Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Machine-learning approaches for the enhancement of 2D and 3D electron microscopy data

A thesis submitted for the degree of
Doctor of Philosophy
School of Physics
Trinity College Dublin

Ph.D. Candidate:
Laura Gambini

Supervisor:
Prof. Stefano Sanvito

2024

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).

Signed: 

ABSTRACT

Electron microscopy allows academic and industrial users to investigate the structure and properties of a variety of materials. However, some limitations are imposed by the instrumentation and the nature of the analyzed specimen. This work proposes machine-learning-based approaches to overcome some of nowadays impediments in the field of 2D and 3D electron microscopy data. In the first project, machine learning is employed to enhance the quality of Scanning Transmission Electron Microscope (STEM) data, effectively reducing noise levels across various electron beam intensities. The algorithm developed undergoes rigorous testing using both synthetic and actual microscopy data. Furthermore, a quantitative and impartial benchmarking protocol for comparing various denoising workflows is proposed, based on the precision of atomic column localization. The second project focuses on STEM data analysis in the context of quantifying vacancies in transition metal dichalcogenides (TMD). Here machine learning improves the quality of STEM-acquired TMD images, facilitating the vacancy-counting process in materials science research. The third project explores the application of a powerful neural network, developed for video-frame interpolation, for the enhancement of 3D tomography. This innovative approach significantly increases the resolution of tomographic images, with applications ranging from materials science, where it aids the study of graphene nanosheets, to medical imaging, where it potentially reduces ionizing radiation doses in Computed Tomography (CT) scans and enhances cardiovascular assessment in coronary angiography videos. Throughout the entire work, a particular effort is dedicated to the development of the techniques needed to quantify the improvement resulting from the application of the proposed methodologies, which are compared to the state-of-the-art approaches. This research development demonstrates the versatility and transformative potential of machine learning in advancing imaging techniques across diverse scientific domains.

Key words: scanning transmission electron microscope, image denoising, autoencoder, neural network, transition metal dichalcogenides, vacancies, FIB-SEM, graphene, MRI, CT, coronary angiography.

LIST OF PUBLICATIONS

- Gambini, L., Mullarkey, T., Jones, L. & Sanvito, S. Machine-learning approach for quantified resolvability enhancement of low-dose STEM data. *Mach. Learn.: Sci. Technol.*, 4(1):010525 (2023).

- Gambini, L., Gabbett, C., Doolan, L., Jones, L., Coleman, J., Gilligan, P. & Sanvito, S. Video frame interpolation neural network for 3D tomography across different length scales. *Under review*

- Gabbett, C., Doolan, L., Synnatschke, K., Gambini, L., Coleman, E., Kelly, A. G., Liu, S., Caffrey, E., Munuera, J., Murphy, C., Sanvito, S. & Coleman, J. 3D-imaging of Printed Nanostructured Networks using High-resolution FIB-SEM Nanotomography. *arXiv preprint arXiv:2301.11046, Accepted for publication at Nat. Commun.* (2023).

- Gambini, L., Douglas-Henry, D., Nicolosi, V. & Sanvito, S. Machine-learning-assisted procedure for vacancy counting in STEM-images of Transition Metal Dichalcogenides. *In preparation*

- Gambini, L., Caldwell, D., Banahan, P., Gilligan, P. & Sanvito, S. Video frame interpolation neural network for the enhancement of coronary angiography videos. *In preparation*

CONFERENCES CONTRIBUTIONS

- **ATOM XVI Workshop**
  Online, May 2021
  Oral Presentation

- **Microscopy Conference (MC) Digital**
  Online, August 2021
  Poster Presentation

- **2nd Joint Symposium of Microscopy Society of Ireland (MSI) and Scottish Microscopy Society (SMS)**
  Galway, Ireland, April 2022
  Flash Talk and Poster Presentation

- **DeepLearn Summer School**
  Las Palmas, Spain, July 2022
  Oral Presentation

- **Microscopy Society of Ireland (MSI) Symposium**
  Cork, Ireland, January 2023
  Oral Presentation

- **Microscopy Conference (MC)**
  Darmstadt, Germany, February 2023
  Poster Presentation

- **AMBER's Internal Conference**
  Limerick, Ireland, April 2023
  Oral Presentation

ACKNOWLEDGEMENTS

This Ph.D. thesis would not have been possible without the contributions and support of many people, which I would like to thank here.

First of all, my supervisor Stefano Sanvito, who guided me during these years with high expertise, dedication and patience. Thank you for having enthusiastically supported the exploration of unexpected paths.

I'm deeply thankful to all my collaborators, from Trinity and outside, who allowed me to develop such an interdisciplinary project. Thank you Lewys Jones, for the uncountable advice and for introducing me to the microscopy community. I would like to extend my acknowledgement to other valuable Trinity members: Cian Gabbett, Luke Doolan, Jonathan Coleman, Valeria Nicolosi and Danielle Douglas-Henry. Thank you all for sharing your expertise and ideas with me, it has been an honour for me to develop projects with you and benefit from your knowledge.

From the Mater Misericordiae University Hospital of Dublin, I would like to thank Paddy Gilligan, David Caldwell, and Paul Banahan. Your passion and expertise in applying AI in healthcare have and will continue to have a profound impact on my career.

I would like to acknowledge Science Foundation Ireland and AMBER for the financial support, and TCHPC and Nvidia for providing computational resources.

Thank you to all the members of the Computational Spintronics Group, including Stefania Negro, for the constant support. Thank you to all the friends I made during my time in Dublin. In particular, thanks to Hugo, Luke, Matteo, Michelangelo, and Mike. I'm grateful we've grown from the original "Ph.D. Newbies" to "Ph.D. Oldies" together. I'm glad to have you all in my life.

Finally, I want to express my gratitude to everyone who stood by my side during this journey, even if doing this Ph.D. meant being apart from them. I am especially thankful to my Mum, my Dad, my sister Giulia, my grandparents, and Giorgio. Thank you for everything you've always done for me.

CONTENTS

CHAPTER 1

INTRODUCTION

ELECTRON microscopes are powerful imaging tools that have contributed to countless discoveries in a wide range of research fields and industrial applications. Much progress has been made since the assembly of the first electron microscope, in 1931 [1]. This prototype, built by Ernst Ruska and Max Knoll, was a Transmission Electron Microscope (TEM) with a resolution (i.e. the smallest distance at which two points can be identified as distinct) of hundreds of nm. Today, aberration-corrected TEMs, operating in scanning mode (AC-STEM), provide the highest resolution of all imaging instruments, below 0.1 nm, and allow one to investigate the structure and chemical composition of materials at the atomic scale [2]. Some of the fields that benefit the most from electron microscopy are material science, where this technology is used for characterizing the structure, composition, and properties of materials at the nanoscale [3]; biology and life sciences, with important results in the observation and study of viruses [4]; pharmaceutical industry, where electron microscopes are widely used to characterize drug structures [5].

Despite these unprecedented achievements, there are still constraints that affect the extent to which electron microscopy can be applied successfully. These limitations are mainly related to the technological state of the microscopy instrumentation and the fragility of the analyzed specimen, with impacts on the quality of the information retrievable from microscopy data. Researchers and engineers are constantly working on possible expansion of electron microscopy capabilities, mainly from the instrumentation perspective. However, it is important to note that these efforts are both time-consuming and expensive, requiring significant investments in hardware production and integration. The development and implementation of aberration

correction in TEM is an example of technological instrumentation advancement in electron microscopy. Before this upgrade, traditional TEM instruments suffered from optical imperfections that degraded image quality and limited resolution. The first aberration-corrected TEM images were published in 1998 [6], obtained using specialized electromagnetic lenses. These correctors allowed explorations at the nanoscale with unprecedented detail and are now standard features in state-of-the-art instruments. Another illustration of progress in the microscopy field is represented by the introduction of Cryo-Electron Microscopy (Cryo-EM) [7]. This is a hardware advancement designed for biological samples, consisting of specialized specimen holders and cooling systems, which allow researchers to freeze biological specimens in vitreous ice. The goal is to preserve the sample's native structure and minimize electron beam-induced damage, making this implementation ideal for studying biological macromolecules and cellular structures. Both the mentioned technological advancements have been instrumental in significant improvement in electron microscopy capabilities, and are, therefore, widely employed. Nonetheless, the financial cost associated with these state-of-the-art infrastructures must be considered [8].

In this research project, a different type of approach will be explored, based on improvements that do not require hardware modifications, a practice that is becoming increasingly more popular in the microscopy community [9]. Specifically, machine-learning-based strategies will be investigated to enhance today's electron microscopy capabilities. The application of some of these methodologies will be also extended to other imaging instruments, belonging to the medical area, in order to demonstrate the model's potential. The aim of this project is to serve as an interdisciplinary overview of the power of computer-aided solutions for the enhancement of imaging instruments.

In a nutshell, machine learning involves developing algorithms able to extract patterns from data of various nature [10]. The learned patterns are then used by these models to make predictions on unseen datasets. Significant advancements in computational power and the increased availability of data have facilitated substantial progress in machine learning, with meaningful impacts on several aspects of contemporary society. Examples of machine-learning tools that can be encountered on a daily basis are recommendation systems for streaming services and e-commerce platforms [11], chatbots and virtual voice assistants [12, 13], email filtering [14].

Notably, applications to visual data, such as images, videos, and volume rendering, are particularly advanced. In fact, machine-learning models can expedite the interpretation and understanding of visual data, leading to several practical applications. For instance, algorithms have been developed for tasks such as image classification (i.e. classifying images into predefined categories), image

segmentation (i.e. dividing images into meaningful regions), object detection (i.e. locating and recognizing multiple objects within an image), video analysis (i.e. tracking objects across video frames), and many others [15]. The majority of these models are trained on the so-called *natural* data, meaning visual representations involving people, animals, and everyday objects. As a consequence of the success of these approaches, an increasing number of algorithms are currently being developed also for applications in other contexts, such as electron microscopy [16] and healthcare [17]. Some of the architectures have been developed from scratch for these specific fields, while others have been adapted from algorithms built for the more general *natual scene* context. This is common practice in the development of machine learning models, which often involves reusing and adapting existing architectures for new purposes. In this project, both development approaches are explored.

An important aspect of any machine-learning scheme is the development of validation methods for the assessment of the model's capabilities. In many cases, a complete understanding of the decision process behind this technology is not obtainable, due to the complexity of the algorithms. Therefore, meticulous results evaluations are needed, to provide an objective measure of model performance and to allow for comparisons between different approaches. Depending on the field of applications and the nature of the algorithms, numerous strategies can be pursued. However, all the approaches share some crucial steps: goal definition, metrics selection, data preparation, and results interpretation. In certain cases, defining evaluation procedures is straightforward. For example, in classification models used to assign labels to input data, such as distinguishing between images of cats and dogs, one can compare model predictions with manually assigned labels. This allows for easy testing of unseen images and unequivocal comparison with human perception. Nonetheless, in other circumstances, it can be quite challenging to identify adequate methods for results assessment. Focusing again on computer vision applications, this is certainly true in the case of algorithms developed to improve image quality. This is a consequence of several factors. Firstly, there may be a lack of a well-defined ground truth, namely the absolute and objectively known information used as a reference for model assessment. Moreover, subjectivity should be considered when evaluating an image quality, as perception can vary from person to person. In general, in this context, there are no standardized metrics that can be used for assessing image quality enhancement. Therefore, the evaluation strategies are developed depending on the specific application, which might require different features from the data.

This can be particularly challenging when dealing with datasets from the medical context, where machine-learning frameworks aim at improving the diagnostic value of the data and assisting doctors in the decision-making process. Indeed,

good-performing algorithms may not necessarily lead to better patient outcomes, which is the ultimate goal of this type of application. In the medical imaging field, it is sometimes difficult to establish objectively whether one image provides more information than another one, and the opinion of certified medical experts is often required.

The approach pursued in this work, aimed at providing a quantitative results assessment, is to focus on the information retrievable from the data, intrinsic to the examined application. Therefore, particular attention is devoted to the identification of adequate metrics, which should correspond to quantities that are commonly used to extract insights from the analyzed data. Throughout the entire research development, the advantages and limitations of different assessment techniques will be discussed.

With the primary objectives and strategies outlined, attention can now shift toward the actual implementation of the individual projects.

Firstly, a description of all the methodologies involved in the research study will be provided in Chapter 2, which is divided into three main sections. Starting from a general description of electron microscopy, the first section will focus on Scanning Transmission Electron Microscope (STEM) and on Focussed-Ion-Beam Scanning-Electron-Microscope (FIB-SEM) tomography. In order to motivate the projects finalized within this research development, some of the limitations of these imaging instruments will be detailed. The second section will introduce the topic of machine learning and some of the related concepts, with particular attention on the models that will be useful for the purposes of this work, namely autoencoders and neural networks developed for video frame interpolation. Lastly, an overview of some medical imaging techniques will be presented, including a discussion on the advantages and drawbacks of the application of machine-learning approaches in the healthcare context.

The purpose of the first project, presented in Chapter 3, is to overcome one of the main limitations of Scanning Electron Transmission Microscope (STEM) data. These are extremely powerful imaging instruments, which allow achieving atomic resolution images, as mentioned previously. However, this usually involves the use of a high electron dose, which can lead to specimen damage and affect the observation. As a consequence, the application of high-resolution microscopy is usually limited to non-beam-sensitive materials. The solution proposed in this work is to use a strategy, based on machine learning, to significantly improve the quality of STEM data acquired at low electron dose, strongly affected by Poisson noise, an effect that cannot be corrected at the instrumentation level. The developed algorithm, namely an autoencoder, is trained on synthetic data and it is subject to rigorous testing using both synthetic and actual microscopy data. Following a first qualitative model evaluation, achieved through visual comparison, more objective

assessment approaches will be investigated. This analysis will demonstrate that the presented framework can effectively reduce noise levels and approximate ground-truth precision across a wide range of electron beam intensities. Importantly, no human data pre-processing or explicit dose knowledge is required, and it operates at a speed compatible with real-time data acquisition. Furthermore, a quantitative and unbiased benchmarking protocol will be introduced, based on the evaluation of the atomic column localization. This is in accordance with the evaluation strategy introduced above, whose goal is to employ metrics related to the physical properties intrinsic to the information content that can be retrieved from the data. The main goal of this first project is to propose a scheme, rooted on machine learning, that can be ideally applied to any STEM investigation. Having demonstrated that this approach is valuable and facilitates obtaining insights from the data, the following project, detailed in Chapter 4, describes a possible practical use of this type of strategy.

The aim of this next chapter is to illustrate how a machine-learning model can assist STEM data analysis, with a practical application to vacancies investigations in transition metal dichalcogenides (TMD). These materials present many interesting properties, which make them promising candidates for applications in several fields. However, these properties can be altered, positively or negatively, by the presence of vacancies, which should therefore be properly quantified. A common strategy to assess the quality of TMD is to image them with a STEM. However, it can be challenging to identify light atoms, such as chalcogens, even when the images are acquired with technologically advanced instruments. The approach suggested in this study involves employing a machine-learning model to enhance the quality of STEM-acquired TMD images, and therefore ease the vacancy-counting process. A procedure for the quantification of these defects will be proposed and discussed. The major factors affecting the results will be also reviewed.

The following project revolves around a different imaging technique, namely three-dimensional (3D) tomography, with the main focus on 3D volumes achievable with FIB-SEM (Focussed-Ion-Beam Scanning-Electron-Microscope) technology. 3D tomography represents a powerful investigative tool for many scientific domains, going from materials science, to engineering, to medicine. This is realized through a multitude of experimental techniques, that can image objects of the most diverse nature and across many length scales. Many factors can limit the 3D resolution, which often remains spatially anisotropic. This undermines the ability to achieve cubic-voxel definition in the three dimensions, a fact that hampers the precision of the information that one can extract from the 3D reconstructions. The solution proposed in this work is to use a powerful neural network, developed for video-frame interpolation, to augment tomographic images and to bring them to cubic-voxel resolution. The aim of the numerous neural networks devel-

oped for video frame interpolation is to increase the frames per second (*fps*) of a video by generating one or more frames between the existing ones, resulting in a smoother and visually fluid motion. In the proposed work, this method is applied to radically different situations. The ground truth is not available for all circumstances and, as a consequence, different assessment strategies are used. As a first application, the morphology of ink-jet printed networks of graphene nanosheets will be investigated, obtained by milling a specimen with a Focused Ion Beam (FIB), while imaging with a Scanning Electron Microscope (SEM). For this FIB-SEM technique, the resolution is in the range of nanometers to tens of nanometers and it is limited by the destructive milling. This work will demonstrate how a neural network developed for a different purpose, namely the interpolation of video frames, can be successfully implemented to increase the resolution of FIB-SEM-generated data, at different levels of complexity. In order to quantitatively validate the results, several metrics will be implemented. Specifically, in addition to conventional computer-vision metrics, physical parameters that can be derived from the resultant 3D reconstructions, such as network porosity and tortuosity, will be evaluated and compared across different image interpolation techniques. For this analysis, images are removed from the original dataset and used as ground truth.

The extensive applicability of the proposed method will be demonstrated by implementing it on datasets of different scales, namely medical images of various types. For these cases, the resolution is in the millimeter range and it is limited by both the instrumentation and the necessity to keep the acquisition time low, primarily for the patient's benefit. The first case consists of magnetic resonance imaging (MRI) acquisitions of the human brain. Being the resolution already isotropic, frames can be removed from the original dataset and used as ground truth, for validation purposes. Specifically, metrics such as the gray-matter volume variation [18] will be investigated. Then, X-ray computed tomography (CT) scans of the abdomen region will be considered. In this context, the use of image interpolation methods could potentially imply a reduction of the released ionizing radiation dose, which is known to have several negative biological effects on humans [19]. For this example, being the resolution of the 3D volume significantly anisotropic, no ground truth can be extracted from the original data source. Therefore, the noise power spectrum is evaluated for correspondent areas of the original and artificially augmented datasets, to investigate the image quality enhancement. Furthermore, the same video frame interpolation strategy will be applied to videos of coronary angiography, a medical procedure using a contrast dye and real-time X-ray imaging to assess the cardiovascular system. The main limitation of this practice is related to the release of ionizing radiation associated with each X-ray-generated frame. Integrating a video frame interpolation technique into this procedure could allow

a reduction of the harmful radiation delivered to both patients and practitioners.

Lastly, Chapter 6 will provide a comprehensive overview of the key findings and outcomes derived from the study, offering insights into the possible applications of this research. Moreover, it will outline potential areas for future exploration.

To summarize, the aim of this project is to demonstrate how machine-learning-based approaches can expand today's capabilities in the field of electron microscopy, with some applications extended to the medical imaging area. The goal of the proposed tools is to enhance various aspects of the imaging systems, such as expanding the range of analysable materials, accelerating the acquisition process, and improving the quality of the extracted information. Ultimately, this work wants to contribute toward the integration of machine-learning algorithms into imaging devices, enhancing their performance and leading to new discoveries.

CHAPTER 2

METHODOLOGY

THIS chapter aims at introducing the primary methodologies involved in this research project, namely electron microscopy, both from the experimental and simulation point of view, and machine learning. An overview of medical imaging techniques is also provided, in support of the last section of Chapter 5.

2.1 Electron microscopy

The term microscopy defines the discipline of inspecting small objects that would not be visible by the naked eye, but require using an instrument named microscope, from the Greek words *mikros* (= *small*) and *skopein* (= *to look at*). These devices are generally classified into two types, depending on the nature of the wave employed to interact with the specimen: light and electron microscopes.

The invention of the first microscope is not certain, but it is usually attributed to Hans and Zacharias Janssen, who built, in the 16th century, the first so-called compound light microscope, namely a magnifying instrument with more than one lens [20]. In a light microscope, also known as *optical* microscope, a set of lenses is used to focus visible light on a sample and to bend the light, allowing the magnification of the image [21]. The resolution of light microscopes, which is defined as the shortest distance between two points identifiable as distinct, is limited by several factors. A theoretical value for the resolution limit can be found by using the Abbe diffraction limit relationship [22],

$$d \approx \frac{\lambda}{2 \times NA}, \tag{2.1}$$

where d is the resolution limit, λ is the light wavelength, and NA is the numerical aperture of the objective lens. This last term is a dimensionless quantity that describes the range of angles the system can accept for the incoming light and depends on the refraction index, n, of the medium between the objective lens and the sample, and the half angular aperture of the objective, μ, according to the formula,

$$NA = n \times \sin \mu. \tag{2.2}$$

In the case of air between the lens and the specimen, and green visible light, the resolution is limited to approximately 200 nm. It is worth mentioning that modern microscopes can achieve higher resolution, by employing techniques that belong to the super-resolution optical microscopy field [23]. However, the main limiting factor remains the wavelength of the light, a limitation that can only be overcome by using a different beam, made of electrons, i.e. by using electron microscopes. In fact, electrons have a typical wavelength 100,000 shorter than that of visible light, and therefore they show a much higher resolving power, of the order of 0.1 nm. The invention of the first electron microscope dates back to 1931, when Ernst Ruska and Max Knoll built the first prototype, which achieved a resolution of hundreds of nm [1].

Both light and electron microscopes present advantages and drawbacks. They are chosen depending on the specific needs of the application. For instance, light microscopes are more beneficial for studying live specimens and larger structures, such as in the context of biological research [24]. In contrast, electron microscopes are more suitable when atomic-level details are needed, with applications ranging from material characterization to nanotechnology [25]. For this project, the focus will be on electron microscopes.

Electron microscopes are usually classified into two main categories: Scanning Electron Microscope (SEM) or Transmission Electron Microscope (TEM). This distinction is mainly based on the type of electrons used to generate the image. In fact, different families of electrons are involved in the imaging process of electron microscopes and carry different types of information. The so-called *primary electrons* are the high-energy constituents of the incident electron beam. They can interact with the specimen following different mechanisms. The primary electrons can interact elastically with the atoms from deeper regions of the sample and generate backscattered electrons, meaning that the primary electrons undergo a change in trajectory and are reflected back, with approximately the same energy; they are highly sensitive to differences in atomic number and therefore carry information on the sample's composition. The primary electrons can also interact inelastically with the atoms of the specimen, mainly on the surface or near-surface region. They have lower energy than the backscattered electrons and provide information about the topography of the specimen surface. If the sample is sufficiently thin, the

electrons can also be transmitted through it, after interacting with the internal atoms.

Secondary electrons and backscattered electrons are collected and used to reconstruct images in the case of SEM, which operates by scanning a focused electron beam across the surface of a specimen. State-of-the-art SEMs achieve a 0.4 nm resolution [26], which is limited by factors such as the electron probe size and the volume of interaction between the electron beam and the specimen. SEMs are widely used to study the external morphology and chemical composition of solid objects. When combined with a FIB (Focused Ion Beam) instrument, they can be employed to investigate the 3D internal structure of materials at the nanoscale, as it will be described later in this chapter.

TEMs are used to investigate the crystal structure and their operation is based on transmitted electrons. Specifically, an electron gun produces an electron beam, which is accelerated and transmitted through a thin specimen. Subsequently, the electron beam passes through a series of electromagnetic lenses, which play the same role as optic lenses in light microscopes [27]: they are employed to produce a magnification of the specimen. Lastly, the information carried by the electrons can be recorded, in the form of an image. Depending on the technique used to address the specimen, two types of transmission electron microscopes can be distinguished: conventional TEM (CTEM) and TEM operating in a scanning mode (STEM) [2]. CTEM is characterized by a wide-beam approach, in which a close-to-parallel beam invests the entire area of interest [28]. This apparatus will not be considered in this project.

2.1.1 Scanning Transmission Electron Microscopes (STEMs)

A more sophisticated technique is implemented in STEM, which operates by focusing a convergent electron beam on a small area and by scanning it across the sample. This approach presents some advantages, such as highly controlled positioning of the electron beam, the possibility to collect additional signals such as secondary electrons and scattered beam electrons, improved resolution, and easily interpretable data. One of the STEM available at Trinity College Dublin, namely the state-of-the-art Nion UltraSTEM, is displayed in Fig. 2.1, both from the outside (on the left-hand-side panel) and the inside (on the right-hand-side panel). Fig. 2.2 displays a schematic of the STEM structure. Electrons are emitted from the heated tip of an electron gun, made of a material with a high melting point, such as Tungsten. Depending on the population of electrons considered for the image formation, the main imaging modes are the bright field (BF) and the dark field (DF). As shown in Fig. 2.2, the BF detector is placed in the path of the electrons transmitted through the specimen; in this case, only the unscattered electrons contribute to the image formation. In bright field images, crystalline or high-mass

Figure 2.1 – Pictures of the state-of-the-art Nion UltraSTEM available at Trinity College Dublin, displaying the microscope from the outside (on the left-hand-side panel) and from the inside (on the right-hand-side panel). The pictures are courtesy of Danielle Douglas-Henry.

density areas appear dark on a white background. In contrast, DF detectors collect the electrons scattered out of the path of the electron beam, therefore the atom areas appear bright on a dark background. Examples of a sample image taken with the two imaging modalities are displayed in Fig. 2.3, which shows Graphene data simulated with the simulation software Prismatic [30, 31], described in the following section. DF images are particularly valuable due to the fact that the signal is chemically sensitive; the generated images show different levels of contrast, which depend on the chemical composition of the analyzed specimen [32]. Specifically, the image intensity is proportional to the atomic number, Z, of the atoms in the specimen: elements with a high Z have strong scattering interaction with the incident electron beam, and therefore the image intensity is higher for the correspondent pixels. This mechanism is known as Z-contrast and allows easy interpretation of the data, making it possible to distinguish different atomic species [33, 34].

Limitations of STEMs

Despite being the instrument that provides the highest resolution, below 0.1 nm, STEMs present some limitations. Sample preparation and environmental constraints are critical factors that can significantly impact the feasibility and success of experiments in electron microscopy. For instance, the specimen should satisfy

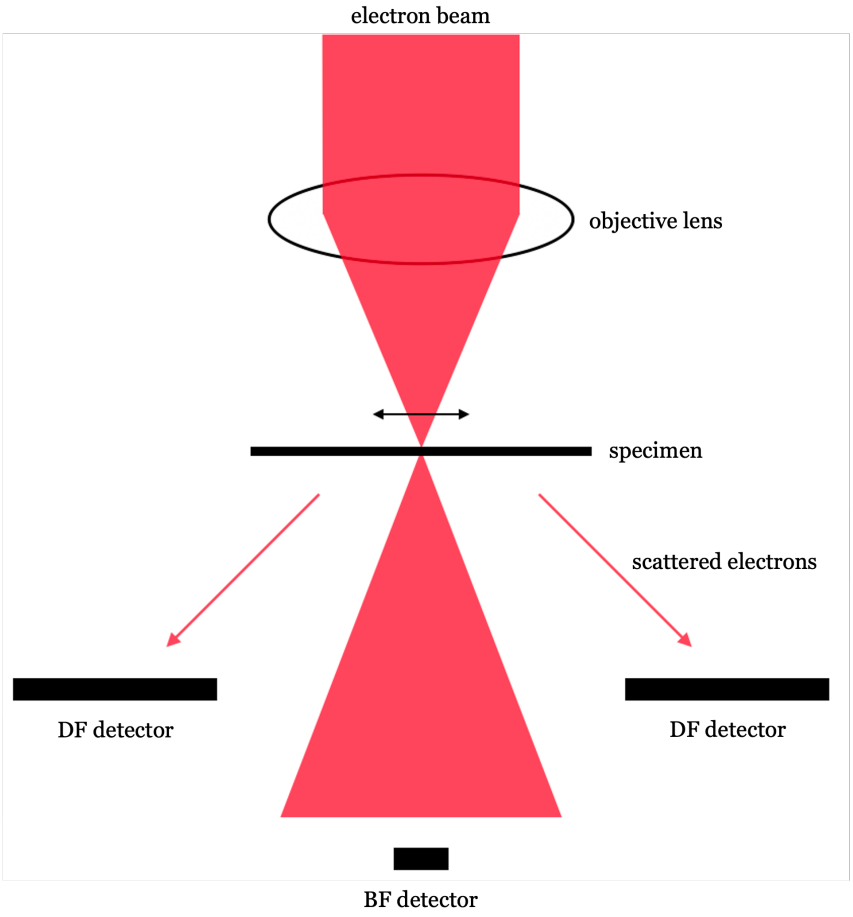

Figure 2.2 – Schematic representation of a STEM. Simplified version of Fig. 2.3 of reference [29]. An electron beam passes through some objective lenses, which focus the beam on a small area of the specimen. The beam is then scanned across the entire sample. The electrons, which pass through the specimen, are collected by the bright field detector (BF), while the scattered electrons are collected by the dark field detectors (DF).

specific thickness requirements and a high-vacuum environment should be maintained for the experiments, to ensure image quality. Certainly, these conditions pose some challenges on the range of materials that can be analysed. Nonetheless, even when these demands are fulfilled, there are additional constraints that can hinder the analysis. One of the major restrictions is that atomic resolution is achievable only when the specimen is illuminated by a very intense electron beam. This maximizes the signal-to-noise ratio, but may damage the sample and the observation. In fact, the damage is a function of the electron dose, defined as the total number of electrons per unit area hitting the specimen, and it is caused by various energy-loss mechanisms. The most common ones are knock-on damage and radiolysis or ionization damage. In the case of knock-on damage, the atoms of the sample are displaced from their sites due to a transfer of momentum from the incident electrons [35]. The specimen can be also affected by radiolysis, which is

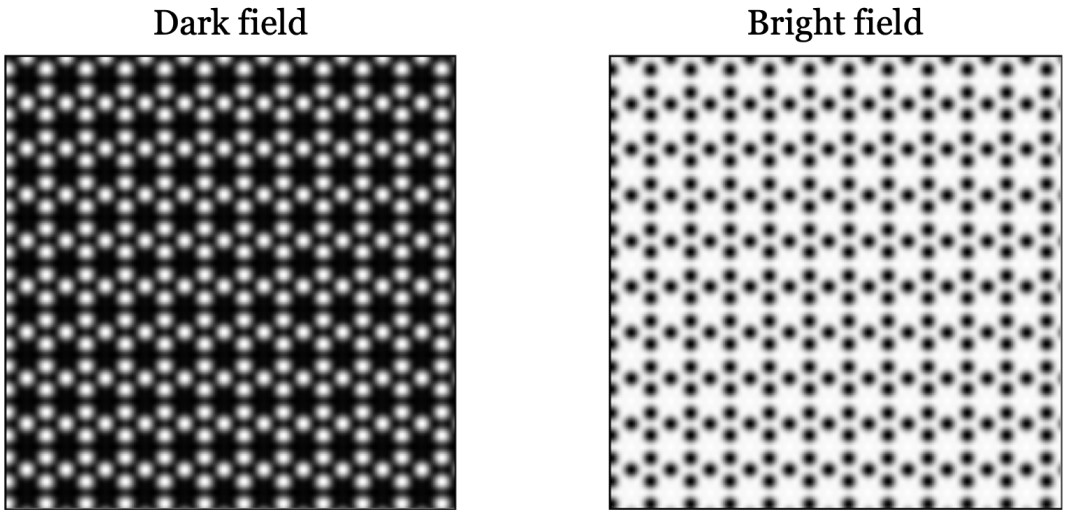

Figure 2.3 – Example of Graphene STEM data, imaged with the dark-field (on the left-hand-side panel) and bright-field (on the right-hand-side panel) techniques. The images are simulated using the simulation software Prismatic [30, 31].

a consequence of the inelastic scattering of the incident electrons: the transfer of energy from the incident electrons to the specimen electrons can lead to the formation of excited states and consequent structural and chemical modifications [36]. In both these conditions, the specimen under investigation changes in time as the measure progresses, compromising the experiment result.

The sample integrity can be protected by decreasing the electron dose, but this leads to a deterioration of the image quality, thus reducing the chance of extracting useful information from the data, such as the atom's position and the identification of defects. The reason behind such a loss of resolution can be ultimately identified with the presence of Poisson noise, which increases upon reducing the number of incident electrons [37], according to the relation,

$$f \propto \frac{1}{\sqrt{\rho}}\,, \tag{2.3}$$

where f is the noise and ρ the dose. Poisson noise is related to the quantized nature of the electron beam, meaning that it is related to the discreteness of the electrons. Indeed, the electron beam can be described as a movement of discrete packets (i.e. the electrons). During the imaging process, a discrete number of electrons, independent from each other, reaches a specific location of the sample, at a constant rate. Processes of this type are described by a Poisson distribution.

A Poisson distribution is expressed by the relationship:

$$p(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{2.4}$$

where $p(k)$ is the probability of an event happening k times, and λ is the mean number of events, which is assumed to be constant. It is a discrete probability distribution, which means that involves the probability of a discrete outcome. In this case, it is the number of times a specific event happens. It is also required that the events are independent from each other. For a high value of the mean number of events, the Poisson distribution can be approximated with a normal (i.e. Gaussian) distribution. In practice, in the electron microscopy context, this means that if the electron dose is increased (each sample location is illuminated by more electrons) the collected signal will be less noisy.

In contrast to other types of signal distortions, such as Gaussian noise, scan noise, and drift [38], Poisson noise cannot be eliminated by improving the instrumentation or by changing the working conditions, due to its nature. For instance, it is possible to eliminate Gaussian noise completely by replacing standard acquisition with electron counting, a strategy that itself represents one of the latest electron microscopy advancements [39]. Within this thesis, in accordance with the work [39], the standard acquisition method will be referred to as *analog* acquisition, while the electron-counting process will be referred to as *digital*. The digital approach was developed with the purpose of obtaining high signal-to-noise images even at low-dose settings. It consists in pulse-counting the individual electrons that are scattered to the dark field detector. Among the advantages of this strategies, detailed in [39], it is worth mentioning the ability to generate images, where the pixel intensity corresponds to the number of electrons detected at a given pixel, namely the pixel intensity is a directly observable quantity. In our opinion, digital data acquisition represents the ultimate future of electron microscopy, a consideration that motivated our choice of considering only Poisson noise in the work presented in Chapter 3.

Simulation of STEM data

Over the last decades, numerous tools have been developed for the simulation of STEM images, which plays an essential role in several aspects of microscopy, such as planning, optimization, and interpretation. In fact, the simulation of data under different imaging conditions allows one to investigate and optimize the microscope's setting parameters for the experimental measurements. In the case where the sample structure is known, it is possible to use simulations to compare the experimental data with the theoretical representation provided by the simulation. This helps the interpretation of the investigated structure.

Furthermore, simulation data are particularly useful for the training and testing of machine-learning models. The use of experimental datasets in this context is not advisable for two main reasons: first of all, the dataset preparation could require unsustainable time and cost. Secondly, in many cases, the so-called ground truth,

namely the goal *perfect* image is needed to evaluate the model's performance during training, and this is not achievable with experimental acquisition. Some papers propose simple linear methods to generate a synthetic dataset [40]. However, despite being fast approaches, electron microscope images do not follow a simple linear image model and therefore any linear method cannot be quantitatively precise [29]. In fact, in order to have a more realistic dataset, it is advisable to use simulation techniques that implement the Bloch-wave method [29] or the multislice algorithm [41]. These quantum mechanical techniques employ a detailed description of the specimen and of the instrument settings to generate the images, meaning that they implement a more faithful simulation of an actual measurement. Some of the information provided by the user are: location and atomic number of each atom in the test structure, specimen thickness, energy of the electron beam, and others.

The mentioned simulation techniques propose two different approaches for modelling the effect that the beam-specimen interaction has on the probe wave function, which can be found by solving the Schrödinger equation for fast electrons travelling in the z-direction,

$$\frac{\partial \psi(\mathbf{r})}{\partial z} = \frac{i\lambda}{4\pi}\nabla^2_{xy}\psi(\mathbf{r}) + i\sigma V(\mathbf{r})\psi(\mathbf{r}), \tag{2.5}$$

where $\psi(\mathbf{r})$ is the electron wavefunction, describing the electron beam, $\mathbf{r}$ is the 3D spatial coordinate with components (x, y, z), λ is the incident electrons wavelength, σ is the interaction parameter, and $V(\mathbf{r})$ is the sample's electrostatic potential [29, 31]. The term on the left side of the equation represents the evolution of the electron wavefunction with respect to the distance along the beam axis. The first term on the right side of the equation is a free-space propagator operator, which measures the gradient of the wavefunction at every location and describes the behaviour of the wavefunction in free space, in the absence of electrostatic potential. Lastly, the second term on the right side is the electrostatic potential.

According to the Bloch-wave method, the interaction between the electron beam and the specimen can be described following the Bloch's theorem [42], valid under the assumption of crystal-structure periodicity. This states that the electron waves can be modelled as a product of a plane wave and a periodic function. In practice, the application of this method requires calculating a scattering matrix which, multiplied by the incident wavefunction, returns the exit wavefunction. Building the matrix involves the eigen-decomposition of approximations of Eq. (2.5), this will provide us with a basis set, which is used to simulate different probe positions on the sample surface, by means of weighting coefficients. This process can be extremely time-consuming for simulations of large samples and it is therefore only applied to the case of small STEM simulations.

The multislice method is more advanced and flexible for applications to specimens with defects and to amorphous materials [29]. According to this algorithm, the sample structure is split into several thin slices, whose interaction with the electron beam is described by the weak phase object approximation, valid for thin samples [29]. According to this approximation, the electrons in the imaging beam undergo a small deviation of their wavelength when passing through the specimen, due to the higher energy they have compared to the specimen's electrons. This step corresponds to solving the second term on the right-hand side of Eq. (2.5), meaning to compute the 2D projected potential for each slice. After the electron wave passes through one of the slices, the propagation to the next one is described by Fresnel diffraction in the free space between slices. This is the diffraction (i.e. the deviation of a wave from its propagation) that happens when the distance between the source and the obstruction is comparable to the obstruction's size. This process is repeated for the entire sample and, under appropriate conditions, the wavefunction exiting from the bottom slice of the sample is the simulated image. This technique can become excessively slow for large simulations, where the transmission and propagation steps are repeated for each probe position.

A more sophisticated simulation technique was developed in 2017, namely, PRISM (plane-wave reciprocal-space interpolated scattering matrix), which incorporates features from both the previously described methods and is implemented in the Prismatic simulation software [30, 31]. This algorithm delivers a significant acceleration in the simulation, a fundamental aspect when dealing with the generation of large datasets. The initial steps of PRISM are the same as in the multislice method: the sample is divided into a series of thin slices along the beam directions and the projected potential is computed for each of them. This describes how the atoms in the slice interact with the electron beam. In traditional multislice simulations, the propagation is independently evaluated for each STEM probe position. In contrast, in the case of the PRISM algorithm, a compact scattering matrix is computed, for a basis set of the incident plane waves. This is a mathematical representation of how the incident wavefunction is transformed when passing through the specimen. Notably, this scattering matrix is only computed once and can be reused for any probe wavefunctions to model their propagation, with great impact on the model speed. Significant acceleration of the computational time is also facilitated by the definition of an interpolation factor, necessary for Fourier interpolation of the scattering matrix.

2.1.2 FIB-SEM tomography

The SEM, briefly described at the beginning of this chapter, can be used in combination with a Focused Ion Beam (FIB) to obtain an imaging system known as FIB-SEM, which allows for 3D imaging of nanoscale specimens. The imaging

procedure, schematically illustrated in Fig. 2.4 [43] involves a FIB, which mills away slices of a specimen, while a SEM takes images of the exposed planes, as displayed in the left panel of the figure. After repeating this process for the entire specimen, the outcome is a stack of hundreds of 2D images. These are then aligned and used to produce a high-fidelity 3D reconstruction [44], schematized in the middle and right panels of Fig. 2.4, respectively. The tomographic reconstruction can be produced by using software such as Fiji [45] or Dragonfly [46].

A FIB is an instrument that utilizes a beam made of ions, typically generated from a Gallium source, for different purposes: imaging, when operated at low beam current, or milling, at high beam current. In the context of FIB-SEM tomography, the FIB is used at a high beam current, as a milling instrument. The Gallium ion beam is generated by an ion gun and then accelerated, usually with voltages in the range of $1 - 30$ kV. Subsequently, the beam goes through condenser and objective lenses, and it is scanned across the sample surface.

The milling phase is based on ion-atom elastic collisions. Specifically, atoms from the specimen are removed, when the incident ion is able to transfer enough kinetic energy to overcome the binding energy of the sample's atoms. The incident ions can also undergo different processes. They can lose their energy after a cascade of collisions and stop inside the specimen, according to a process known as ion implantation, which is correlated to radiation damage. They can be backscattered and deposited on the specimen surface or they can interact inelastically with the target's atoms and produce secondary electrons and other particles [47, 48]. The main factors that impact the milling rate are the angle of incidence, the type of target material, and the voltage that is used to accelerate the ion beam. Notably, the milling capability of FIB instruments is not only used for FIB-SEM tomography, but also for sample preparation, mainly for TEM investigations [48]. This application will not be covered in this work. More advanced FIB-SEM systems are also combined with instruments for chemical and crystallographic analysis in 3D data, such as electron dispersive spectroscopy (EDS) and electron backscatter diffraction (EBSD) [49]. FIB-SEM tomography is widely used in material science, biology, natural sciences, semiconductor industry, nanotechnology, and other fields [50, 47, 51, 52]. Recent studies showed that FIB-SEM tomography is particularly suited for the analysis of printed nanostructured networks [44]. In this case, due to the length scale involved, it is more appropriate to describe the technique as FIB-SEM nanotomography (FIB-SEM NT). The main dataset used for the project described in Chapter 5, belongs to this category.

Limitations of FIB-SEM tomography

The presented imaging technique is characterized by some limiting aspects, which motivated the work described in Chapter 5 of this research project. Firstly, FIB-SEM

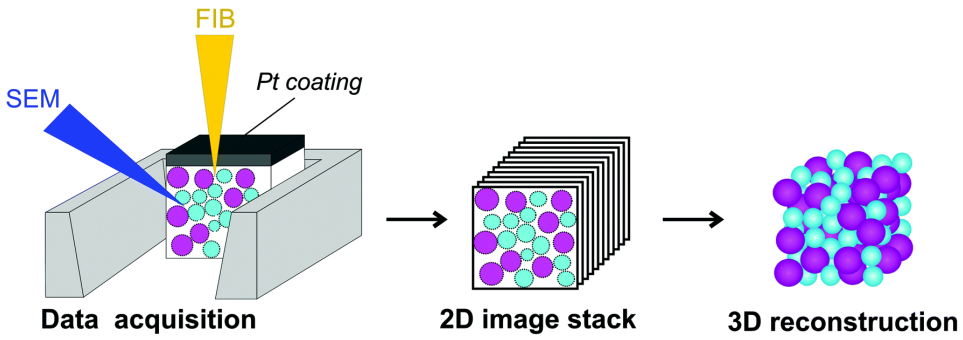

Figure 2.4 – Schematic illustration of FIB-SEM tomography, from reference [43]. The panel on the left shows the acquisition process, which consists of serially removing slices of material with the FIB (yellow beam in the figure), and imaging the resulting cross-section with the SEM (blue beam in the figure); the black layer on top of the specimen is a platinum coating, necessary to protect the sample and prevent artefacts during the imaging process. The middle panel displays the stack of 2D images obtained at the end of the experimental procedure. The panel on the right illustrates the 3D reconstruction of the specimen, obtained from the 2D stack.

tomography is a destructive procedure, meaning that the sample is destroyed at the end of the experiment and cannot be recovered. This implies that, in case not enough data was collected, it is impossible to experimentally acquire additional 2D frames.

Another limitation of this technique is that the resolution of the generated 3D volume is often anisotropic, especially when working at high resolution. In fact, while the cross-section (also referred to as the xy-plane) is imaged at the SEM resolution, about 5 nm in the case of high-resolution images of the graphene dataset used for the project developed in Chapter 5, the resolution along the milling direction (the z-direction) corresponds to the slice thickness and it is usually around $10 - 20$ nm. As a consequence, the reconstructed 3D volume will not be characterized by cubic voxels (the 3D equivalents of pixels). Note that cutting thinner slices is hindered by the instrumentation, the nature of the specimen, and by economic constraints. Moreover, a reduction of the slice thickness implies a detriment of the resolution in the xy-plane, since the damage produced during one cut can propagate to the following one. This limitation is linked to another problem of FIB-SEM instruments, namely the slow imaging speed [53], which has an impact on the overall economic cost associated with the procedure. Therefore, considering all these issues, one would then desire a method to interpolate images, which preserves and possibly enhances the information quality and ideally reduces the number of milling steps to perform. A solution to this problem is proposed and discussed in Chapter 5.

2.2 Machine learning

Since ancient times, philosophers believed that the process of human thinking could be automatized and reduced to mechanical calculation [54, 55]. This ambition of replicating human intelligence in machines was and still is driven by the desire to enhance decision-making, automate tasks, improve efficiency, and tackle complex challenges. The attempts to perform, with the aid of computers, tasks usually done by humans are enclosed in the discipline of Artificial Intelligence (AI), which was officially founded in 1956 [56]. Since then, the development of computer science and the availability of data allowed much progress in the field, with extensive impact on a variety of areas of nowadays society: robotics [57], healthcare [58], finance [59], education [60], and many others.

Machine learning (ML) is a subcategory of the more extensive field of AI. It refers to the practice of using algorithms to extract patterns from data and use this information to *learn* how to perform specified tasks. Despite being often used interchangeably, ML and Deep Learning (DL) do not express exactly the same concept. In fact, DL is a subfield of ML. It does not require structured data as input and automatizes the feature extraction pre-processing necessary for ML models.

The most common algorithms, which are the key component of DL, are Neural Networks (NNs), models that try to mimic the neurons of the human brain and their connections. NNs are universal function approximators, meaning that, given the right architecture and complexity, they can approximate any mathematical function.

The architecture consists of one input layer, one or more hidden layers, and one output layer. The number of hidden layers defines the depth of the algorithm; in order to be identified as a deep-learning algorithm, a neural network must have at least three hidden layers. Each layer consists of a set of individual units known as nodes or neurons, whose task is to receive some inputs from other neurons and produce an output, after processing the inputs. Many neural network architectures have been designed in the past decades. In the next sections, the models employed for the development of this research project will be described.

The main difference between classical programming and ML lies in the kind of input and output involved. In fact, classical programming accepts some data and some rules as input and delivers some answers as outputs. In contrast, machine learning receives data and answers as input, and from this information is able to retrieve, during a process called *training*, the rules that link them. The described process belongs to the so-called *supervised* type of learning, where labels (namely the answers) for the input data are available; this is used for classification, regression, and other applications. When the dataset is unlabeled, the machine-learning model can be trained to find patterns by clustering the data. In this case,

the learning is called *unsupervised* and can be used for dimensionality reduction and other tasks. The last main type of learning is *reinforcement* learning; this does not need labelled data and is used to find the best behaviour, in specific contexts, to obtain the maximum reward, for example when performing autonomous driving tasks.

In order to explain the ML workflow, let us consider the case of supervised learning. A fundamental aspect of every machine-learning project is the dataset, which usually undergoes some pre-processing steps to ensure it is suitable for the model's purpose. The first step is usually *data cleaning*, which consists of removing input data that may hinder the training of the models because they are corrupted, duplicated, or for example formatted incorrectly. A process called *data augmentation* is commonly performed before (or during) the training of the algorithms. This is a strategy used to artificially increase the size of a dataset, by generating new elements from the modification of existing ones. Depending on the type of data, different approaches are available. For instance, in the case of images, elements of the training set can undergo rotation, shift, magnification, change of brightness, and others. Finally, before the training phase begins, some datasets are subject to *data transformation* (sometimes referred to as *data preparation*), to facilitate the learning process. For instance, data can be normalized to ensure that all elements are on the same scale of values.

For our example, each element of the dataset $D(x, y)$ is a pair of an input x and a target or label y. This collection of data is split into three subsets, namely the training set $D(x_{\text{train}}, y_{\text{train}})$, the validation set $D(x_{\text{val}}, y_{\text{val}})$, and the test set $D(x_{\text{test}}, y_{\text{test}})$. The training set is used by the model to learn patterns from the data that will be used to make predictions on new examples (a phase that is known as *training*). Specifically, it is used to feed a model M and get an output y^*_{train} in the form $y^*_{\text{train}} = M(x_{\text{train}}, \theta)$, where θ is a set of parameters that needs to be optimized. The parameters θ must be optimized in order to minimize the discrepancy between the computed y^* and the expected value y. This discrepancy cannot be evaluated on the training set, because it would prevent the model from generalizing[1] properly to unseen data. It is necessary to use an unbiased dataset, known as validation set. Specifically, the training set is used by the model to identify a set of parameters θ to fulfill the model's purpose. The suitability of these parameters is validated by using the model on a different dataset, namely the validation set. The discrepancy between the expected y_{val} and the computed $y^*_{\text{val}} = M(x_{\text{val}}, \theta)$ is evaluated in terms of a loss function, whose definition varies depending on the purpose of the model and the type of data. Some examples of commonly used loss functions are,

[1] In the machine-learning context, the term generalization refers to the ability of a trained model to perform well on new data, not encountered during the training process.

- Mean Squared Error (MSE), which can be defined as,

$$MSE = \frac{\sum_i^N (o_i - r_i)^2}{N}, \tag{2.6}$$

  where N is the number of data points under evaluation, o_i is the original data and r_i is the reconstructed data.

- Binary Cross Entropy, often used for classification problems and mathematically expressed as,

$$Loss = \frac{1}{N} \sum_i^N - \left[y_i * log(p_i) + (1 - y_i) * log(1 - p_i) \right], \tag{2.7}$$

  where N is the number of data points under evaluation, y is a binary indicator of the class (0 or 1), p_i is the probability of class 1, $(1 - p_i)$ is the probability of class 0.

The most commonly used process for training a neural network is a two-step procedure consisting of forward pass and backpropagation. At first, the input data is passed through the neural network and each layer performs specific calculations using an initial set of parameters, to produce an output. This prediction is then compared to the target value or ground truth, resulting in a loss that quantifies the discrepancy between the predicted and the expected results. In the subsequent step, known as backpropagation, this loss is used to calculate gradients, which indicate how much each parameter in the network should be adjusted to minimize the loss. The backpropagation process begins from the output layer and moves backwards through the network layers to compute the gradients. These are used to guide the optimization algorithm, a mathematical function used to adjust the model's parameters, in the direction that reduces the loss. It is worth mentioning that during training, data is often divided into batches (subsets), and parameter updates are calculated based on the average gradient computed from each batch. This approach helps the training process converge efficiently. Some examples of commonly used optimizers are: stochastic gradient descent, Adam, RMSProp, and AdaDelta. More details on the advantages and disadvantages of the different kinds of optimizers can be found in [61]. The described procedure is repeated for multiple iterations, also known as epochs, until a specific stopping criterion is met. This criterion depends on the tackled problem. It often involves monitoring the convergence of the loss or achieving a desired level of accuracy on a validation dataset.

Once this process, namely the training, is completed and all the parameters have been optimized, the model can be used on unseen data, namely the test set, to evaluate its performance and limitations. In the case the results are not

satisfactory, one can follow different approaches, such as increasing the training and validation set size, and changing the model's hyperparameters (the optimizer, the loss function, the number of layers, the number of nodes, etc.).

The operation used to find the most advantageous set of hyperparameters for a specific model and dataset is known as *hyperparameter tuning*. It consists in assigning different values to the set of hyperparameters and separately training the model, with each different set of values. Several techniques are available for this purpose, the most commonly used are:

- grid search, which consists of creating a set of possible discrete hyperparameter values, namely a grid, and training the model with every possible combination of those values. This investigation is comprehensive, but it can be extremely slow when several hyperparameters need to be optimized;
- random search, which, in contrast to grid search, selects only a limited number of combinations of the possible hyperparameters value chosen randomly. This ensures less computational time and avoids biases linked to the user's choice of values. However, the resulting combination of hyperparameters might not be the best possible, being this method not exhaustive;
- Bayesian optimization, which takes into consideration the results of the previous evaluation and uses a probabilistic function to select a new combination of hyperparameters. This method allows one to find an adequate set of values after a relatively low number of iterations. However, it is applied sequentially, due to the need to consider the results of the previous iteration, and therefore the optimization time might be longer.

Another common practice in ML applications is the so-called *fine tuning*. In fact, sometimes it can be convenient to re-use an existing model instead of training a new one from scratch. This is particularly useful when one wants to perform a task that is similar to the one for which the original model was trained. For instance, a model trained to recognize cars can be fine-tuned to recognize trucks. The fine-tuning process consists of using the parameters of the original model as starting values for training the new model.

2.2.1 Autoencoders

The deep-learning model chosen for the first part of this project is a neural network structure known as autoencoder (AE), traditionally used for dimensionality reduction and feature extraction [10]. This consists of two main parts: an encoder and a decoder, which usually have the same number of layers, with the number of nodes per layer in reverse order. Therefore, the majority of autoencoders present a symmetrical structure.

As its name suggests, the role of the encoder is to encode the input and produce a latent representation of it. In contrast, the output of the decoder is a reconstruction of the input obtained using only the latent representation. The goal of an autoencoder is to generate an output as equal as possible to the input.

It is crucial to impose constraints on the copying task, which act as forms of regularization[2], so to prevent the neural network from learning the identity function (i.e. copying the input into the latent representation and into the output) and to extract useful properties from the data. The most common solution is to impose that the size of the latent representation is smaller than the size of the input. In this case, the model is called undercomplete autoencoder and it is forced to learn only the most important features of the input during the encoding task. Another possibility is represented by denoising autoencoders [62]. In this approach, the model is trained to reconstruct the input given a noisy version of it. The encoder receives some noisy data as an input and the loss function compares the original uncorrupted data with the reconstruction generated by the decoder. By removing the noise from a corrupted input, the model is forced to extract the essential features that characterize the original data. A schematic representation of a denoising autoencoder workflow can be found in Fig. 2.5, which displays STEM data as an example. Given a dataset of clean original data (known as *infinite-dose images*, which, in the case of STEM data, refers to ideal images that are not affected by any type of noise), some kind of noise is added to it; these corrupted data are used as input of the autoencoder. The encoder generates a latent representation of them and the reconstruction of the decoder is then compared to the original uncorrupted data, by means of a loss function. After the training, aside from feature extraction, denoising autoencoders can be used to remove noise from data that are generated corrupt. In order to exploit this functionality, it is crucial to train the model with noisy data that are similar to the real corrupted data of the chosen context.

It should be noted that denoising autoencoder can be undercomplete or overcomplete. In the latter case, the dimension of the latent space is the same or larger than the dimension of the input. As mentioned before, having a latent space with a size smaller than the input, as in the case of undercomplete autoencoders, forces the model to learn a compressed representation of the data, which ideally retains only the most important features. This can be considered a form of regularization, which leads to better generalization. Moreover, in general, undercomplete autoencoders are less computationally expensive than the alternative, namely overcomplete autoencoders. However, the compressed representation may not

[2]Regularization is a tool used to improve the generalization performance of machine learning models. It aims at reducing the complexity of a model, making it less likely to fit noise in the training data and more likely to capture underlying patterns that generalize well to unseen data.

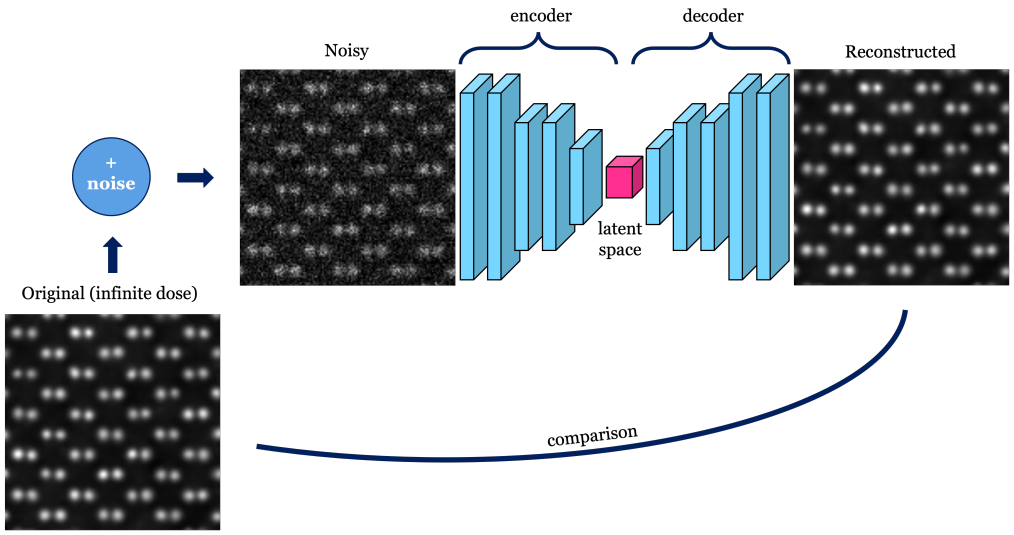

Figure 2.5 – Schematic representation of a denoising autoencoder workflow, applied to a dataset made of STEM data. Some noise is added to a clean image (named *Original, infinite dose*, in this case). The noisy image is fed to the encoder, which produces a latent representation of it. The decoder part of the model receives the latent representation as input and uses it to reconstruct the noise-free data. This is compared to the original image through a loss function.

capture all the essential information in the data. Overcomplete autoencoders have increased capacity from this point of view, being able to capture intricate data patterns and fine details. Nonetheless, some drawbacks are associated with this type of architecture, namely the higher demand for computational resources and the risk of low generalization ability. This last disadvantage can be prevented by including regularization techniques, such as adding noise. The choice between overcomplete and undercomplete autoencoders depends on the specific task and dataset.

In the case of autoencoders developed for image processing, the so-called *convolutional layers* are commonly employed to extract features from the data [63]. The output of convolutional layers is usually called *feature map*. In order to obtain these maps, a linear operation is performed between the input and an array of weights (known as *filter*), whose size is smaller than the input. A dot product is performed between a filter-size area of the input and the filter. This operation is repeated for each overlapping filter-size area of the input, moving the filter along the width and the height of the image, according to a defined stride. The resulting feature map summarizes some features detected in the input (i.e. edges, corners, objects). During the training process, the numerical values encoded within the filters are optimised. In fact, different values in the filters allow the identification of different features from the data. Convolutional layers require also the choice of an activation function, which determines the output of such layers. Two examples

of activation functions, which will be also used for this research project, are:

- *ReLu*, which stands for Rectified Linear Activation Unit and applies element-wise non-linearity. It is expressed by the function,

$$f(x) = max(0, x). \tag{2.8}$$

- *Sigmoid*, which requires more computational effort compared to *Relu*, but it is commonly used in the output layer of autoencoders applied to grayscale images because it guarantees an output in the range $[0, 1]$. For this reason, it can be interpreted as a grayscale image and makes the learning process more stable. It can be written as,

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{2.9}$$

Another argument that needs to be specified when working with a convolutional layer is the padding. Padding refers to the addition of pixels, usually of value zero, to the borders of the images. When padding is specified as *valid*, no additional pixels are added. In this case, some information on the border of the image will be lost and the output will be smaller than the input, depending on the filter size and the stride. In contrast, the padding value *same* implies the addition of as many pixels as required to have the output of the convolutional layer of the same size as the input. Moreover, it will be possible to extract more information from the border of the input. A powerful feature of convolutional layers is the possibility to apply multiple filters in parallel and therefore to learn various features at the same time. However, this will imply an increase of the time required to train the model, being the filters made of the weights that are learned during the training process.

The variation of the size that the input undergoes when passes through an autoencoder is achieved by using *MaxPooling* and *UpSampling* layers. Specifically, *MaxPooling* layers are used to downsample the input (the feature map generated by a convolutional layer) by taking the maximum value over a region smaller than the input that is shifted across the image, according to a certain stride. The role of this layer is to reduce the size of its input and to summarize the features observed in it. When passed through an *UpSampling* layer, the rows and columns of an input are repeated as many times as specified in the *size* argument of the layer. A common choice is to set the repeating factor to two for both directions.

2.2.2 Neural networks for video frame interpolation

As mentioned in the introductory chapter, video frame interpolation neural networks are widely developed and can also be used in applications not related to the

Figure 2.6 – Schematic representation of a video frame interpolation process, applied to a video of a dog running in a garden. The original and the augmented videos are presented as sequences of frames, along the time direction indicated by the blue arrow. The top row displays the original video, of only three frames. The video on the second row, obtained by applying the video frame interpolation algorithm RIFE, is made of five frames. The movements of the dog appear more visually fluid in the augmented video segment.

original purpose of their development. Video frame interpolation is a computer-vision, whose aim is to generate more visually fluid videos by creating additional frames between consecutive existing frames, namely by increasing the so-called *frame rate* or *frame per second* of videos [64]. This approach is widely used in the generation of slow motion videos [65] and video prediction [66].

The concept of adding frames between existing frames is depicted in Fig. 2.6, where two videos are displayed in terms of consecutive frames, ordered according to the time direction indicated by the blue arrow. The top row shows the original video segment, made of only three frames, of a dog running in a garden. The bottom row presents the video segment obtained after applying a video frame interpolation technique, namely RIFE [67], which will be introduced shortly. In this case, the video consists of five frames, which makes the dog's movements more visually fluid. A variety of deep-learning frameworks have been proposed to pursue this task. One of the state-of-the-art algorithms for video frame interpolation is RIFE (Real-Time Intermediate Flow Estimation), developed by Huang et alii [67]. Five main methodologies are usually employed for video-frame interpolation, namely flow-based methods, convolutional neural networks (CNN), phase-based approaches, GANs, and hybrid schemes. These typically differ from each other because of the network architecture and their mathematical foundation [68]. RIFE belongs to the flow-based category, whose focus is the determination of the nature of the flow between corresponding elements in consecutive frames, namely, the optical flow. When compared to other popular algorithms [69, 70, 71], RIFE performs better both in terms of accuracy and computational speed. Models belonging to the same class usually involve a two-phase process: the warping of input frames in accordance with the approximated optical flow, and the use of

CNNs to combine and refine the warped frames. The outcome of the intermediate flow estimation often requires the presence of additional components, such as depth-estimation [69] and flow-refinement models [70], so to mitigate potential inaccuracy. Unlike other methods, RIFE does not require supplementary networks, a feature that impacts significantly the model speed. In fact, the intermediate flow is learned end-to-end by a CNN. Specifically, RIFE adopts a neural-network architecture, IFNet, which directly estimates the intermediate flow adopting a coarse-to-fine approach with progressively increased resolution. This allows for capturing finer motion details and producing high-quality intermediate frames. Moreover, a privileged distillation scheme is introduced to train the model. This means that a teacher model, that has access to the ground truth (the intermediate frame), guides a student model during the learning process. The input of the RIFE model can either be a video or a sequence of two images. For this project, the most straightforward solution is to use images. Therefore, the model was adapted to accept a series of any number of images, instead of only two at a time. As with any large machine-learning model, RIFE updates regularly. At the current moment in time, the best version available is the HD model v4.6 referred to as RIFE HD [72]. This is trained on the Vimeo90K dataset, which covers a large variety of scenes and actions, involving people, animals, and objects [73]. Clearly, none of the datasets examined in Chapter 5 is included or is similar to the data of the Vimeo90K dataset. For this reason, fine-tuning of the available pre-trained model was performed for one of the proposed applications. Since fine-tuning of RIFE HD is currently not possible, the second-best model is here considered, namely RIFE_m [74]. More details about the training set used for the fine-tuning will be provided in Chapter 5.

Validation and limitations of video frame interpolation algorithms

Despite being extensively used in the context of video frame interpolations, these kinds of algorithms are associated with several challenges. Firstly, the outcome quality is degraded in the case of abrupt changes in lightning, fast-moving objects, occlusions, and other factors, which hinder the ability to predict the optical flow accurately.

Moreover, the models are trained on a limited amount of data and they need to be able to generalize to unseen types of motion. The resolution of the generated frame should be at the same level as the original frames, in order to ensure smooth transitions among them and to guarantee the extraction of precise information from the data, hampered in the case of outcomes with blurred features.

Another aspect that should be taken into account is computational efficiency, meaning the time and resources needed to achieve satisfactory results; near-real-time performance would be preferable.

Finally, objective evaluation of the results is often not achievable. In fact,

the majority of the algorithms rely on traditional full-reference computer vision metrics [75]. A full-reference metric is based on a direct comparison between the analysed output and the ideal outcome, namely the reference data. In the case of video frame interpolation applications, these metrics are evaluated on pairs of frames. Some examples are:

- MSE (Mean Squared Error), defined in Eq. (2.6), provides pixel-by-pixel comparison; for two identical images the MSE is zero;
- PSNR (Peak Signal to Noise Ratio) is a variation of MSE, defined as,

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX_I^2}{MSE}\right), \tag{2.10}$$

  where MAX_I is the maximum signal in the image and the log is used to express PSNR as a logarithmic quantity in the decibel scale; the PSNR approaches infinity as the MSE approaches zero;
- SSIM (Structural Similarity Index Method) aims at being more correlated with the human quality perception, by taking into account three factors: luminance, l, contrast c, and structure, s. The SSIM between two images or windows x and y is expressed as a weighted combination of these three measurements, with weights α, β and γ, according to the relation,

$$SSIM(x,y) = l(x,y)^\alpha \cdot c(x,y)^\beta \cdot s(x,y)^\gamma, \tag{2.11}$$

  with:

$$l(x,y) = \frac{(2\mu_x\mu_y + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)}, \tag{2.12}$$

$$c(x,y) = \frac{(2\sigma_x\sigma_y + c_2)}{(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{2.13}$$

$$s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}. \tag{2.14}$$

  Here μ_x (μ_y) is the pixel sample mean of x (y); σ_x^2 (σ_y^2) is the variance of x (y); σ_{xy} is the cross-correlation of x and y; c_1, c_2, and c_3 are values used to stabilize the division ($c_3 = c_2/2$). As their names suggest, the luminance factor compares the pixel intensity or brightness between the two images; the contrast term measures the loss in terms of contrast, and the structure element correlates with the spatial arrangement of pixels. A simplified version has been proposed [75], obtained by setting the weights $\alpha = \beta = \gamma = 1$:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \tag{2.15}$$

SSIM can assume decimal values between -1 and 1: 1 means perfect similarity (i.e. same image), 0 corresponds to no similarity, and -1 indicates complete dissimilarity or inversion between the two images, a rare circumstance in real-world scenarios, usually requiring deliberate manipulation to be achieved.

- MS SSIM (Multi-Scale Structural Similarity Index Method) is an extension of the SSIM. It considers image details at multiple scales and is therefore useful for capturing both local and global structural information in images. This makes the comparison more robust. To calculate this metric, the original data are downsampled to several scales, typically using a Gaussian pyramid[3], in order to obtain a set of images at different resolutions, both for the generated and the reference (ground truth) data. The traditional SSIM is then calculated separately for each pair of corresponding images. Finally, these values are combined together to obtain a single Multi-Scale SSIM score. For this combination, different weights are assigned to each SSIM value. Usually, these weights are inversely proportional to the scale, meaning that, in the final MS SSIM score, more importance is given to finer scales (higher resolution).

These metrics are not necessarily consistent with human visual perception [76] and they often fail in evaluating image quality attributes such as sharpness, noise, distortion, contrast, and artefacts (i.e. blurring, flickering, blocking, and others). Clearly, simple visual comparison, always presented for the assessment of video frame interpolation methods, cannot be exhaustive for an objective quantitative evaluation.

Some reduced-reference [77, 78] and no-reference [79, 80] metrics have been proposed. Reduced-reference metrics, used when the reference video is not fully available, assess the model by comparing features extracted from the two analyzed videos. In contrast, no-reference metrics do not require any reference. Due to the lack of exhaustive information from a reference video, both these categories of metrics provide inaccurate results, compared to full-reference metrics. Notably, some methods to assess the quality of video by using neural networks have been proposed [81, 82]. However, these practices are rarely employed in the video frame interpolation papers that can be found in the literature, and they are usually only applicable to the so-called *natural* video scenes, namely videos involving people, animals, and objects. Therefore, this solution must be excluded for the purpose of this research project, which does not involve such kinds of videos.

The approach proposed in this work is to identify some features related to the information that is usually extracted from the investigated dataset. For instance,

[3]A Gaussian pyramid is a multi-scale representation in which an image undergoes repeated Gaussian blurring and downsampling operations. The final result is a stack of images representing the original data at different levels of detail. This representation is widely used in various image processing and computer vision applications.

in the case of nanostructured graphene networks, we compare the target and reference outcome in terms of network porosity. This feature is regularly studied by experts in the field of nanostructured specimens to understand the material's properties. This approach allows one to confirm that the outcome of the proposed model preserves or improves the quality of the information characterizing the original dataset. In fact, for the applications investigated in this work, the focus is not on the image quality per se, but on the knowledge that we can retrieve from the data.

2.3 Medical imaging techniques

Medical imaging allows investigation of the inside of the human body and plays a fundamental role in diagnosis and treatment. It encompasses several modalities, a few of which will be covered in this chapter. Specifically, a description of the techniques used to extend the application of the video frame interpolation neural network RIFE is provided. Additionally, a brief overview of AI applications in healthcare is proposed, including a discussion of the limiting aspects of this field.

2.3.1 X-ray medical imaging

X-rays are electromagnetic waves with wavelengths ranging from 0.01 to 10 nm, which can be used to image the inside of the human body [83]. X-rays are produced by an X-ray tube, where a difference of potential is used to accelerate electrons toward a target material (typically Tungsten) and turn them into electromagnetic radiation.

When a beam of X-rays passes through the body, it undergoes different levels of absorption, depending on the atomic number and the density of the tissues encountered. The beam attenuation is determined by two main types of phenomena: Compton and photoelectric interactions. The first occurs when X-rays interact with the outer electrons in the body's soft tissues, such as muscles and organs. The consequence is energy loss and a change of direction for the scattered particles. Photoelectric interactions describe the process happening between the X-ray beam and the inner electrons of the invested tissue: the X-ray photon is absorbed and a photoelectron is released. This happens for body parts containing atoms with high atomic numbers, such as bones and radiocontrast agents, in addition to Compton interactions. Radiocontrast agents, usually iodine-based or barium-based, are substances that are sometimes administered to patients undergoing X-ray imaging procedures. By absorbing external X-rays, they improve the visibility of internal structures and, therefore, enhance diagnostic capabilities.

Measurements of the attenuation profile of the transmitted beam can be used

to form an image on a film or a digital detector. The pixels in the image will be brighter in areas corresponding to dense tissues (e.g. bones), while darker pixels will represent less dense tissues.

This imaging modality is mainly used to assess bone fractures, detect certain types of tumours, and examine lung conditions. Among the advantages of this technique, the speed and the relatively low cost of the procedure should be mentioned. The disadvantages include the limited performance in distinguishing soft tissues with similar density and the ionizing radiation exposure. This is a known risk factor for cancer development [19]. However, due to the speed and the number of procedures people require on average during their lives, this imaging modality can be considered safe in most cases. In other words, in the majority of the circumstances, the benefits of this technique largely outweigh the risks.

2.3.2 X-ray computed tomography

The X-ray imaging modality described in the previous section can also be used in a more advanced setup, which allows the generation of 3D reconstructions of internal parts of the human body, namely X-ray computed tomography (CT). The functioning of CT is based on a rotating X-ray beam that invests the body from different angles. The attenuation profile is measured for each position and processed by a computer. The outcome is a stack of cross-sections of the body, that can be used to produce a 3D representation.

In CT scans, the radiodensity of different materials is quantified according to the Hounsfield unit (HU) scale, which measures the level of X-ray absorption by these materials. The term radiodensity is commonly used in the medical context and refers to the level to which a material absorbs X-rays passing through it. A radiodense material absorbs more X-rays and will appear whiter in CT scans. A material with lower radiodensity interferes less with the passage of X-rays and will be represented by darker shades in the CT image. By applying a linear transformation to the X-ray attenuation coefficient measurement, the HU scale expresses the radiodensity of tissue relative to the radiodensity of distilled water, which is set to be 0 HU. Values in the HU scale can either be positive or negative. For instance, bones have a Hounsfield value of $+1,000$ HU, while air is around $-1,000$ HU.

Compared to standard X-ray imaging, CT provides better visualization, both in terms of contrast and resolution. With this imaging modality, it is possible to obtain a detailed representation of organs and soft tissues, whose visualization is limited in the case of standard X-rays. In fact, the algorithms used for image reconstruction can enhance the attenuated differences between tissues with similar densities. Applications of this imaging modality include the investigation of complex fractures, tumours, vascular diseases, and others. Notably, the relatively

small amount of time needed to acquire detailed information makes this imaging procedure particularly beneficial for decision-making in emergency contexts. However, CT involves higher costs and higher risks linked to increased radiation exposure, compared to standard X-ray and other imaging modalities, such as magnetic resonance imaging.

Dose reduction strategies have been and are currently investigated, with the goal of reducing the dose delivered to the patient by preserving the image quality [84]. First of all, unnecessary radiation can be reduced by tailoring the procedure needs to the patient's characteristics. The X-ray tube output can be varied during the procedure by modulating the tube current with the goal of reducing the dose for low-attenuation areas and increasing it for high-attenuation areas, such as bones.

Another possible solution is the so-called *Dual-Energy* CT [85], where two different X-ray energy levels are used, depending on the addressed material. The choice of the algorithms used to generate the 3D representation has an impact on the dose delivered to the patient: iterative methods [86] provide better reconstruction compared to traditional filtered back projection methods, allowing a reduction of the radiation dose needed for high-quality results [87]. Furthermore, noise reduction techniques can be used during the 3D reconstruction to improve the accuracy of areas acquired at lower doses.

It should be noted that this measurement technique is not limited to the medical environment, but it is also widely used in industrial settings and, in general, as a research tool across materials science [88, 89].

2.3.3 Fluoroscopy and coronary angiography

X-rays can also be used in medical practices like fluoroscopy [90] and coronary angiography [91], widely employed in cardiology to investigate heart and blood vessel conditions.

Fluoroscopy is an X-ray imaging technique that allows continuous visualization of moving body structures. It involves using an X-ray machine with a fluoroscope, which is a specialized X-ray detector that can display images in real time on a monitor. During this medical practice, a continuous X-ray beam is passed through the patient's body, and the transmitted X-rays are detected by the fluoroscope. The detector sends the X-ray images to a monitor, where they appear as moving frames. This real-time imaging technique enables physicians to observe the movements of organs, blood flow, and the progression of medical procedures.

In particular, fluoroscopy is employed during a procedure known as coronary angiography [91], used to assess the cardiovascular system. Specifically, a contrast dye is injected into the coronary arteries through a thin flexible tube, namely a catheter, usually inserted in the groin or wrist. X-ray imaging is needed during

the procedure to guide the wire inside the patient's body and assess how the blood flows. In fact, the dye is visible on X-ray images and outlines the blood vessels, allowing the medical practitioner to identify possible blockages, which are evidence of coronary artery diseases [92].

Contrary to the other medical operations described in this chapter, coronary angiography is an invasive procedure. Risks and complications associated with it are both patient- and procedure-related. Some examples are bruising, allergic reactions to the contrast dye, hematoma, but also artery damage and heart attack, in some extreme cases. However, it is considered a safe procedure in the majority of circumstances [93]. Furthermore, in many cases, it allows one to avoid surgery, which involves higher invasiveness and costs [94].

It should be noted that, for fluoroscopy-assisted protocols, there are some risks also for the physician performing the procedure, mainly related to X-ray exposure [95]. In fact, contrary to the case of CT scans, the practitioner needs to be in the same room as the patient, to guide the wire inside their body. Modern equipment can guarantee a reduction of radiation exposure. However, some radiation risks cannot be controlled. Prolonged and repeated exposure over time poses some serious risks for practitioners. According to current regulations [96], the dose limit for medical professionals is 20 mSv per year averaged over 5 consecutive years, with a maximum of 50 mSv for each year. This value is larger than the one for the general public, which is 1 mSv per year. However, it still limits the number of procedures that professionals can perform during their lifetime.

One of the factors affecting the amount of radiation dose released to the patient and the operators is the frame rate. This indicates how often X-ray images are collected and displayed on the monitor during the procedure. As in the case of any other type of video, a high frame rate provides a more continuous transition between frames. In the context of coronary angiography, this translates to a smoother visualization of the movement of the contrast dye through the blood vessels, which can ultimately facilitate the identification of cardiovascular irregularities. However, higher frame rates also imply higher X-ray exposure. Therefore, a trade-off should be identified between radiation release and diagnostic information. The most commonly employed frame rates are 30 fps (i.e. frames per second), 15 fps, and 7.5 fps, chosen depending on the specific procedure's requirements and the patient's condition. Notably, the most advanced X-ray technologies allow one to adjust and optimize the frame rate during the medical procedure [97].

2.3.4 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is a non-invasive imaging technique, used for diagnostic purposes [98]. Contrary to X-ray imaging and related practices,

in this case, no radiation risk is involved. In fact, MRI is based on the use of strong magnetic fields and low-energy radio frequency signals, interacting with the hydrogen atoms in the human body. Precisely, the patient is placed in a strong static magnetic field, between 1-3 T, generated by superconducting magnets. In this condition, the spin of the protons in the body tissues aligns with the magnetic field, parallel or antiparallel to it, leading to a magnetization vector with the same direction as the external magnetic field. Precisely, the spin of the atoms shows a precession behaviour (namely a change of the direction of the rotation axis) around the direction of the magnetic field. The frequency of this precession is known as Larmor frequency. By applying energy to the precessing spins it is possible to alter the net magnetization vector and, consequently, retrieve information from the spins. The energy is applied in the form of radiofrequency, at the Larmor frequency, and alters the orientation of the spins. When radiofrequency is switched off, the spins return to their original position, emitting back the radio waves. A computer receives these signals and converts them into an image.

The realignment time and amount of released energy depend on the environment and the characteristics of the tissues, namely thickness, hardness, and amount of water molecules. Therefore, these features allow one to distinguish the location and shape of different body components.

Compared to CT, MRI achieves better contrast, especially in imaging soft tissues. It is therefore widely used to investigate anomalies in the human brain, tumours, and several other diseases [98]. Sometimes, contrast agents are used to enhance the contrast among body features. However, the MRI acquisition time is quite long and therefore this procedure is not suited for emergency diagnosis. The long time required, together with the loud acoustic noise created by this technology, makes this medical exam uncomfortable for some patients. Notably, modern MRI instruments have made significant progress in improving patients' well-being during the procedure [99].

2.3.5 AI for medical applications

An increasing number of studies involve the application of AI to healthcare. The reason behind this technological advancement is linked to the ability of neural networks to process vast amounts of data and identify complex patterns among them, which might be too subtle for the human eye. As a consequence, decision-making processes can be accelerated and human errors can be reduced.

It has been demonstrated that neural networks can support and assist medical practitioners in performing several tasks and ultimately enhance diagnostic and treatment capabilities. Some examples include drug discovery [100], disease detection [101], hospital operations and management [102], remote monitoring of patients [103], robot-assisted surgery [104], and many others. A large and

widespread branch of this field involves the use of deep-learning models for medical imaging, including both data generation and data analysis. Several models have been developed to perform tasks such as segmentation, feature extraction, classification, and visualization, just to name a few [105].

In the case of CT data, neural networks can be, for instance, used to facilitate the 3D reconstruction process, especially when a limited number of views is available. This is achieved either with an end-to-end [106] or help-to-end [107] modality. The term end-to-end refers to models that go directly from the sinogram space (i.e. the complete set of attenuation profiles) to the image space. Help-to-end solutions are developed to assist traditional reconstruction processes by using AI to define optimal parameters and post-process the data. Neural networks have also been developed to overcome one of the main challenges of MRI data generation, namely the long acquisition time [108]. Interestingly, some deep-learning models have been developed to generate synthetic MRI and CT data, which can be for instance used as a data augmentation tool [109], but also to convert MRI data into CT scans. This is particularly useful when the patient's radiation exposure needs to be limited, such as in the case of pregnancy [110].

The development of AI-powered medical tools has led to high expectations but has also raised concerns from the medical community, the regulatory bodies, and the general public. These issues include data security concerns, ethical debates, and the need for scrupulous assessment. Recent studies [111, 112] demonstrated some scepticism among clinicians and patients towards the use of AI in healthcare. The lack of trust is often linked to difficulties in understanding how algorithms perform the tasks for which they are built and what are the motivations behind their inference. From an ethical perspective, some controversy can arise when, for example, a treatment suggested by an AI model is not beneficial for the patient. Current regulations are not able yet to provide guidance on the responsibility of the algorithms, in these circumstances. Regulatory challenges also involve considerations about patient safety and data privacy. In fact, since such algorithms require access to extensive databases, it is crucial to guarantee secure storage of sensitive information and avoid possible patient re-identification.

In order to overcome these concerns, it is essential to improve the communications between practitioners and AI experts and to provide a rigorous assessment of the results. Two main types of validations are possible for AI-assisted medical devices (including software): internal and external validation. The former is performed on a subset of the dataset used for training the model, which can be selected randomly or according to some criteria. External validation, which is preferable when possible, consists of assessing the model's capabilities through a dataset totally unrelated to the training set. For instance, a dataset provided by a different hospital using an imaging instrument from a different manufacturer

would be an ideal solution.

In the case of binary classification models, used for example to diagnose diseases, it is relatively easier to identify quantitative metrics to assess the results. In fact, in the majority of the cases, the ground truth is provided by human-generated labels or can be obtained with additional medical tests. Some examples of quantitative metrics include [113]:

- Accuracy, namely the percentage of correctly assigned labels, expressed as,

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \tag{2.16}$$

  where TP (TN) is the number of true positives (negatives) and FP (FN) is the number of false positives (negatives)[4].

- Precision, which expresses the percentage of actual positive cases out of all the cases that are predicted as positive, calculated according to the formula,

$$Precision = \frac{TP}{TP + FP}. \tag{2.17}$$

- Sensitivity or *true positive rate*, which is the ability of a model to correctly identify a patient with a disease, and it is expressed as,

$$Sensitivity = \frac{TP}{TP + FN}. \tag{2.18}$$

- Specificity or *true negative rate*, which is the ability of a test to properly identify a patient without a disease, expressed as,

$$Specificity = \frac{TN}{TN + FP}. \tag{2.19}$$

- ROC (Receiver Operating Characteristic), this metric provides a graphical representation of a model's performance, depending on the discrimination threshold used for the classification task. In fact, when performing classification, the output of the model is a continuous value of probability. A threshold must be selected to assign predictions to different classes (two in the case of binary classification). The ROC metric displays the true positive

[4]The term true positives (TP) refers to the cases where the model correctly predicts the positive class, which could be the presence of a disease, in the case of medical applications. True negatives (TN) are the cases where the model correctly assigns the negative class (e.g. g. absence of disease). The term false positives (FP) identifies the cases where the model assigns a positive label and the correct prediction would be a negative label. Contrarily, the cases where the model predicts a negative class when the correct answer would be positive, are referred to as false negatives (FN).

rate (Sensitivity) on the x-axis and the false negative rate ($1-$ Specificity) on the y-axis, across various threshold values.

- AUC (Area Under the ROC curve), this is a scalar value in the $0-1$ range, used to quantify the overall performance of a classification model. It is obtained by integrating the ROC curve. The value AUC $= 1$ identifies a perfect classifier, while a random classifier has AUC $= 0.5$.

The assessment of a deep-learning model can be achieved easily also in the case of segmentation tasks, where the model-inferred segmentation can be compared to manual segmentation provided by experts or obtained with different techniques. In these circumstances, the Dice similarity coefficient [114] can be used to compare segmentation results, also known as *masks*. It ranges from 0 to 1, where 1 indicates perfect overlap and can be computed according to the formula,

$$Dice = \frac{2 \times \text{Area of Overlap}}{\text{Total number of pixels in both masks}}. \tag{2.20}$$

It is evident that the metrics described so far cannot be used when working with AI-assisted tools, whose goal is to improve image quality. In this context, it is more challenging to define valuable metrics to validate the quality of synthetically-produced images. Most of the time, physical parameters are not available to objectively compare images and uniquely state, which one provides more relevant information. When ground-truth data is available, which is not always the case, commonly used objective metrics comprise the computer vision analysis measures described in the previous section, such as MSE, PSNR, and SSIM. Additionally, other metrics have been introduced with the advent of generative models [115], aiming at assessing the quality and realism of artificially-produced images. A commonly used metric for this purpose is the Fréchet Inception Distance (FID) [116], which measures the similarity between the distribution of real images and the distribution of generated images, in a feature space. As the name suggests, this is a combination of the Fréchet distance [117] and Google's inception model [118]. The latter is a pre-trained convolutional neural network used to extract features from the images. This latent space retains properties of images at different scales and locations in the data. The mean and covariance matrix of the features in each image are then computed. Finally, the Fréchet distance is calculated, to compare the difference between each image's mean and covariance matrixes. Obtaining a lower FID score indicates that the generated images are more similar to the real images in terms of feature distribution, suggesting higher image quality and realism, with FID = 0 implying two identical images. Another commonly used metric in the field of generative models is the Wasserstein Distance Score (WDS) [119]. This is based on calculating the Wasserstein distance between the pixel values of real and

generated images. It quantifies the *cost* required to transform one distribution into another. Lower WDS values suggest that the distributions are more similar.

One of the limitations of these metrics is that they can only be used when ground-truth data is available. When this requirement is not fulfilled, it is even more challenging to evaluate the generated medical images. In some contexts, the noise power spectrum [120] can be calculated to compare different datasets (namely the original and the AI-improved version of it). Contrary to pixel standard deviation, the noise power spectrum describes both the magnitude and spatial frequency characteristics of the noise in an image, which affects the visibility of structures. This measurement is generally used for the assessment of the image quality of CT scanners and other imaging instruments, and it is evaluated over uniform regions of interest in phantoms [121]. In the medical context, a phantom is a specialized object that can be used for different purposes, including medical equipment testing. Phantoms are fabricated with the intent of mimicking characteristics of the human body or specific body tissues. They allow for standardized testing and comparisons across different medical facilities and equipment. They are simplified models of the human body features, which cannot replicate exactly the clinical data. Therefore, the applicability of the same metric to clinical data is limited to small uniform regions of the examined image.

It should be noted that all the described metrics present some limitations. Importantly, the results of these analyses might not be consistent with human perception. Furthermore, in general, an image quality improvement may not necessarily imply a positive impact on the diagnostic capabilities.

This issue can be partially overcome by involving medical experts in the assessment process. These professionals can provide an evaluation of medical imaging data through surveys and subjective scores. This process is crucial to quantify the perceived image enhancement and identify inconsistencies in the algorithms' outcomes. Moreover, the involvement of practitioners is also important to ensure that the AI-driven results are clinically relevant. A commonly used strategy is to include both quantitative metrics and experts' opinions, when possible.

Importantly, different characteristics may be required for certain imaging data, depending on the specific application case. Therefore, extensive analysis should be developed depending on the particular use.

In conclusion, the integration of AI models into healthcare can enhance today's medical imaging capabilities and improve diagnostic accuracy, in a variety of applications. However, some challenges must be taken into consideration, including ethical concerns, data privacy regulations, and the need for scrupulous validation.

CHAPTER 3

DENOISING OF LOW-DOSE STEM DATA

This chapter presents and discusses the results of applying an in-house-developed deep-learning model to low-dose STEM data to enhance their resolvability and extract useful information from them. The work has been published in [122]. The first author and main contributor of this paper is the same as this PhD thesis.

3.1 Problem description and state-of-the-art methods

At present, aberration-corrected scanning transmission electron microscopes (STEM) provide the highest resolution of all imaging instruments, below 0.1 nm, and allow to investigate the structure and chemical composition of materials at the atomic scale [2]. However, a strong electron beam is needed to achieve atomic resolution and maximize the signal-to-noise ratio. This requirement often implies sample damage and alteration of the observation, with consequent limitations on the applicability of high-resolution electron microscopy. A reduction of sample damage could be achieved by decreasing the electron dose, but this would imply difficulties in extracting useful information from the data, due to the presence of Poisson noise, which increases when the electron dose is reduced [37]. More details about noise affecting STEM images can be found in Chapter 2.

Notably, noise removal is a common practice in image processing. However, most of the denoising techniques provide accurate results only in the case of additive noise, such as Gaussian. On the contrary, Poisson noise is signal-dependent and requires more advanced techniques [123]. One of the most commonly used

strategies for noise removal in electron microscopy is the application of a smoothing Gaussian filter [124, 125]. However, as it will be shown in this work, such a method can sometimes lead to results that are less precise than the original noisy version of the image. A more sophisticated technique is known as *block matching and 3D filtering* [126]. In this case, images are decomposed into fragments, which are grouped by similarity, and then the fragments are passed through filters. Unfortunately, such a denoising scheme assumes that the images to process correspond to a 2D periodic structure, an assumption potentially leading to artefacts, such as the inability to identify genuine vacant sites. Furthermore, this technique is usually more suitable for the removal of additive noise, such as Gaussian noise, and, when adapted to Poisson-noised data, the computation time increases significantly.

An alternative to these methods is provided by deep-learning techniques, which are becoming increasingly popular in the microscopy field across several applications [16]. The majority of the available denoising algorithms involve Gaussian noise only, so that they are useful exclusively for analog data acquisition. For instance, the state-of-the-art neural networks for Poisson-noise removal have been proposed [127]. These provide significant noise reduction, but their performance rapidly degrades at low doses. The reason behind such low-dose accuracy loss can be identified in the nature of the training set, made of simulated STEM images obtained by using a simple linear imaging model. In fact, in order to have a more realistic dataset, it is advisable to use simulation techniques that implement the multislice algorithm [41] or the Bloch-wave method [29], described in Chapter 2. These techniques incorporate information about the specimen and the instrument setting, and therefore achieve a more realistic simulation of an actual measurement. Furthermore, such state-of-the-art technique requires inputs from the users, who should decide whether or not to apply some level of up-sampling/down-sampling before processing the image. This condition hinders the feasibility of real-time denoising.

Other methods have been proposed before [128] and after [129] the publication of our paper [122] upon which this chapter is based. However, to the best of our knowledge, our method is the only one that fulfils the following requirements: 1) the denoising algorithm does not require any human pre- and post-processing of the images, allowing the model to be employed during live data acquisition; 2) our scheme removes the noise from digital images and excludes any types of noise that can be corrected at the instrumentation level, which is not required for state-of-the-art image acquisition instrumentation; 3) we propose a fully quantitative approach to evaluate the results of any denoising scheme; 4) the result of our model preserves the intensity information retained in the pixel value of the digital images. With respect to this last claim, it is worth mentioning that the intensity values in dark-field STEM images are indicative of critical information about the

quantity of atoms of each atomic column, as well as their composition. Future studies will involve an analysis of the scattering cross sections [130] to quantify the model's performance in preserving such physical information.

As mentioned in the previous chapter, only Poisson noise was considered for this project, since the main focus is the improvement of images acquired with the state-of-the-art digital acquisition (i.e. electron counting procedure, more details in Section 2.1.1 and [39]). Strategies to eliminate other types of distortions are not investigated within this project.

3.2 Model construction and training

The deep-learning model chosen for this denoising project is an autoencoder, whose general characteristics are described in Chapter 2. The model was built using the Keras library [131], which runs on the open-source machine-learning platform Tensorflow [132]. The architecture, schematized in Fig. 3.1, consists of ten layers in total: five for the encoder and five for the decoder. Images of size (n, n, 1) are provided to the input layer, where n represents the number of pixels in the two spatial directions, which can be any size, while the last value identifies the number of channels of the image. This is 1 in the case of a grayscale input, as requested by this model (it would be 3 in the case of RGB images). The input image then goes through a sequence of two blocks made of one Convolutional layer and one Pooling layer. Each Convolutional layer consists of 32 filters of size 3×3, moved with a stride of 1 pixel. The padding chosen for this model is *same*, with value 0. Finally, the activation function used to apply element-wise non-linearity is *ReLu*. Regarding the Pooling layers, used for dimensionality reduction and to summarize the feature map generated by the Convolutional layers, MaxPooling was selected, with size (2, 2) and a stride of 2 pixels. The outcome of these steps is the encoded version of the input, also known as *latent representation*. For an input with size (128, 128, 1), the latent space representation size is (32, 32, 32). Subsequently, the data is processed by the decoding part of the model, made of two blocks consisting of one Convolutional layer and one UpSampling layer, which presents a repeating factor of 2 for both directions and is used to increase the data dimension. The hyperparameters of the Convolutional layers in the decoder are the same as the encoder part, with the only exception of the last layer, which has only one filter, in order to ensure that the output size coincides with the input size.

The definition of the loss function is particularly significant in constructing the autoencoder. The commonly used Mean Squared Error (MSE), described in Chapter 2, is not suitable for training the proposed dataset. In fact, in the standard MSE equal importance is given to both black (low intensity) and non-black (high intensity) pixels, even if the interest sits mainly with the non-black pixels, which

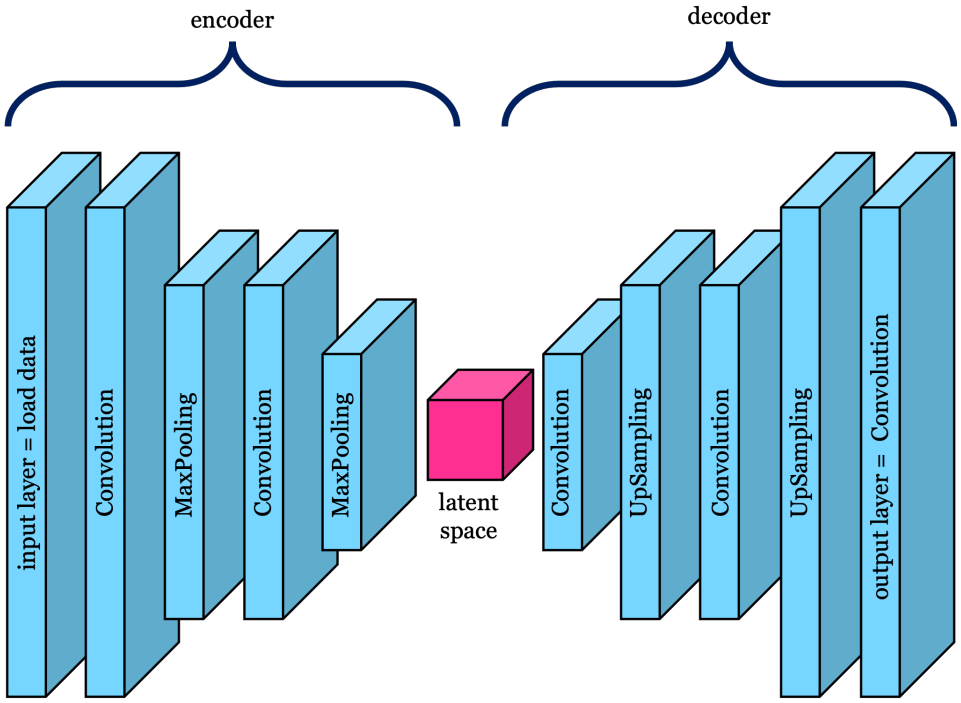

Figure 3.1 – Schematic of the proposed autoencoder, made of ten layers. The input data goes through an encoder, whose output is the so-called *latent space*. This representation then is decoded by the decoding part of the model.

indicate the presence of the atoms (here we consider dark-field images). Note that the importance of black and white pixels is reversed when the image is taken in the bright field. It should also be noted that Poisson noise varies according to the pixel intensity. Therefore, when a pixel is black no noise is detected and the corrupted and uncorrupted images are identical. For this reason, we employ a customized loss function, which gives more importance to the pixels that are not black in the original images contained in the training set. This loss function can be described as a weighted MSE (WMSE), with a 1000:1 weight ratio for non-black pixels. It can be expressed as

$$\text{WMSE} = \frac{1}{n}\sum_{i=1}^{n} w_i(\widetilde{y}_i - y_i)^2 \tag{3.1}$$

$$w_i = \begin{cases} 1 & \text{if } i = min(y); \\ 1000 & \text{otherwise.} \end{cases}$$

where n is the number of pixels in each image, w_i is the weight associated to the i-th pixel, $\widetilde{y}_i$ the predicted value and y_i the true value.

The model architecture was optimized by performing hyperparameter tuning over a validation set, with the random search approach (more details can be found in Section 2.2). A list of all the hyperparameters of the model and the range of

Table 3.1 – List of all the hyperparameters defining our model together with the corresponding range of values investigated.

Hyperparameter	*Investigated values*
Number of layers	6, 10, 14
Number of filters in the Conv2D layers	8, 16, 32, 64
Loss function weight	10, 100, 1000, 10000
Optimizer	Adam, Adadelta
Batch size	16, 32, 64

values investigated can be found in Table 3.1. It should be noted that, in the table, the number of layers include also the MaxPooling and UpSampling layers, not only the convolutional ones. Depending on the number of layers and the number of filters in each convolutional layer, the tested models present a different number of trainable parameters. The number of parameters for each convolutional layer can be computed as $(input\ channels \times filter\ size \times output\ channels) + output\ channels$. For instance, in the case of the first convolutional layer of the selected model, the input channel size is 1, the filter size is 3×3, the output channel size is 32, and therefore the number of trainable parameters is 320. In total, the chosen model presents 28353 trainable parameters. The investigated parameters space goes from a minimum of 737 parameters to a maximum of 185857 parameters.

The model has been trained on Quadro RTX 8000 GPUs, for 500 epochs, with a batch size of 64 and the Adam optimizer, for a total time of 2 h and 30 min. GPUs, provided by Nvidia, allowed us to significantly speed up the training time. In fact, the time required to train one epoch over the Quadro RTX 8000 GPUs is about 15 s, while with a Tesla K40c GPU approximately 120 s are needed. The training of one epoch of the same dataset on a CPU would require 717 s.

3.3 Training set

The autoencoder is trained on a dataset made of about 27,000 simulated images, all generated by using the Prismatic software [30, 31]. As described in Chapter 2, the use of this software ensures a faithful simulation of actual measurements.

Different materials have been considered in the creation of the training set, namely, graphene, graphite, GaAs, InAs, MoS_2, $SrTiO_3$, and Si, generated across various imaging conditions. Several parameters need to be defined to perform simulations using the Prismatic software, whose details can be found in the related documentation [30, 31]. The values selected for the generation of this training set are:

- the pixel size varies from 0.08 to 0.3 Å;
- the size of the simulated potential varies from 0.04 to 0.16 Å;

- the sample thickness range is $250 - 630$ Å;
- the electron beam energy range is $80 - 200$ keV;
- the maximum probe angle range is $25 - 32$ mrad;
- the inner and outer detector collection angles are 30 and 60 mrad, 30 and 70 mrad, 50 and 180 mrad, 60 and 180 mrad, 65 and 180 mrad, 75 and 180 mrad, 80 and 180 mrad;
- the potential bound is 2 Å;
- the probe tilt is 0 mrad in both horizontal and vertical directions;
- the probe defocus is 0 mrad;
- the number of frozen phonon configurations to calculate is one;
- the thermal effects are included in the simulation;
- the occupancy values for likelihood of atoms existing at each site are included.

These intervals and values correspond to realistic imaging and materials parameters, commonly encountered in actual STEM experiments, and were defined following the recommendation of microscopists. For each simulation, values for the specified parameters were randomly selected from a uniform distribution within the predefined ranges. For each material, 10 to 12 images were generated, including pristine and defective structures incorporating vacancies. In order to avoid reconstruction bias (the autoencoder learning the periodicity of the lattice), each simulated image has been rotated at two random angles to increase the variety of the dataset. Undoubtedly, the choice of training data poses limitations on the model's generalizability. Images with some features non included in the training set are investigated within this chapter, such as samples of different materials, sample placed on amorphous substrate, and samples affected by additional kind of noise. A complete analysis of the model's applicability will be conducted in future studies.

The choice of the dose values included in the training set highly affects the performance of the model and the ability to denoise images taken over a wide range of doses. As such, one wishes to keep that range as wide as possible. However, both the time required for training and the computational costs also scale with the size training set and the diversity of the data, so that a compromise between data variety and computational effort must be found. For this reason, we have selected a dose range going from 500 $e^-/\text{Å}^2$ to 10,000 $e^-/\text{Å}^2$. In the construction of the dataset, the dose value for each image is randomly selected within the defined range, so that there is no dose bias across the various materials.

It is worth mentioning that a different dose distribution across the dataset has also been tested. In this case, the number of images N_i at a certain dose value, ρ, was proportional to the selected dose value, according to the relation:

$$N_i(\rho) \propto \frac{1}{\sqrt{\rho}}\,, \tag{3.2}$$

which is the same relation that occurs between Poisson noise and dose [see Eq. (2.3)]. However, this solution did not improve the model's performance. Another option for the dose distribution in the training set consists of generating different datasets for the different dose levels and using them to train separate models. However, this would imply the requirement for the user to choose which model to use depending on the dose level of the test data, a quantity that is often unknown or not precisely defined in the case of an experimental acquisition. This goes against the desire to make the tool feasible for real-time application and requires prior knowledge. Moreover, in the test performed, we experienced a deterioration in the model reconstruction performance, and therefore this option was discarded.

All images have been cropped into 128×128-pixel plots, which is the format used to train the autoencoder. Note that, common real images are usually larger (at least 512×512 pixel), but dealing with reduced size allows one to increase the computational efficiency and to identify more details in the reconstructed data. It should be noted that the model accepts images of any size, however, it was trained only on 128×128-pixel plots to simplify the dataset construction. The result is a collection made of about 30,000 images: 10 % was used as validation set and 90 % as training set, as commonly done in machine-learning projects.

The pixel intensity of the images generated by Prismatic corresponds to the fractional intensity of the entire electron beam that is scattered to the specified STEM detector at a given pixel. This takes values comprised between 0 and 1, and usually corresponds to around $10 - 15$ % of the entire beam intensity. Measuring the signal in fractional units does not allow us to directly retain information on the actual physical electron dose used. As such, before applying the Poisson noise, it is necessary to convert the pixel values into integers, representing the physical number of electrons at a given pixel. This can be obtained by multiplying the original pixel value by the total electron dose (in units of electrons per Å^2) and then by multiplying by the pixel size (in Å^2). Such conversion is important to generate a training set with an intensity distribution compatible with that of typical experimental data. As a consequence, any type of pre-processing of the test data can be avoided. This is a crucial condition enabling one to use our machine-learning tool during the actual microscope real-time data acquisition. It is also important to remark that, in doing so, the pixels value, namely the outcome of our autoencoder, will describe a directly measurable quantity with proper physical

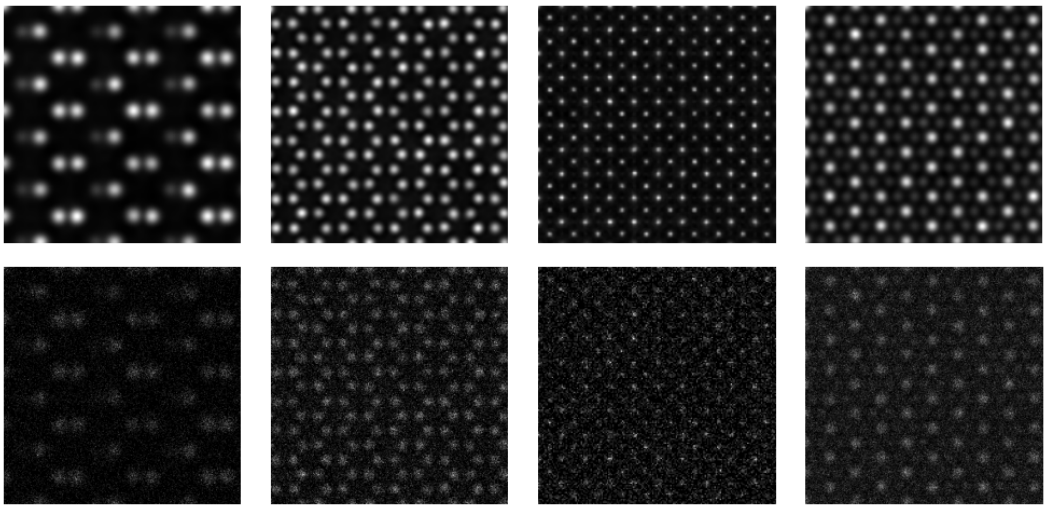

Figure 3.2 – Examples of elements of the training set. The first raw displays the original noise-free data and the second raw displays the Poisson-noised data. It should be noted that different levels of noise are applied for each one of the original images. Moreover, the images are rotated are random angles and cropped to 128×128-pixel plots before training. The materials displayed in this illustration, from left to right, are: Si, graphene, $SrTiO_3$, MoS_2.

meaning. This is not common practice in the case of analog-acquired data, where the pixel value cannot be directly associated with an observable, and the images are usually scaled between 0 and 1 before training and testing.

Some examples of the images of the training set are displayed in Fig. 3.2, where the original and Poisson-noised data are presented in the first and second raw, respectively. It is important to mention that these are just some examples of the considered levels of noise (multiple levels of noise were considered for each original image) and that these data still need to undergo the rotation and cropping processing.

3.4 Qualitative assessment of the results

The most straightforward qualitative evaluation consists of a visual comparison between the reconstructed and the corrupted images. The improvement brought by the denoising process is easily recognizable even by researchers, who are not experts in electron microscopy. A visual example is shown in Fig. 3.3, which displays the reconstruction of the digital experimental images of a Gold nanoparticle deposited on an amorphous Carbon substrate, obtained at different dose levels (note that Gold is not included in the training set). The data, acquired by Tiarnan Mullarkey and Clive Downing at the CRANN Advanced Microscopy Laboratory (AML www.tcd.ie/crann/aml/), Trinity College Dublin, are provided in the form

of twenty different frames of the same sample region; by rigidly aligning and summing up the signals of an increasing number of frames one can obtain multiple images at different doses. It is worth mentioning that, while the single acquisition is affected by Poisson noise only, additional types of noise characterise the images obtained by summing the individual frames. These can be due to a combination of factors, including particle rotation during the image series, which are not investigated in this work. In the experimental acquisition, the dwell time was 2 μs for each frame and the beam current was approximately 5 pA. This means that the dose of a single frame is 62 e^-/pixel. The dose of each image is then 62 e^-/pixel multiplied by the number of frames used for the image. For the sake of brevity, only three of the twenty consecutive sums are presented here. In particular, 128×128-pixel portions of the images acquired at 62 e^-/pixel, 372 e^-/pixel, and 744 e^-/pixel are shown in Fig. 3.3. The dose values expressed as $e^-/\text{Å}^2$ are 968 $e^-/\text{Å}^2$, 5808 $e^-/\text{Å}^2$, and 11616 $e^-/\text{Å}^2$. The top row displays the noisy images, the second one corresponds to the reconstructions obtained after the application of the autoencoder, the third row contains the difference between the noisy and the reconstructed images (referred to as *Residual*), and the bottom row shows the Fast Fourier Transform (FFT) of the Residual. Although Gold is not classified as a beam-sensitive material, the example provided is effective, since it tracks the model performance across different dose levels, a task that remains challenging when dealing with experimental data. The reconstructed data always display a significant quality improvement over the original noisy images. In fact, from the reconstructions, one can immediately recognize the five crystallites forming the Gold nanoparticle, regardless of the noise level of the original image. The most notable difference between the three reconstructions is in the shape of the individual atoms, which appear progressively more round as the dose increases. Note that our algorithm is not trained to necessarily return round atoms, but only to denoise the signal. This is why at a very low dose, as in the case of 62 e^-/pixel, there is significant atomic distortion. It is also worth noting that no further progress is found when adding additional frames to the 744 e^-/pixel case (namely when increasing the dose). This result allows one to conclude that, upon autoencoder reconstruction, an increase in the dose is not needed to obtain a satisfactory reconstruction. As a consequence, the beam damage to the sample can be reduced.

The two bottom rows of Fig. 3.3 show some periodicity in the removed noise. This is an expected feature of the Poisson noise, which scales with the pixel intensity, and does not imply a loss in crystallographic information. Such periodicity in the residual appears to be more evident for high-dose images, a fact that simply demonstrates that the autoencoder can remove the noise at high doses more efficiently than at low doses. It is worth mentioning that crystal structure

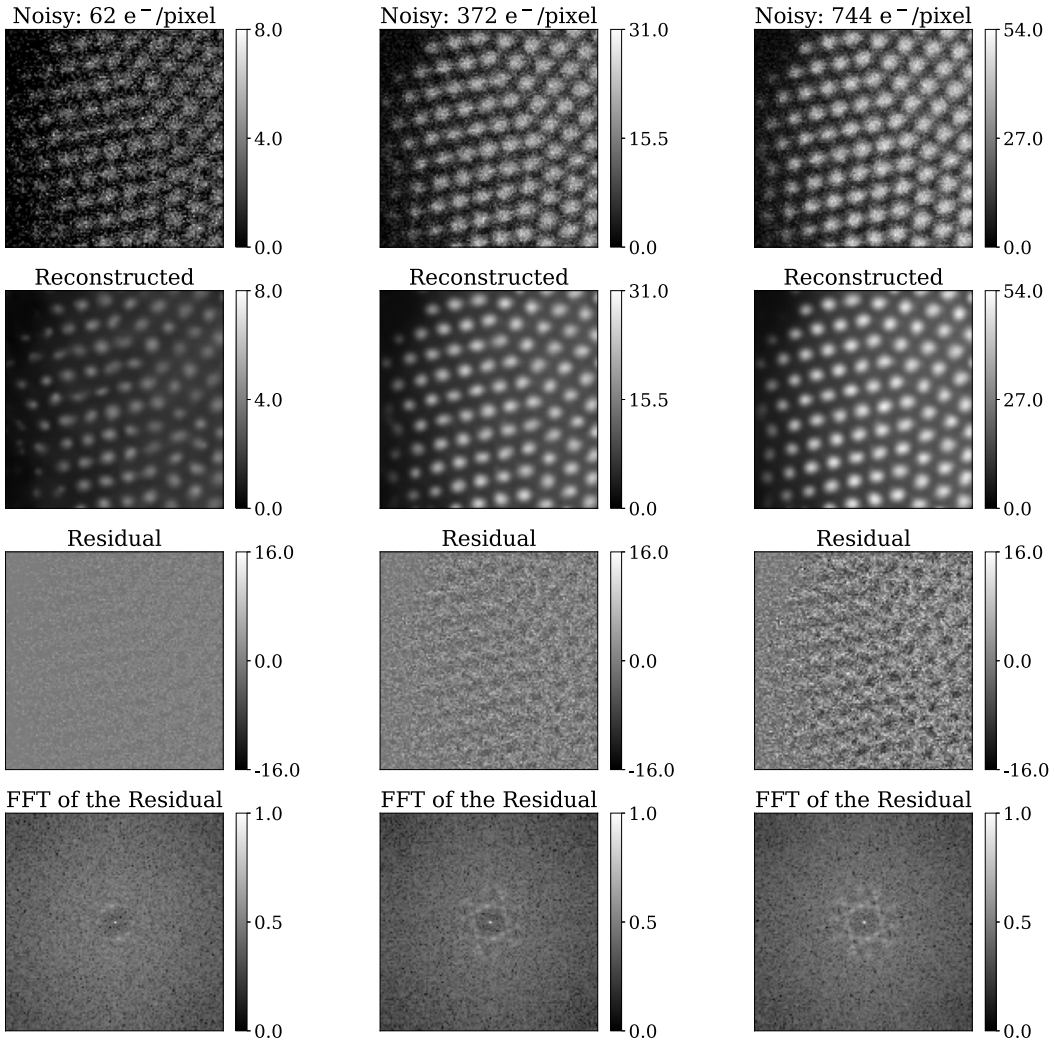

Figure 3.3 – Image reconstruction of a Gold nanoparticle deposited on an amorphous Carbon substrate, and imaged at different dose levels. Top row: the original noisy images; second row: the reconstructions obtained after the application of the autoencoder; third row: the difference between the noisy and the reconstructed images (called *Residual*); bottom row: Fast Fourier Transform (FFT) of the Residual. The three columns correspond to different dose levels, respectively from left to right 62 e^-/pixel, 372 e^-/pixel, and 744 e^-/pixel.

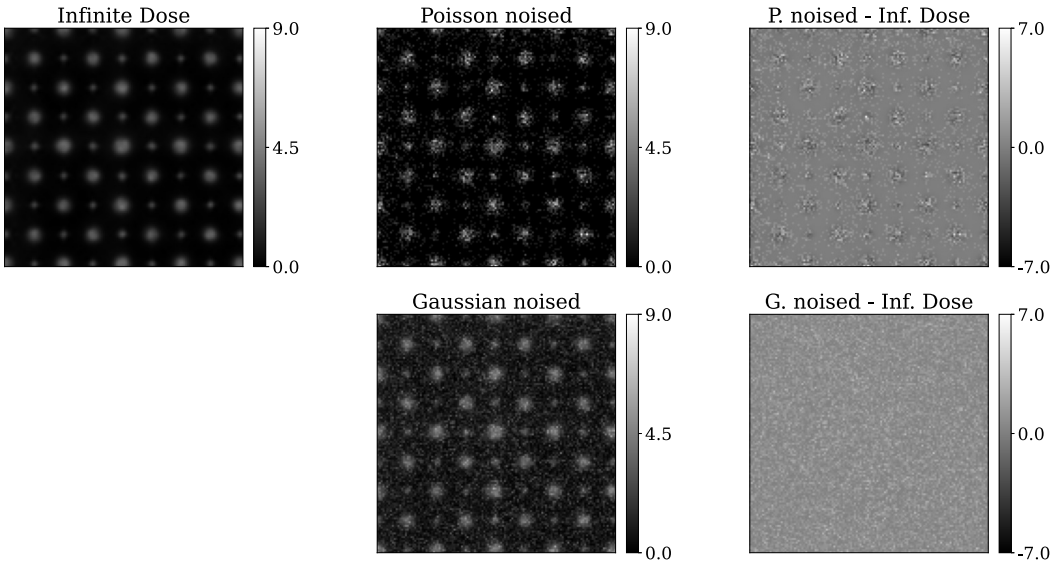

Figure 3.4 – Effect of Poisson and Gaussian noise on the difference between original and noised data. The first column displays a simulated infinite-dose image of simulated TePb. In the second row, noised data are shown, with Poisson and Gaussian noise in the first and second rows, respectively. The difference between the noised and the original data can be found in the third row, from which it is evident that Poisson noise is intensity-dependent, while Gaussian noise is just additive.

information in the residual would not be found in images affected by Gaussian noise only. Gaussian noise, in fact, is not intensity-dependent and therefore does not map onto the crystal structure. Fig. 3.4 can be used as an example to explain the difference between the effect of Poisson and Gaussian noise on the Residual. In the first row, going from left to right, is shown an infinite dose (i.e. noise-free) simulated image of TePb, the same image including Poisson noise, and the difference between the two (noisy-original). In the second row, Gaussian noise has been added in place of the Poisson one and both the noised image and the difference are shown. Focusing on the third column, some periodicity can be seen in the case of the Poisson noise, contrary to the Gaussian case.

A final important consideration is that, for the presented test, the model is able to efficiently denoise the data despite the presence of an amorphous substrate, which generates diffuse scattering and hence additional non-Poisson noise. Substrate scattering was not included in the training set, so that the reconstructed images should be considered as Poisson denoised but still inclusive of substrate scattering.

3.5 Quantitative assessment of the results

Despite being a valuable and rapid practice to inspect the results, visual comparison does not suffice for the purpose of objectively assessing the method's capability. Therefore, some quantitative approaches are proposed in this section, applied to simulated data. Indeed, the ground truth image is needed to quantitatively evaluate the denoising power of the proposed method, and this is not available in the case of experimental microscopy data.

3.5.1 Line profile analysis

A quantitative evaluation of the model performance can be achieved through the so-called line profile analysis. This consists of selecting one line of pixels, along the horizontal direction in this case, and by plotting the pixel intensity at each position. One thus obtains an intensity scan profile that can be used to distinguish atoms of different elements, as shown in Fig. 3.5. The line profile analysis is here conducted on a synthetic image of a 252 Å-thick TePb specimen oriented along the 001 direction, taken at the low dose of 1,000 $e^-/\text{Å}^2$, with a pixel size of 0.18 Å. This corresponds to a pixel dose of 32 e^-/pixel. As we can see from the figure, although the intensity profile of the reconstructed image is not identical to that of the infinite-dose one, the reconstruction appears to be accurate enough to localize and distinguish Te and Pb atoms, namely it contains the same content of information of the ground-truth case. In contrast, when the same line profile analysis is conducted over the noisy image the two different species appear indistinguishable so that the chemical information cannot be extracted. The original images used to perform the line profile analysis can be seen in the bottom row of Fig. 3.5, where the red line specifies the line selected for the study, approximately crossing the atoms at their centre.

3.5.2 Precision of atomic column localization

Another technique that can be used to quantitatively validate the model, involves the determination of atomic column localization. This essentially defines the position of the various atomic columns, thus allowing one to extract quantitative structural information from the STEM measurements. By performing atomic column localization, one can quantify possible lattice strain and measure its error. This is a technologically useful information, since strain affects many physical characteristics of a material, such as the mechanical and electronic properties [134]. Several computational schemes and associated software are available for this purpose, one of them is the Matlab-based package StatSTEM [135]. StatSTEM is based on the principle that, in STEM images, the intensity peaks are located

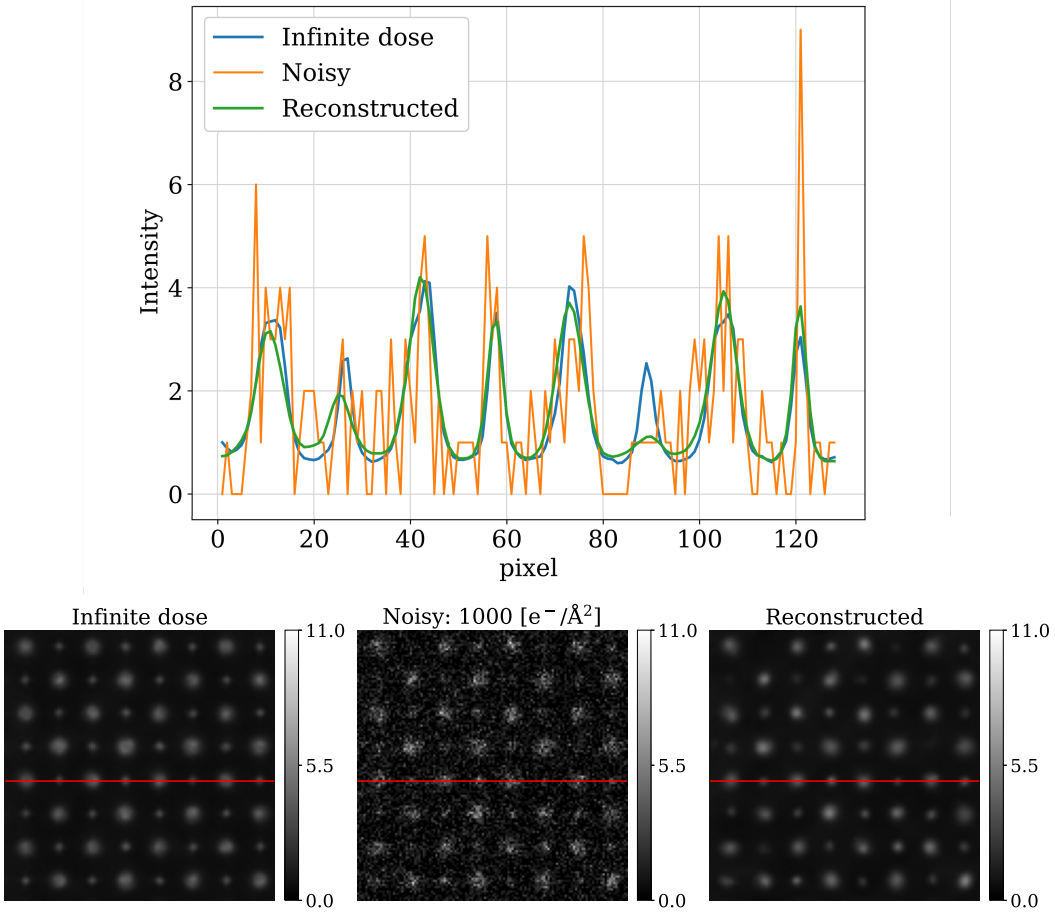

Figure 3.5 – Line profile analysis. On the top row, the image intensity is shown as a function of the horizontal position for the infinite-dose, the reconstructed and the noisy image. A comparison of the peaks intensity allows one to distinguish atoms corresponding to different elements. The high peaks correspond to Pb, which has the highest atomic number, while the lowest peaks correspond to Te atoms, the species with the lower atomic number. In fact, the pixel intensity in dark field images increases with the atomic number [133]. The test is conducted on a simulated image of TePb. The dose value of the noisy image is 32 e^-/pixel with a pixel size of 0.18 Å. The image corresponds to a 252Å-thick TePb slab oriented along the 001 direction. In the bottom row, we show the original images used to conduct the line-profile analysis. The scanning line is plotted in red.

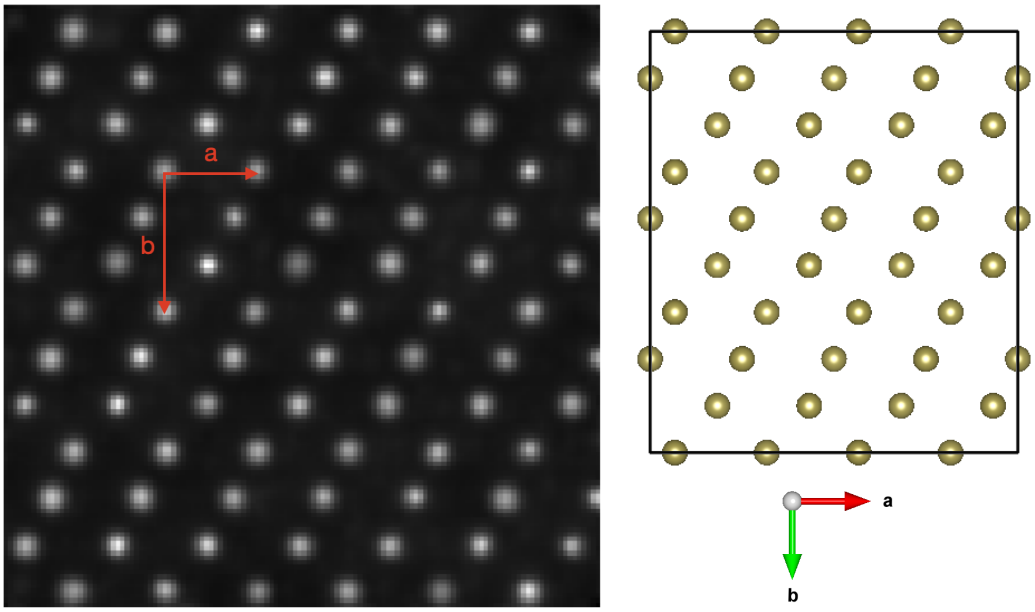

Figure 3.6 – Left-hand side panel: simulated image of a 252 Å-thick Tellurene sample oriented along the 001 direction. The image is taken with a pixel size of 0.2 Å and a dose of 10,000 $e^-/\text{Å}^2$. The red arrows represent the horizontal and vertical distances measured to verify the precision of the atomic column localization. Right-hand side panel: schematic of the crystal structure.

at the atomic column position and these can be approximated by a Gaussian function [136]. Since several localization methods can be used to determine the position of the Gaussians, it is important that our quantitative analysis considers both the localization method and the denoising algorithm, in order to provide a quantitative benchmark of the various possible image-processing workflows.

The precision of the atomic column localization can be estimated by measuring the distances between the various atoms, in both the horizontal and vertical directions. As these are determined solely by the crystalline structure, a statistical distribution of the distances provides a quantitative measure of the accuracy of the combined denoising and localization algorithm. Thus, the standard deviation of the computed distances can be taken as a measure of the accuracy of the localization process and one can compare results obtained for the infinite-dose, the noisy, and the reconstructed images. A reduction in the distance standard deviation corresponds to an enhancement in the image resolvability. The strain error along the horizontal and vertical direction can then be found by dividing the standard deviation in the position by the reference horizontal and vertical distances, respectively.

Figure 3.6 shows the simulated image used for this investigation. This corresponds to a 252 Å-thick Tellurene sample oriented along the 001 direction and imaged with a pixel size of 0.2 Å. The reference horizontal and vertical distances, a

and b respectively, are marked by the red arrows. The easiness of identifying these distances makes Tellurene a good candidate for strain error analysis. The plot in Fig. 3.6 corresponds to the highest dose considered for this analysis, namely 10,000 $e^-/\text{Å}^2$, while denoising is also performed for images taken at 500 $e^-/\text{Å}^2$, 1,000 $e^-/\text{Å}^2$, 2,500 $e^-/\text{Å}^2$, 5,000 $e^-/\text{Å}^2$ and 7,500 $e^-/\text{Å}^2$.

Our denoising autoencoder is then tested against a commonly used algorithm for image processing, namely a Gaussian filter [137]. The procedure to measure the column localization by using StatSTEM is as follows. Firstly, one needs to define the starting coordinates for the atomic column positions, namely the local maxima in the image. This can be achieved by using one of the two available peak-finder routines, which include techniques to smooth the image, to ease the atom localization. The difference between these routines lies in the way the smoothing of the image can be achieved. Specifically, in *Peak-finder routine 1*, there is the option to apply three different filters: an average filter, a disk filter, and a Gaussian filter. In contrast, *Peak-finder routine 2* does not offer an explicit filter adjustment feature, but this can be indirectly achieved by modifying the estimated radius parameter for the atomic columns. For both peak-finder routines, it is possible to specify a threshold value to eliminate undesired pixel intensities from the background. Additionally, *Peak-finder routine 2* provides an additional option, allowing users to define the minimum distance between projected atomic columns in the image. In this work, the *Peak-finder routine 2* was used. The estimated radius parameter was kept to the default value (10 pixels), for each image. The definition of a threshold value to remove nuisance pixel intensities from the background is a necessary step to avoid the identification of too many fictitious atoms in the case of images characterized by strong noise. In order to set the same value for each image and to make the analysis coherent, the intensity is normalized to 1. Then we find that the minimum threshold value compatible with the algorithm memory requirement is 0.12. The minimum distance between atomic columns is kept to the default value of 0 pixels, to avoid the input of material-related information.

After this first step, one has the option to manually add atom positions, an operation that is not considered in this case in order to avoid any human bias in the workflow. Once the starting coordinates are identified, a fitting procedure can be used to model the image as a superposition of Gaussian peaks. In this step, it is possible to specify the width of the atomic columns, by choosing between the *Same* and *Different* options. In the first case, all Gaussian are taken with the same width. In contrast, the second more computationally demanding option, chosen for this analysis, makes the approach more general, since it does not assume that all the Gaussian peaks have the same width. As such, it avoids the introduction of any a priori knowledge of the image. The final result of this procedure is a set of coordinates, which correspond to the atomic columns in the image. These values

can be used to measure the horizontal and vertical distances a and b between the atomic columns, as shown in Fig. 3.6.

The data used for this analysis are displayed in Fig. 3.7. The first column shows 128×128-pixel images of Tellurene simulated at various dose levels, ranging from 500 $e^-/\text{Å}^2$ to 10,000 $e^-/\text{Å}^2$. The reconstruction obtained by using the proposed autoencoder (*Reconstructed AE*) and that obtained with a Gaussian filter (*Reconstructed GF*) can be seen in the second and third columns, respectively. The resolution enhancement is more evident for higher dose levels, for both reconstruction techniques. However, the autoencoder appears to be significantly more successful in the reconstruction task at low doses, and in general at any dose value. Once the atomic columns are localized, following the described procedure, the horizontal and vertical distances, displayed in Fig. 3.6, are measured and they can be represented as histograms, for each image. Fig. 3.8 shows the histograms for the horizontal distance, a, while Fig. 3.9 shows the histograms for the vertical distance, b. The results are here represented as probability histograms, which means that the data are normalized to one. The data associated with the noisy images are placed in the first column, while those associated with the autoencoder and the Gaussian filter reconstructions are in the second and third columns, respectively.

The distribution appears to be more uniform for the autoencoder-reconstructed data (*Reconstructed AE*), compared to the other results. In the noisy and Gaussian-filter-reconstructed (*Reconstructed GF*) images some atoms are misplaced and incorrectly localized, therefore the spread in the histograms appears wider. It should be noted that the scale of the x-axis in the case of the autoencoder-reconstructed data is different from that of the other two columns. The distances reported in the central column of both Fig. 3.8 and Fig. 3.9 appear to be more localized around a single value with a distribution following an approximately normal distribution. Such a feature facilitates the atoms' localization. From the histogram corresponding to each image, the strain error can be computed and compared at different dose values. This is shown for the distances a and b of Tellurene (see Fig. 3.6 for the definition of the distances) in Fig. 3.10. In the figures, the strain error for the infinite-dose case is represented as a blue line and, by definition, it is dose-independent. This, in fact, represents the ultimate theoretical precision achievable by noise-free STEM. In contrast, the noisy images have a strain error that grows with reducing the dose, in an approximate exponential behavior. The autoencoder is able to drastically improve over the noisy images and returns us a strain error that trails closely that corresponding to the infinite-dose case. In more detail, we find that the denoised images present a strain error, which is approximately dose independent when the electron dose is higher than 2,500 e-/Å^2. For lower doses a sharp error increase is reported. Such an increase, however, leaves the strain

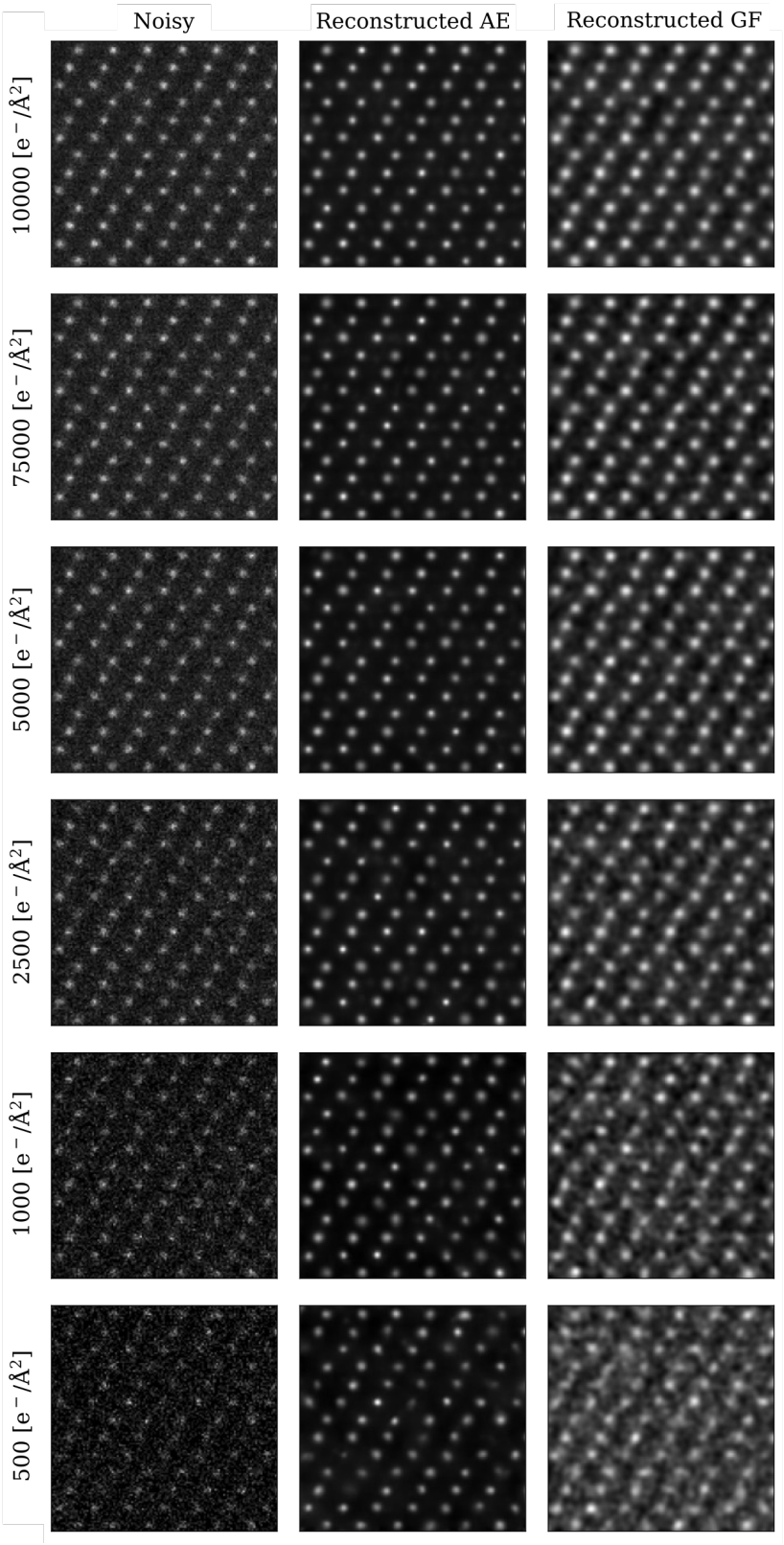

Figure 3.7 – Noisy, autoencoder-reconstructed and Gaussian-filter-reconstructed images of simulated Te at various dose levels.

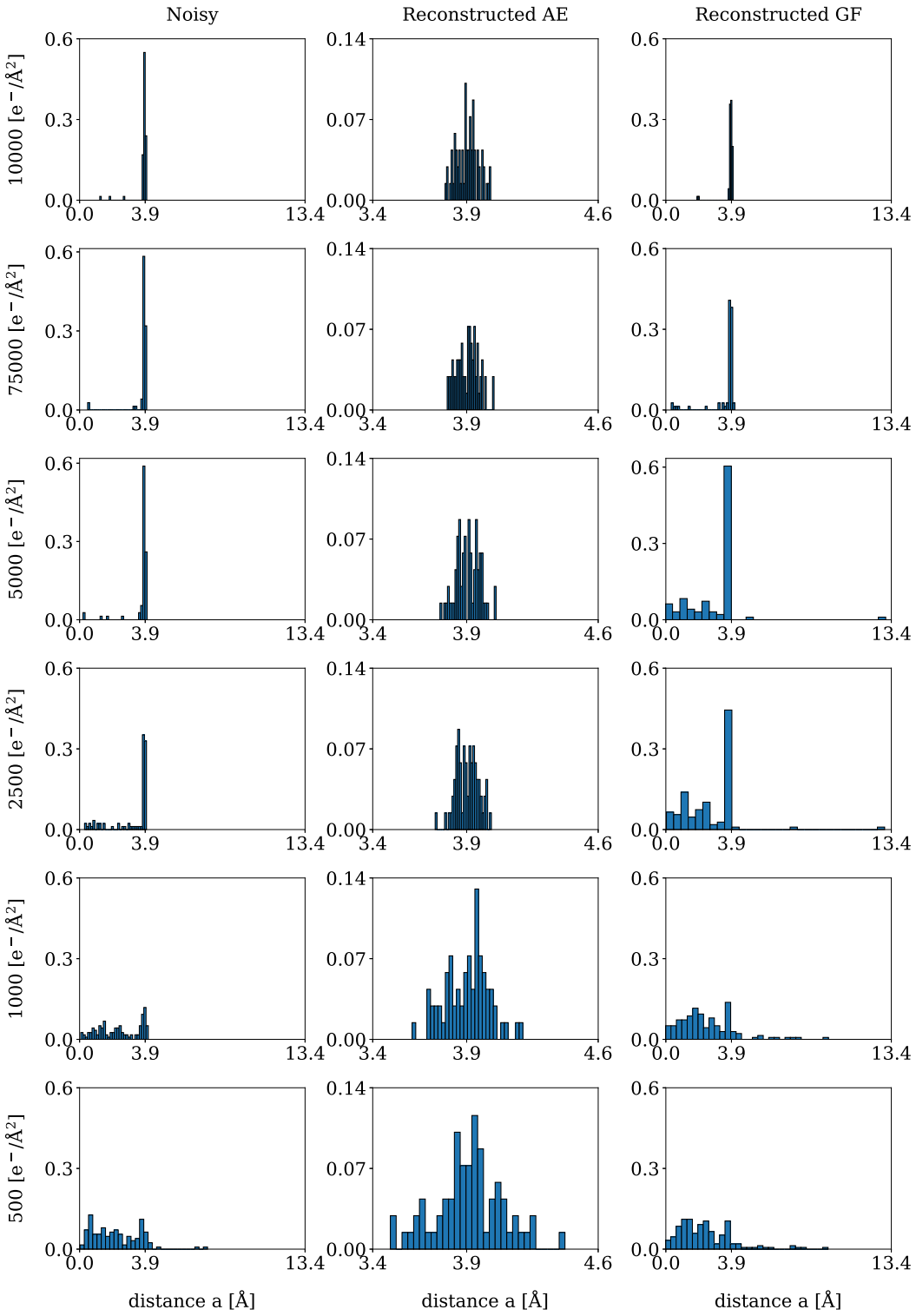

Figure 3.8 – Probability histograms obtained from the measurement of the horizontal distance a between atomic columns for noisy, autoencoder-reconstructed, and Gaussian-filter-reconstructed images of simulated Te at various dose levels.

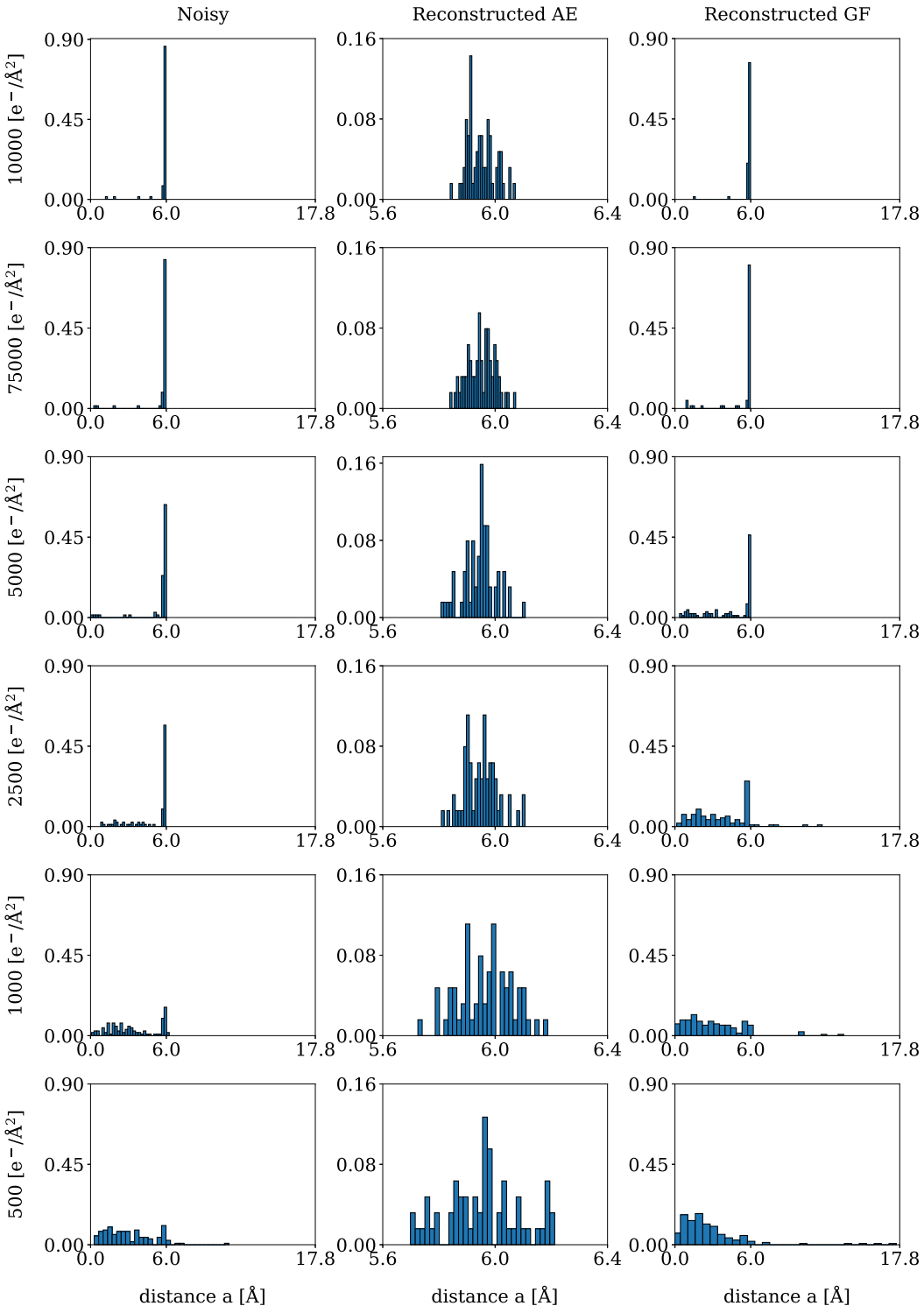

Figure 3.9 – Probability histograms obtained from the measurement of the vertical distance b between atomic columns for noisy, autoencoder-reconstructed, and Gaussian-filter-reconstructed images of simulated Te at various dose levels.

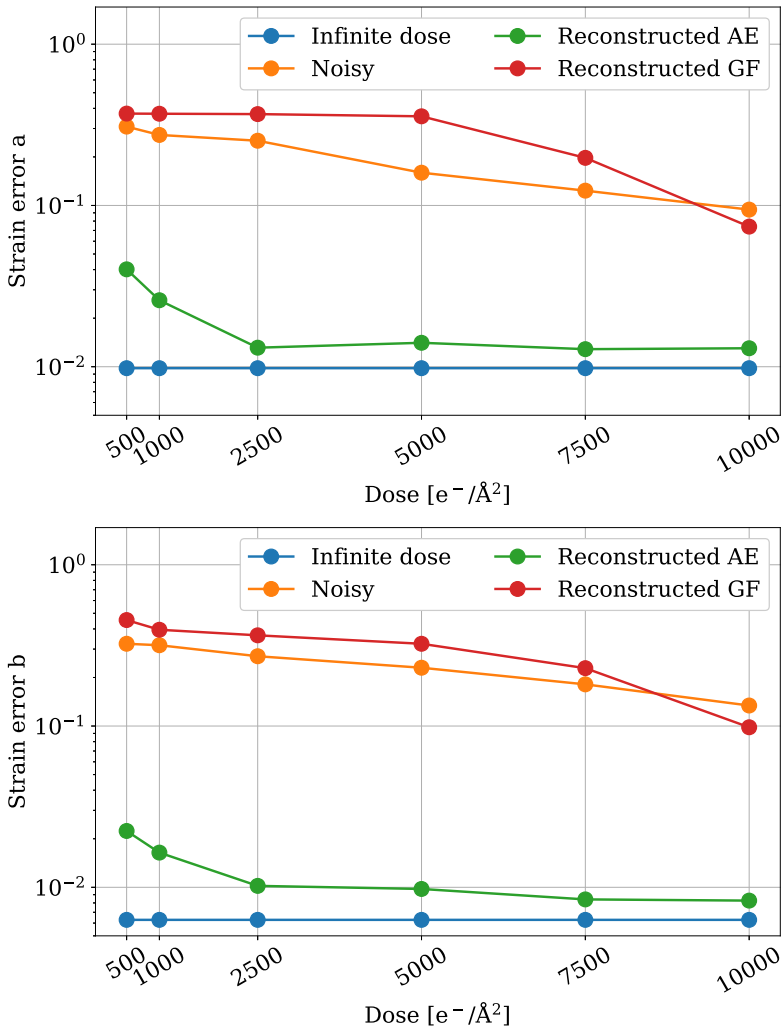

Figure 3.10 – Comparison of the strain error along the horizontal (top panel) and vertical (bottom panel) direction at various dose levels for simulated images of Te. The strain error is computed for the infinite dose (blue line), the noisy image (orange line), and the reconstruction obtained by using the proposed autoencoder (AE - green line) and the Gaussian filter (GF - red line).

error far below what is computed for the noisy images. We then conclude that the autoencoder is less effective at ultra-low doses, but still remains significantly performing across the entire range. Interestingly, simple Gaussian filtering (red lines in the figures) appears unable to improve the column localization of noisy images, with the only exception at very high doses. Therefore, the commonly used Gaussian filter should be avoided when performing accurate quantitive measures of atomic positions, at least when the data-processing workflow remains completely users unbiased. In this situation, our results suggest that even the untreated images provide a better estimate of the column positions, unless rather large doses are used.

We believe that the comparison provided here should represent a general benchmarking scheme to compare different denoising workflows in a completely unbiased way.

3.6 Application to experimental analog data

As a further test, the model was also applied to experimental analog data, characterized by the presence of Gaussian noise in addition to Poisson noise. Specifically, the test was conducted on an image of MoS_2 hold on an amorphous Carbon substrate, acquired at a low dose on a STEM, provided by Valeria Nicolosi (Trinity College Dublin). Therefore, the data is characterized by two additional kinds of disturbance, compared to the dataset used to train the model: Gaussian noise and the noise caused by the presence of the amorphous substrate. Nevertheless, the reconstructed MoS_2 image, displayed on the right-hand side of Fig. 3.11, shows that the denoising process is still successful. Specifically, in the denoised image, it is possible to observe some of the Sulfur atoms that were not visible in the original noisy image. Importantly, despite the presence of MoS_2 in the training set, some of the Sulfur atoms are missing from the reconstruction. This is a significant proof that the model is not biased by the knowledge of the crystalline structure. In fact, if the signal is too low, we do not expect the autoencoder to display any atoms.

The physical reason behind the absence of some of the Sulfur atoms could be correlated to the sample preparation process, which impacts the number of vacancies present in the crystalline structure. Quantification of the percentage of vacancies related to each preparation procedure is often hindered by the presence of noise, which characterizes the low-dose measurements. Fig. 3.11 demonstrates that, with the help of denoising machine-learning techniques, this quantification can be pursued. This result motivated the development of the project described in Chapter 4.

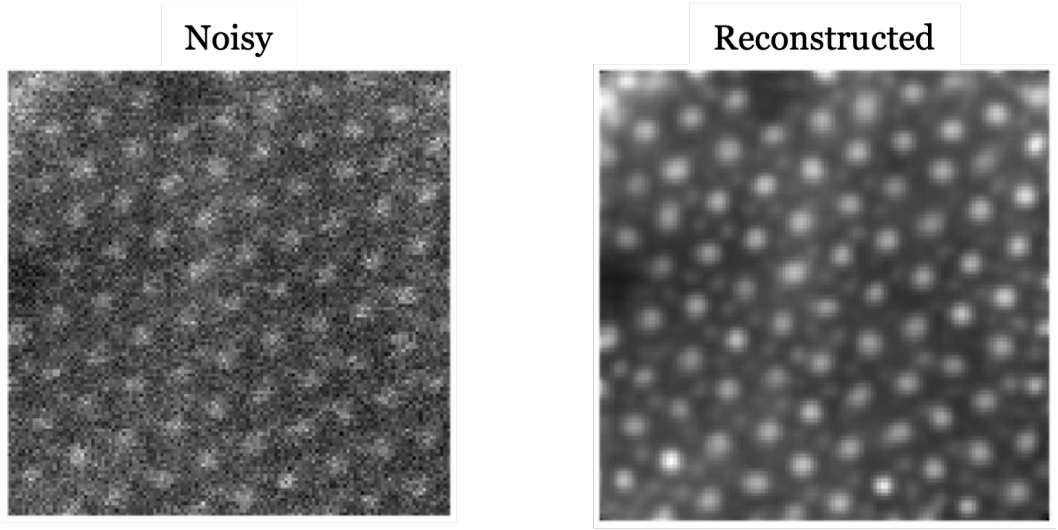

Figure 3.11 – Example of application of the proposed model to an analog experimental image of MoS_2 deposited on an amorphous Carbon substrate, imaged at low dose on a STEM. This example is particularly challenging due to the presence of Gaussian noise, which characterizes analog data acquisition, and the amorphous substrate: none of these features was included in the training set. The left-hand-side panel shows the original noisy image and the right-hand-side panel displays the reconstruction after the application of the autoencoder. Notably, some of the Sulfur atoms that were not visible in the original data, appear in the reconstructed image. Although MoS_2 was included in the training set, the absence of some of the Sulfur atoms demonstrates that the model is not biased by information about the crystalline structure.

3.7 Conclusions

This chapter proposed a solution to one of the limiting factors of STEM technology, namely the requirement of a high electron dose to achieve atomic resolution images, which can damage the analyzed specimen. In fact, a high electron beam maximizes the signal-to-noise ratio, but, due to knock-on and radiolysis damage mechanisms, can alter the specimen during the experimental measurement. Depending on the level of electron dose employed, the resulting image will be affected by a different level of Poisson noise. Other kinds of noise can be neglected in the case of digital acquisition in state-of-the-art imaging instruments.

To tackle this problem, we have proposed a neural network trained on simulated dark-field STEM images, which allows one to successfully remove Poisson noise from low-dose digital data. The proposed dataset reproduces realistic data acquisition conditions and includes variety in terms of specimen and microscope settings, which prevent biases in the model, such as awareness of materials symmetries.

Tests on both simulated and experimental images have been conducted to thoroughly validate the model. The results obtained from these tests show a clear improvement in the image resolution and in the possibility of extracting

useful information from the data in a completely unbiased way. The use of this model may allow a drastic reduction of the dose level to be employed in real-life measurements, making it possible to analyze very beam-sensitive compounds, that would otherwise be challenging to study due to their susceptibility to radiation damage.

Notably, our proposed denoising algorithm is completely autonomous and does not require any human input or knowledge of the actual beam intensity, being trained over a broad range of doses. This ensures its adaptability to a wide range of experimental conditions, and the possibility to use it during live acquisition.

Crucially, the denoising process of a 128×128-pixel image can be performed within approximately one second.

Future developments of this project involve an integration of this tool into the microscope setup, in order to obtain denoised data during live acquisition. This would allow the user to denoise part of the image during the measurements and adjust the electron dose depending on the quality of information retrievable after the application of the autoencoder.

Another interesting study would concern the investigation of the latent representation generated by the encoder. This type of analysis is mainly conducted in the case of generative models. However, for this case, it could allow further understanding of the feature extraction process carried out by the autoencoder to remove the Poisson noise. Due to the presence of multiple layers in the latent space, simple visualization might not be adequate to explore the data. Clustering algorithms might be an appropriate solution to aid the latent space interpretation. Representations of the first layer of the latent space are displayed in Fig. 3.12, for two cases, Si and MoS_2, indicated as *Case 1* and *Case 2*, respectively. From these representations, it is evident that the association of features in the input and output cannot be automatically linked to the latent space.

Finally, to further validate the proposed model, it would be important to develop additional assessment procedures applicable to the experimental data. In this case, full-reference metrics cannot be used, due to the lack of ground-truth data. No-reference metrics for the investigation of the reconstruction results might involve microscopy experts, who could provide quality scores for images before and after processing them.

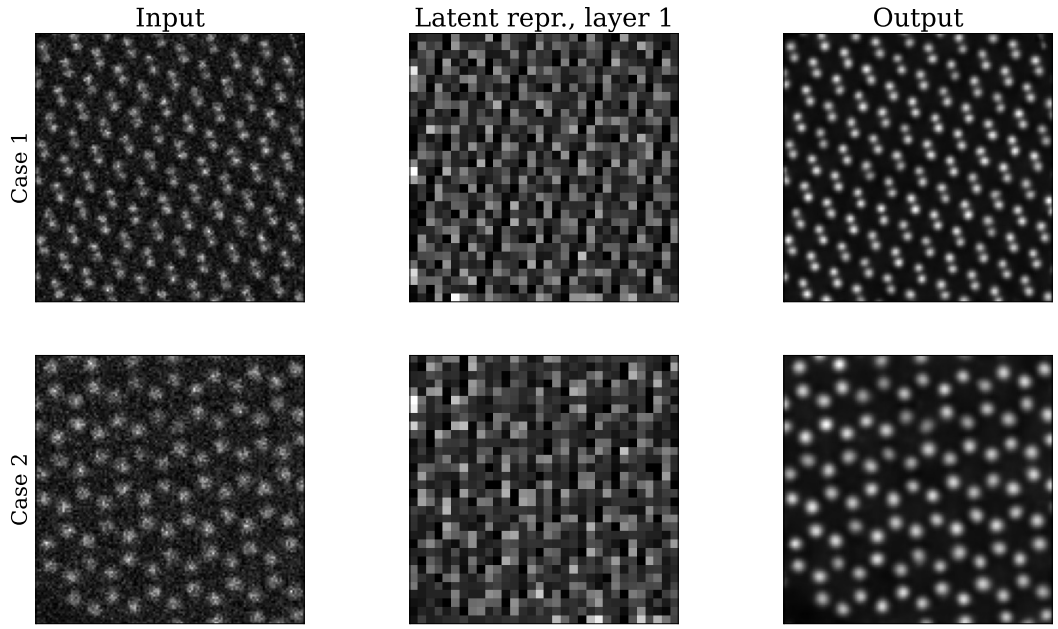

Figure 3.12 – Visualization of the first layer of latent representation for two different materials (on the top row Si, on the bottom row MoS_2), in the central column. The autoencoder's input and output data are in the first and third columns, respectively.

CHAPTER 4

VACANCIES COUNTING IN STEM-IMAGED TMDS

This chapter can be considered an extension of the previous one. The main aim of this research project is to estimate the amount of vacancies present in exfoliated transition metal dichalcogenides (TMDs) samples, imaged with scanning transmission electron microscopes (STEMs). In order to make the atoms more visible and facilitate the counting process, the same denoising algorithm described in the previous chapter will be used. However, being the data affected by different types of noise, as explained briefly, the model will be trained on a different training set.

4.1 Problem description and state-of-the-art methods

One of the most significant achievements in the field of material science, in the last century, is the isolation and characterization of graphene [138]. This achievement has led to great effort and advancements in the study of layered materials, due to their potential applications across a wide range of research and industrial domains. Among others, the class of materials known as transition metal dichalcogenides (TMDs) has gained particular attention in the materials science and condensed matter physics communities. These are layered compounds consisting of transition metal atoms, which are placed between two layers of chalcogen atoms. Transition metals are chemical elements located in the central part of the periodic table, specifically, in groups 3 to 12. Chalcogen atoms, instead belong to group 16. These atomically thin layers are stacked together through weak van der Waals forces

(i.e. weak electrostatic interactions induced by transient dipoles). The general formula of these compounds is MX_2, where M represents the transition metal, such as Molybdenum or Tungsten, and X identifies the chalcogens, which can for example be Sulfur or Selenium. They exhibit several interesting properties, from the electronic, optical, and mechanical points of view. These features make them suitable for numerous applications in photonics, electronics, catalysis, energy storage, and others [139]. Contrarily to graphene, which usually exhibits semi-metallic properties, TMDs predominantly present semiconductor behaviour. This means that by applying an external electric field or by varying their thickness it is possible to modify their electrical conductivity. Clearly, the presence of vacancies can significantly impact the material's properties and, therefore, the possible use. However, the exact correlation between vacancies and property alteration is still not completely established [140].

Two main techniques are used to isolate single-layer TDMs: mechanical exfoliation and liquid phase exfoliation. Mechanical exfoliation [141], also used in the original work that led to the isolation of graphene, allows one to extract atomically thin layers of materials from a bulk sample by using a piece of adhesive tape on the surface of the material. When the tape is peeled off the bulk material, thin layers are obtained. This simple and low-cost method allows the production of high-quality flakes. In this context, the term *flakes* refers to individual atomically thin layers of materials obtained from a bulk source. They can be monolayer or a few-layers thick. The mechanical exfoliation technique presents some disadvantages. In fact, it is not suitable for large-scale production and lacks precise control over layer thickness, size, and contamination. A possible alternative is the so-called liquid phase exfoliation method [142], which requires a liquid dispersion to break down bulk materials into layers. The procedure consists of reducing the material into a fine powder, which is then mixed with a solvent. Ultrasonication is applied to disperse the material in the liquid and obtain individual or few-layer flakes. This method has the advantage of being scalable and therefore suitable for industrial production. Moreover, by adjusting the sonication time and the solvent, it is possible to control the thickness of the exfoliated layers. However, some drawbacks are associated with this procedure. Firstly, the quality and size of the obtained flakes can vary among batches, making the results inconsistent and unpredictable. Additionally, the property of the resulting material can be affected by surfactants or stabilizers used to prevent re-aggregation of the exfoliated flakes in the dispersion. Another disadvantage concerns the choice of solvent, which might not be suitable for all materials.

Regardless of the chosen preparation method, it is crucial to examine the obtained flakes in order to assess their quality, which will affect the material's properties. A valuable strategy is to use a Scanning Transmission Electron Micro-

scope (STEM), whose functioning principle is detailed in Chapter 2. By imaging the samples at the atomic level, it is possible to ascertain the quality of the flakes. Specifically, it is possible to assess how many chalcogens were lost during the sample preparation process. In fact, several mechanisms can cause the loss of these light elements, including oxidation, chemical reactions, and surface contamination [143].

As already mentioned, the presence of defects can highly affect the material's properties, both positively and negatively. In order to have a more complete understanding of this effect, it is important to quantify the amount of vacancies present in the specimen. This can be challenging, even with a high-resolution STEM, due to difficulties in capturing light atoms coexisting with heavy ones, in the presence of several types of noise. It should be noted that Sulfur can also be removed during the characterization process, for instance, due to the electron beam used during STEM imaging, as explained in Chapter 2. The solution proposed in this work is to use a machine-learning model to improve the quality of low dose STEM-acquired images of TMDs and therefore facilitate the vacancy-counting process, while keeping the dose low.

The test set used for this purpose is made of experimental images acquired both on the Titan and Nion microscopes, available at Trinity College Dublin, which will be described in the next section. The data are acquired in an analog mode, not digital. Therefore, the main types of noise affecting them are Gaussian and Poisson. In the case of analog data, the pixel values are usually rescaled between 0 and 1 before being processed by machine-learning models. In fact, in this case, in contrast to what happens for digital data, the units of the pixel intensity are arbitrary intensities and do not have physical meaning. The choice of analog images over digital ones is motivated by data availability. For future studies, digital acquisition will be considered.

It should be noted that, for this project, we are not interested in achieving the best possible reconstruction with the developed neural network. The main goal is to retrieve enough signal to localize the atoms in the image, in order to quantify the vacancies and, consequently, assess the sample quality.

As mentioned before, the sample preparation process is not the only source of possible vacancies, but more factors should be considered. Firstly, the type of microscope can affect the observation from two perspectives: on the one hand using a more powerful microscope, such as the Nion, enables the acquisition of higher-resolution data, easing light-elements identification. On the other hand, using a stronger electron beam to image the material can generate additional vacancies. Using a microscope like the Titan can hinder the procedure, due to the high noise affecting the images. Additionally, the proposed denoising process also has an impact on the results, since it can generate apparent vacancies. This effect needs to be quantified. Finally, results inaccuracy can also be caused by

the vacancy localization and counting procedure, which is performed using the software Atomap [144] in this case.

This study aims to be an example of possible applications of a machine-learning approach to assist investigations in the field of electron microscopy imaging, but further investigations are needed. Importantly, the purpose of this chapter is to present a practical application of the denoising model developed in the previous chapter and to demonstrate the importance of image denoising in microscopy data analysis. Future studies will involve a comparison between the proposed pipeline and neural networks developed for direct vacancies counting in noisy electron microscopy data [145]. This will allow a more complete understanding of the impact of the denoising model within the vacancy counting procedure.

4.2 Model construction and training set

The neural network used for this applicative example is similar to the one developed in the previous chapter. However, being the data affected by noise of a different nature, some modifications are needed, both for the model and the training set.

The model's architecture is identical to the one described in Section 3.2, except for the last layer, which, in this case, uses *Sigmoid* instead of *ReLu*, as an activation function. The reason behind this choice is correlated to the nature of the training set. In fact, in this case, the model is constructed with the aim of being used on analog data. Being the pixel intensities not associated to physical meaning, all images are rescaled in order to have grayscale values, between 0 and 1. This means that the output of the autoencoder should be in the range 0-1, a feature that makes the use of the *Sigmoid* function suitable for this application, since it always returns values within the mentioned range.

Regarding the dataset construction, which is made of 512×512-pixels images, the Prismatic software [30, 31] is used. However, less variety of materials is included compared to the training set presented in Chapter 2, being the main focus of this project on transition metals dichalcogenides. Therefore, the species included are MoS_2, WS_2, $MoSe_2$. Importantly, different versions of each material are generated, with different levels of vacancies, in several spatial configurations. For each material, the following imaging conditions are considered (more information on the meaning of these parameters can be found in the Prismatic documentation [146]):

- the pixel size values are 0.14, 0.18, 0.25 and 0.3 Å;
- the size of the simulated potential is 0.08 and 0.1 Å;
- the electron beam energy is 100 keV;

- the sample thickness range is $6 - 36$ Å.
- the inner and outer detector collection angles are 60 and 180 mrad for MoS_2 and WS_2, 30 and 70 mrad for $MoSe_2$;
- the maximum probe angle is 27 mrad;
- the potential bound is 2 Å;
- the probe tilt is 0 mrad in both horizontal and vertical directions;
- the probe defocus is 0 mrad;
- the number of frozen phonon configurations to calculate is one;
- the thermal effects are included in the simulation;
- the occupancy values for likelihood of atoms existing at each site are included.

The steps followed to generate the different types of noise are displayed in Fig. 4.1, for one example from the training set, namely a simulated image of MoS_2, without any Sulfur vacancies. The various steps of the procedure are displayed in the first column, while the second column shows the different types of noise added at each step. It should be noted that the pixel values are scaled between 0 and 1 after each step. The first image in the first row, designated as *Step 0*, is the infinite-dose image (without noise) generated by Prismatic. Poisson noise is applied to this image, with dose value randomly selected from the range 1,000 - 10,000 $e^-/\text{Å}^2$. One example of this is depicted in the first column of the second row, with label *Step 1*, while Poisson noise is shown in the same row, on the right-hand-side panel. As can be seen from this image and as discussed in Chapter 3, this kind of noise depends on the intensity of the image to which it is applied. Subsequently, Gaussian noise is added to the Poisson-noised image, with variance within the range of 0-0.05. This phase is indicated as *Step 2*, in the panel located in the first column and the third row of the plot, where an image affected by both Poisson and Gaussian noise is displayed. The added noise is represented on the right-hand-side of this panel. In order to obtain more realistic results, as displayed in the last panel of the first column, namely *Step 3*, the presence of a non-constant background is reproduced. To do so, a black image of the same size as the training set elements was created, namely a 512×512-pixels image made of all zeros. Twenty random white points on the black background are then generated by setting random pixel coordinates to white (pixel value set to 1). Gaussian blur is subsequently added to this image, to create a smooth and diffused appearance in correspondence of the white points. Finally, this blurred background (right-hand-side panel of the last row) and the image with Poisson and Gaussian noise are blended together, using a

blending factor of 0.5. It should be noted that different combinations of several levels of noise were considered for each image.

As it can be deduced by this description, the dataset proposed in this case presents less variety, compared to the one detailed in the previous chapter. This is due to the different goals of the two projects. In fact, for this investigation, the aim is not to provide the best-performing model but to develop a model that returns enough signal to distinguish atoms and vacancies in a specific class of materials.

Different data augmentation strategies are implemented, namely rotation and resizing. The final training set consists of about 10,000 images. The model is trained for 200 epochs, using the Adam optimizer and the customized loss function defined in Eq. (3.1).

Some examples of reconstruction achievable with this model can be found in Fig. 4.2, for the case of simulated MoS_2 data with some vacancies, affected by different levels of noise. The first column shows the original infinite dose image, which is the same for all five displayed cases. Some noisy versions of it can be found in the second column. Specifically, different dose values are considered, which determine different levels of Poisson noise, while the intensity of Gaussian noise is selected randomly within the defined range. The background is also simulated randomly, according to the previously described procedure. Going from *Case (a)* to *Case (e)*, the selected values of electron dose are: 1,000 $e^-/\text{Å}^2$, 2,500 $e^-/\text{Å}^2$, 5,000 $e^-/\text{Å}^2$, 7,500 $e^-/\text{Å}^2$, 10,000 $e^-/\text{Å}^2$. The reconstruction obtained with the developed autoencoder is presented in the third column. The first three columns are displayed in scales of gray, with values ranging from 0 to 1. The last column shows the difference between the original image and the autoencoder reconstructed image, for all the examined cases, computed by subtracting the grayscale intensity of each pixel. In this representation, the blue regions indicate that the reconstructed image is lighter than the original image, while the red regions represent the opposite situation, where the reconstructed image appears darker than the original one. As can be deduced from these results, the reconstruction obtained with the proposed autoencoder is not perfect. In particular, the most evident inaccuracy from the difference plots depicted in the fourth column of Fig. 4.2, is the presence of blue circles surrounding the atom locations. The white centre indicates that the atomic column centre is correctly identified. However, the reconstruction seems to generate atoms that are systematically wider than the original data, as indicated by the blue pixels around the atom locations. Nonetheless, this feature does not seem to have a significant impact on the atom localization, which is the main goal of this project. An accurate quantification of the model effect on the vacancy counting process needs to be performed, as will be discussed later in the chapter.

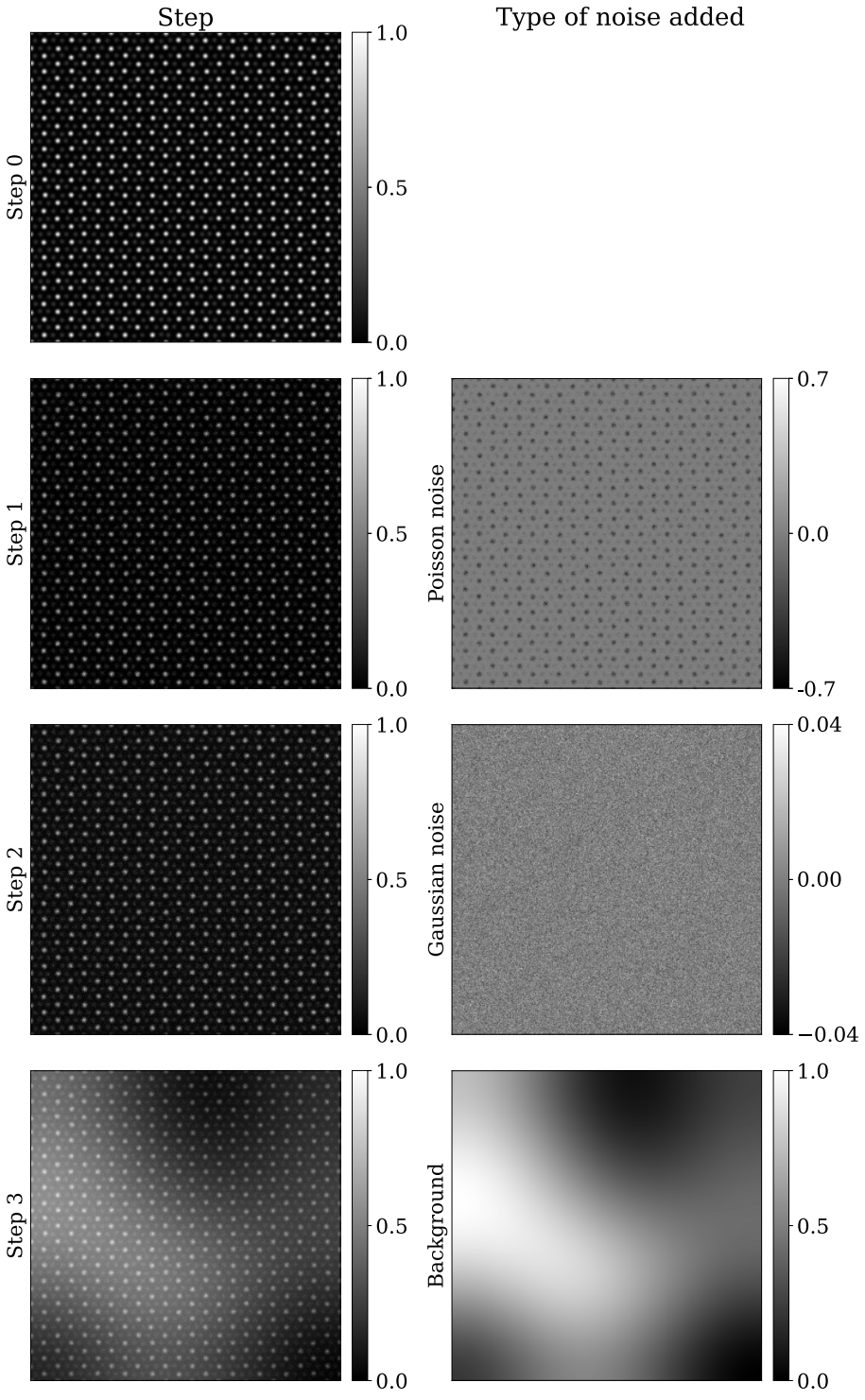

Figure 4.1 – Illustration of the procedure followed to add noise to training set samples. A simulated image of MoS_2 without any Sulfur vacancies is used as an example. The various steps of the noising procedure are depicted in the first column, where in *Step 0* there is no noise, image, in *Step 1* Poisson noise is added, in *Step 2* Gaussian noise is added, and, lastly, *Step 3* is the final result, which includes a background. The images in the second column display the type of noise that is added at each step.

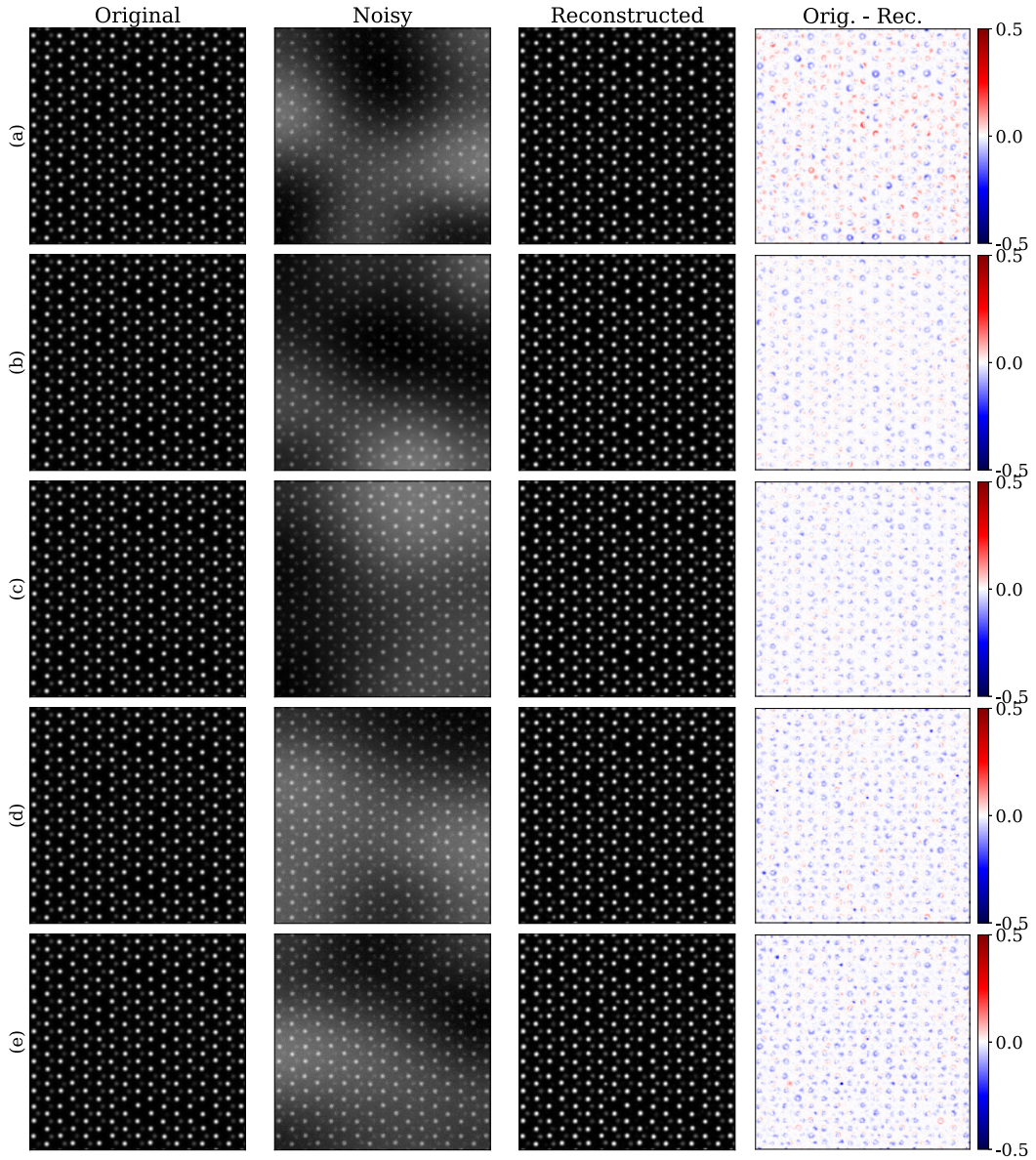

Figure 4.2 – Image reconstruction of simulated MoS_2, with vacancies, at five different noise levels. The first column displays the original data, which is the same for all five examples. The second column shows the noisy version of the data, generated by randomly adding Gaussian noise with variance within the range 0-0.05 and a synthetic background obtained following the procedure described in Section 4.2. Poisson noise is applied according to five dose levels, namely 1,000 $e^-/Å^2$, 2,500 $e^-/Å^2$, 5,000 $e^-/Å^2$, 7,500 $e^-/Å^2$, 10,000 $e^-/Å^2$. The dose level increases going from *Case (a)* in the first row to *Case (e)* in the last row. The reconstruction achieved with the proposed autoencoder is displayed in the third column. Finally, the fourth column presents the difference between the original and the reconstructed data, for all examined cases.

4.3 Vacancies counting procedure on experimental data

The software used in this case for atomic column localization is Atomap [144]. This was chosen, in place of the previously used StatSTEM [135], due to its Python implementation, which is open-source, in contrast to MATLAB, upon which StatSTEM is built. Additionally, Atomap provides a pipeline for localizing atoms of different species separately (namely different sublattices), which is favourable for the set of data investigated in this part of the project. Nonetheless, the localization process appears to be less efficient compared to StatSTEM, requiring additional input from the user, especially in localizing the atoms at the boundary of the images. An interactive interface can be used to remove or add atom positions in the examined data.

The procedure to quantify the vacancies in TMD data consists of several phases, which are now described.

Data acquisition

The experimental images presented in this chapter are acquired on a Nion UltraSTEM and a Titan S/TEM, by Danielle Douglas-Henry, at the CRANN Advanced Microscopy Laboratory (AML www.tcd.ie/crann/aml/), at Trinity College Dublin. The maximum resolution achievable by these instruments is 0.78 Å for the Nion, and 2 Å for the Titan, respectively. The materials considered for this investigation are MoS_2, WS_2, and $PtSe_2$, all placed on a Carbon substrate grid. All the samples were prepared through liquid phase exfoliation, with a resulting thickness of a few layers. Additionally, mechanically exfoliated samples are available for the case of $PtSe_2$.

For the preliminary results presented in this chapter, only the MoS_2 dataset is considered, with images generated with both microscopes. In particular, for the Titan acquisition, a current of about 20 pA and voltage of 300 kV is used, while for the Nion acquisition, the current is about 40 pA and the voltage 40 kV. It is worth mentioning that an exact dose value is not available in the case of analog acquisition.

Data pre-processing

All the experimental images need to be pre-processed before undergoing the atomic column localization with the software Atomap. Specifically, after rescaling the pixel intensities between 0 and 1, the developed denoising autoencoder is used to improve the image quality and therefore facilitate the vacancy counting process. Nonetheless, the quality of some of the images remains low even after the application of the autoencoder. Therefore, these data are discarded. Examples of

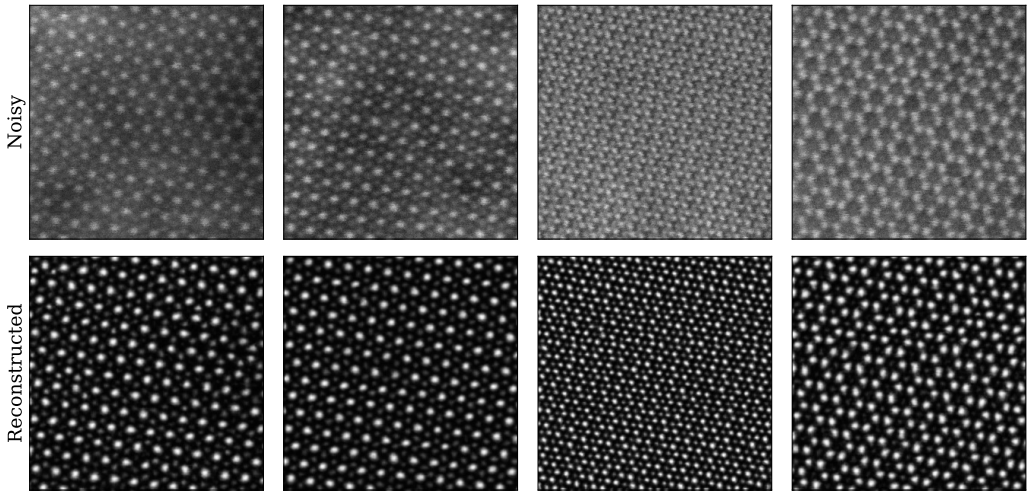

Figure 4.3 – Examples of STEM-acquired images of MoS_2, before (top row) and after (bottom row) denoising them with the developed autoencoder. The first two columns display data captured with the Nion microscope, while the last two columns show Titan-acquired data. These are all examples of images that can be considered for our analysis since the amount of signal enables Sulfur atom localization.

acceptable original noisy (top row) and denoised (bottom row) images are depicted in Fig. 4.3, which show good signal content. Some of the discarded cases are illustrated in Fig. 4.4. In this case, it is evident that for both the original (top row) and reconstructed (bottom row) images the signal content does not allow one to clearly distinguish atoms. In both Figures, the first two columns show data acquired with the Nion microscope, while the last two columns represent data acquired with the Titan microscope.

At the end of this denoising and screening process, for the case of MoS_2, 27 images from the Nion dataset and 13 from the Titan dataset are retained. In order to illustrate the procedure presented in this chapter, an image from the Nion dataset is considered. The original and denoised versions of this example are presented in panels (a) and (b) of Fig. 4.5, respectively. From both images, the presence of scan noise can be detected. However, this does not seem to impact the atom identification. It should be noted that the neural network was not trained to remove scan noise.

First lattice identification and refinement

The selected images are then analyzed with Atomap. The first step consists in localizing the atomic column of the first sublattice, made of the most intense atoms, meaning the atoms with the higher atomic number. In the case of TMD, these are the transition metal atoms, Molybdenum for the analysis of MoS_2. According to the Atomap documentation [147], a peak finding algorithm from the Python

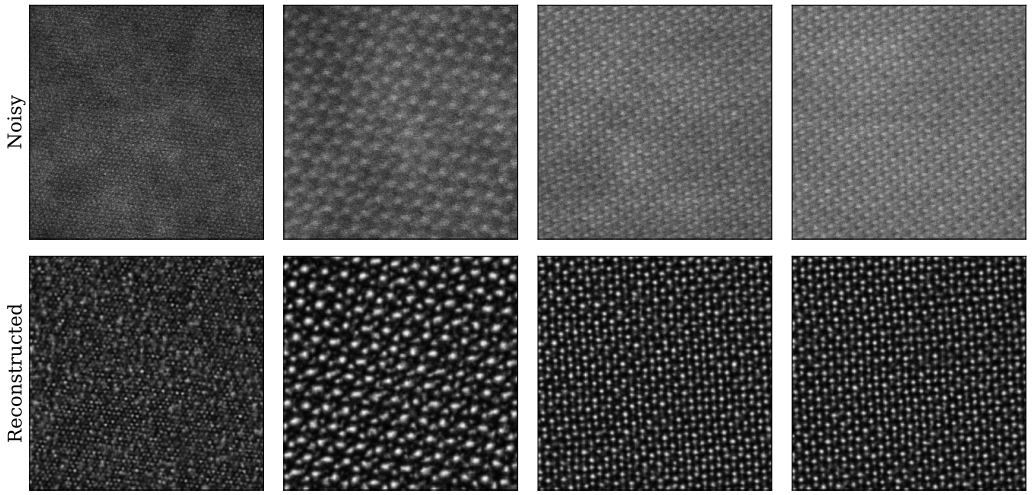

Figure 4.4 – Examples of STEM-acquired images of MoS_2, before (top row) and after (bottom row) denoising them with the developed autoencoder. The first two columns display data captured with the Nion microscope, while the last two columns show Titan-acquired data. These are all examples of images that cannot be considered for our analysis since the amount of signal does not enable Sulfur atom localization. Therefore, these images are discarded.

package *skimage* [137] is used to find the initial atom positions. The minimum separation of the features should be provided, measured in pixels. It is important to select an appropriate value for this parameter. In fact, if the peak separation value is too small, too many atoms are found, while if it is set to a very high value, too few atoms are localized. For the examples of MoS_2 data displayed in this section, this parameter was set to 20 pixels; this value is selected after a visual assessment of the resulting initial atom positions. In fact, the appropriate distance depends on the magnification of the experimentally captured data. If some of the atoms are incorrectly located, which happens principally at the image boundaries, the user can manually add/remove them from an interactive interface. Once the initial positions are established, the nearest neighbours of each atomic column are found, and then they are refined by Atomap by using the centre of mass and 2D Gaussians, as detailed in the Atomap documentation [147]. An example of the first sublattice localization is shown in panel (c) of Fig. 4.5, for a Nion-generated image of MoS_2.

Second lattice identification and refinement

After the localization of the atoms belonging to the first sublattice, these need to be hidden from the image, to allow the identification of the second sublattice. In fact, as mentioned before, the peak finding procedure detects the atoms with the highest intensity. The procedure proposed by the Atomap documentation is

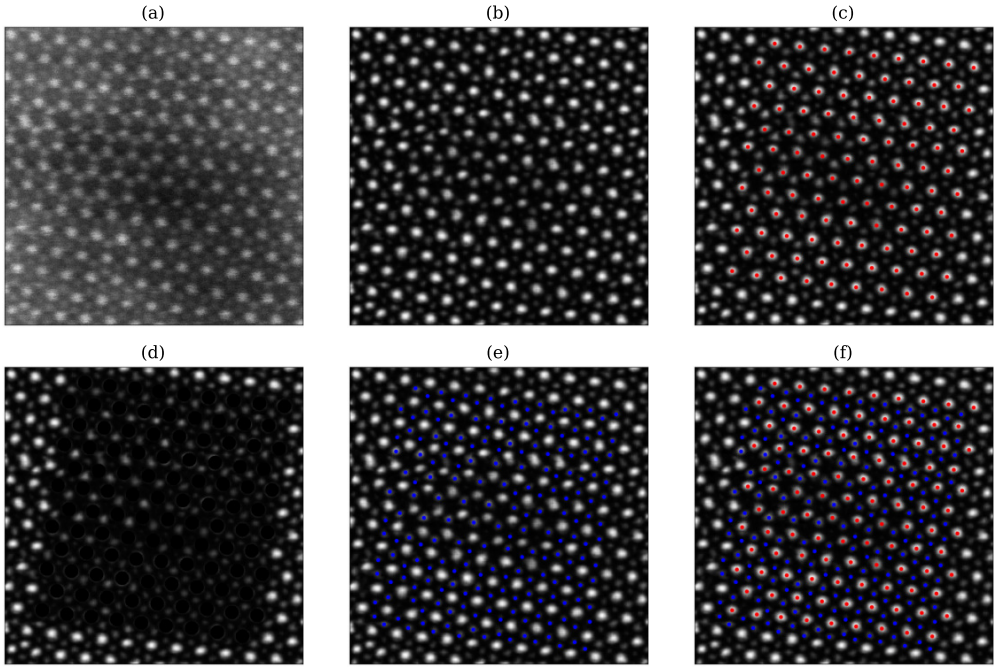

Figure 4.5 – Illustration of the proposed solution for vacancies localization in STEM-captured MoS_2 data. In this example, an image acquired with the Nion microscope is considered. Panel (a) shows the original noisy image. The autoencoder-reconstructed version of the same image can be found in panel (b). This is the input data for the localization procedure performed with the software Atomap [144], whose first step is depicted in panel (c), representing the localization of Mo atoms. Once their position is established [red dots in panel (c)], they are hidden from the image through black circles, as can be seen in panel (d). This facilitates the localization of S atoms, whose locations are depicted with blue dots in panel (e). Manual changes are needed for this step to refine the positions of the atoms. The final lattice is shown in panel (f), from which it is evident that some atoms have been manually discarded, in order to fulfill wraparound lattice conditions.

to mask out these atoms by generating a Gaussian blur in correspondence with their positions. However, an alternative masking process was implemented within this project, which uses black circles of customizable diameter to mask the atoms. Panel (d) of Fig. 4.5 shows the outcome of this masking phase, where a diameter of 12 pixels was chosen for the masks.

Now the atoms belonging to the second sublattice, namely the chalcogens (Sulfur in the case of MoS_2), can be localized, following the same procedure described for the first sublattice. In this case, the peak separation value is set to 14 pixels. The resulting atomic columns are represented with blue dots in the illustration presented in panel (e) of Fig. 4.5. This phase of the procedure is less efficient, due to the small distance and the lower intensity of the atoms. Therefore, numerous modifications from the user are expected.

Importantly, the chalcogen atoms at the top and right border of the image are

intentionally discarded, in accordance to the wraparound lattice approximation. This is done in view of the vacancy counting phase, which can be challenging at the edge areas, since the atoms might not be visible due to the finite field of view. The wraparound lattice approximation can be used to mitigate this edge effect when dealing with images of periodic materials and involves considering the atoms near the edges as if they are connected to the atoms located at the opposite edge. Clearly, this is an approximation and can lead to misleading results, since it assumes that the amount of vacancies at one edge of the image is the same as that at the opposite edge.

Results analysis

The outcome of the image taken as an example of the proposed procedure can be found in panel (f) of Fig. 4.5. The same process is followed for all the available data. A list of coordinates for each sublattice is obtained after the use of the software Atomap. Clearly, the length of these two coordinates lists is equivalent to the number of located atoms for each of the species, namely transition metals and chalcogens. The stoichiometry for TMD, MX_2, indicates that two chalcogen atoms are expected for each metal atom, for an ideal lattice. Therefore, if this condition is not satisfied, it indicates the presence of some vacancies in the examined image, and a percentage of defects can be calculated. According to the investigated data, the images acquired with the Nion present an average percentage of vacancies of 4.3 %, while the average percentage for the Titan is 6.1 %. The distribution of these percentages can be found in Fig. 4.6, where the black dots represent the mean values for the two investigated datasets (Nion and Titan), and the dark gray diamond-shaped markers indicate the outliers in the distributions.

It is important to note that these results are affected by several sources of errors, and some of the vacancies are probably just apparent. Firstly, the level of detail that can be achieved with the two microscopes surely affects the results, with an impact on the denoising process and, ultimately, on vacancy identification. Specifically, the Nion provides better resolution compared to the Titan, due to its more advanced technology. Moreover, the atomic localization procedure implemented with the software Atomap is not fully accurate, even after the manual correction provided by the user. The wraparound lattice approximation should also be included in the possible sources of errors. Furthermore, the denoising performed by the neural network can cause some inaccuracies, especially when high levels of noise affect the images. All the described causes should be further investigated for a more comprehensive study, which also provides an estimation of the impact of each source of inaccuracy. Possible approaches to complete this study are proposed in the next section.

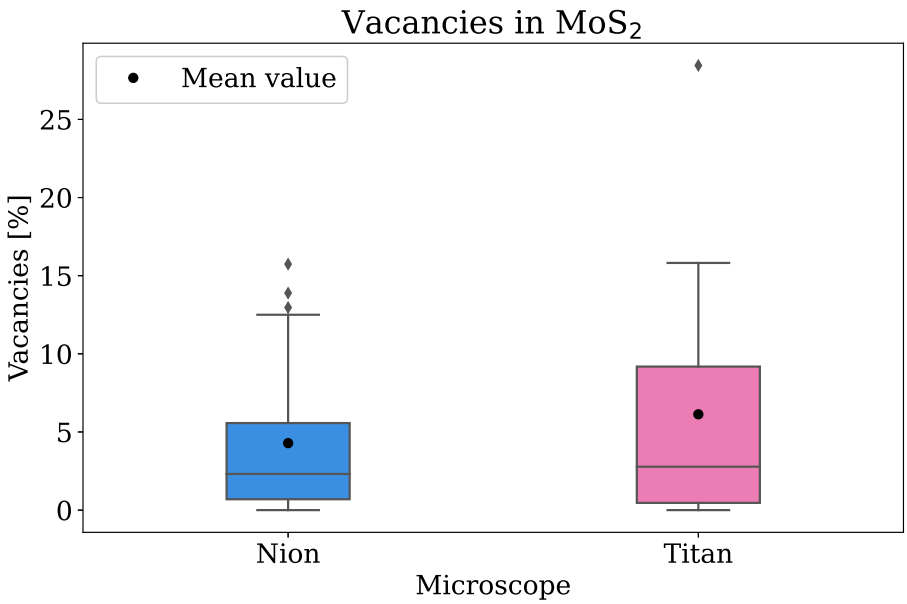

Figure 4.6 – Distribution of the percentage of vacancies found in images acquired with the Nion and Titan microscopes. The black dots show the mean of the vacancy percentage, while the gray diamond-shaped markers are commonly used in box plots to portray outliers.

4.4 Conclusions and outlook

This chapter presented some preliminary results for the investigation of vacancies in STEM-acquired images of TMD samples. As a first investigation, datasets made of MoS_2, imaged with two types of STEM microscopes were considered (Nion and Titan). In order to improve the identification of vacancies, a machine-learning approach was introduced, where a denoising autoencoder was used to enhance STEM images. These denoised images were then processed with the Atomap software to locate atomic columns of transition metal atoms (Mo) and chalcogen atoms (S). A percentage of vacancies was calculated based on the difference between the expected and observed number of chalcogen atoms for an ideal lattice.

The conducted analysis demonstrates an average percentage of vacancies of about 4 % for the Nion dataset and about 6 % for the Titan dataset. Importantly, these results do not only depend on the capability of the considered imaging instruments but it is affected by several factors, which can alter the vacancy counting task. Specifically, the neural-network-based-denoising algorithm, although effective in enhancing image quality, can introduce inaccuracies, particularly when dealing with images affected by high levels of noise. To quantify this source of error, the following approach should be implemented. A dataset of simulated images of TMD should be generated, separately from the training set construction. Noise

should be added to these data, following the same procedure used for training set preparation. Subsequently, the vacancy counting process should be performed on both the original noise-free images and the denoised data. A comparison of the amount of vacancies identified in the two cases should give an estimation of the impact of the denoising algorithm in the vacancy counting process.

Moreover, the atom localization performed by Atomap can also be a source of error. A comparison with the results achievable with a different software, such as StatSTEM should be implemented [135].

Furthermore, to account for edge effects, the wraparound lattice approximation is applied, assuming that atoms near the image borders are connected to atoms on the opposite edge. While a useful approximation, it can introduce inaccuracies, especially in cases with non-uniform edge effects. Importantly, the finite size effect should be taken into consideration. According to this phenomenon, the properties of a system are influenced by its finite size.

Future studies will also involve other TMDs, such as WS_2 and $PtSe_2$. Additionally, a comparison of the vacancies counting between samples prepared following different procedures, such as mechanical exfoliation, will be pursued. Some examples of STEM-acquired images of $PtSe_2$, before (on the top row) and after autoencoder denoising (on the bottom row) can be found in Fig. 4.7. In this case, the samples were prepared through mechanical exfoliation. The data in the first two columns were acquired with the Nion microscope, while the last two columns show data captured with the Titan microscope. The same procedure described in this chapter will be pursued for this dataset.

It is worth mentioning that at the moment the proposed methodology does not allow for distinguishing vacancies located in different layers of the material. For instance, when there are two chalcogen atoms on top of each other, the absence of a chalcogen atom on the top layer can result in reduced intensity in that area of the image. At the moment, the model is not able to identify this lower-intensity atom as a vacancy. A possible solution to this problem is to apply a threshold to distinguish vacancies located in different layers. Different strategies need to be pursued depending on the type of stacking configuration and the number of layers.

Finally, an investigation of digitally acquired data should facilitate the identification of vacancies, eased by the absence of Gaussian noise, which can hinder the atom localization process.

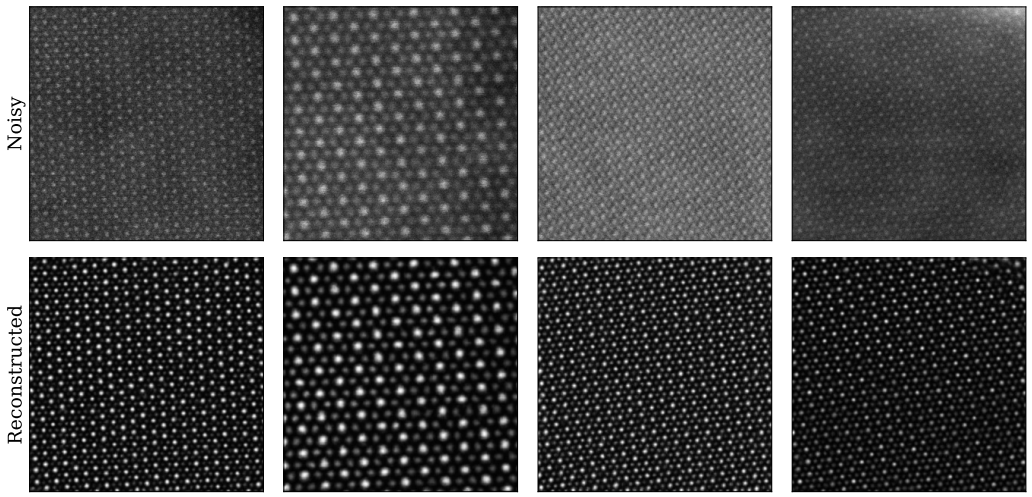

Figure 4.7 – Examples of STEM-acquired images of $PtSe_2$, before (top row) and after (bottom row) denoising them with the developed autoencoder. The first two columns display data captured with the Nion microscope, while the last two columns show Titan-acquired data. These are all examples of images that will be considered for our analysis since the amount of signal enables atom localization.

CHAPTER 5

VIDEO FRAME INTERPOLATION FOR 3D TOMOGRAPHY

The objective of this chapter is to demonstrate how a neural network developed for video frame interpolation can be used to enhance the resolution of 3D tomography data. The examined datasets are generated through diverse imaging instruments and encompass different length scale categories. This work is based on a paper that is currently under review. The first author and main contributor of the paper is the same as this Ph.D. thesis.

5.1 Problem description and state-of-the-art methods

Three-dimensional (3D) tomography refers to a collection of imaging methods used for obtaining 3D representations of the internal structure of a solid object. Generating detailed cross-sectional descriptions is a practice commonly used in many fields, including material science, medicine, and engineering. It is used for the study of specimens, where the material properties are strongly related to their internal structure.

In a general context, the first step of this practice involves a data acquisition process, where two-dimensional (2D) images or *projections* of an object are captured from different perspectives, for example, different angles or different cross-sections. The second step concerns the alignment of the acquired data, which ensures the matching of corresponding points in different images. This is fundamental to guarantee accurate execution of the following phase, namely the 3D reconstruction, which can be achieved by using different types of algorithms. Some examples

include filtered back-projection, algebraic reconstruction techniques, and iterative reconstruction methods [148, 149]. Finally, once the reconstruction is completed, the generated volume can be visualized from different angles and cross-sections.

The morphology of a sample can be investigated over many different length scales, using several imaging instruments to produce a tomographic reconstruction. For instance, in the materials science area, a technique known as FIB-SEM tomography, described in Chapter 2, is used to investigate networks of solution-processed nanomaterials, widely employed in electronics, energy, and sensing [150, 151]. For these, the analysis of their internal structure is a crucial aspect, since the charge transport is determined by the morphology of the contacts between nanosheets, which are defined at a few tens of nanometers length scale. At the opposite side of the length-scale spectrum, one finds medical imaging techniques [152], such as magnetic resonance imaging (MRI) and X-ray computed tomography (CT), where the relevant information is typically available with millimeter resolution.

It should be noted that there are several limitations to 3D tomography, common to many experimental techniques and length scales. Firstly, the resolution achieved must be sufficient for extracting information, but it is often limited by the measuring technique and the necessity to keep the acquisition time short. For instance, in a CT scan one wants to have enough details to make an informed medical decision, but limit the radiation dose the patient is exposed to [153]. Furthermore, in several cases the measurement is destructive, meaning that the specimen being imaged is destroyed during the measurement process [44, 154]. In this situation the 3D resolution is often anisotropic, meaning that cubic-voxel definition in the three dimensions is not achieved. In addition, in a destructive experiment one cannot go back and take a second measurement, should the first have not achieved enough resolution. The mentioned limitations could be overcome by using an image augmentation technique, which, by improving the quality of the available images, should make it possible to extract more accurate information.

This approach can be employed, for instance, in the case of FIB-SEM nanotomography (FIB-SEM-NT), where the term *nano* refers to the length scale involved [44]. FIB-SEM-NT is an imaging technique that involves destructive procedures, where a Focused Ion Beam (FIB) is used to gradually remove sections of a specimen, often a composite material, while concurrently capturing images of the exposed planes using a scanning electron microscope (SEM). This process generates a stack of hundreds of 2D images, which are subsequently employed to construct a high-fidelity 3D representation. However, the resulting 3D volume often exhibits anisotropic resolution, particularly when operating at high magnification. Specifically, the cross-sectional images (referred to as the xy-plane) are obtained at the SEM's native resolution, approximately 5 nm in the cases examined in this work, while the resolution along the milling direction (the z-direction) corresponds to the slice

thickness, usually around $10 - 20$ nm. Consequently, the resulting 3D volume may not feature uniform cubic voxels. It is essential to note that producing thinner slices faces limitations imposed by instrumentation, specimen characteristics, and economic considerations. Additionally, reducing slice thickness can compromise the resolution in the xy-plane due to potential damage propagation between successive cuts. This challenge is closely related to another drawback of FIB-SEM instruments, namely, their relatively slow imaging speed [53]. Therefore, there is a need for an image interpolation method that can effectively preserve and ideally enhance information quality while reducing the number of milling steps required.

The simplest solution is linear interpolation [155]. However, this is reliable only when one can safely assume that the structural variations across consecutive cross-sections are smooth. Unfortunately, when this condition is only approximately met, linear interpolation tends to blur feature edges and generate inaccurate results. This can be partially improved [156] with interpolation strategies that account for feature changes among consecutive images by using optical flow [157], but the performance remains poor at the image borders. As a consequence, such portions of the frame must be discarded, with a consequent loss of valuable information. Alternative solutions involve deep-learning algorithms. For instance, Hagita *et al.* [158] proposed a deep-learning-based method for super-resolution of 3D images with asymmetric sampling. The model was trained on images obtained from the cross-section and applied to frames co-planar to the milling direction and obtained from the 3D reconstruction. Unfortunately, this strategy works only when it is possible to assume that the three directions have the same morphology, but does not provide unbiased reconstruction. In a different effort, Dahari *et al.* [159] developed a generative adversarial network (GAN) trained on pairs of high-resolution 2D images and low-resolution 3D data, aiming at generating a super-resolved 3D volume. The scheme showed success on a variety of datasets. However, generative models are ambiguous to use in this context, since they do not allow one to find a unique solution, due to the nature of this deep-learning architecture.

In this work, an alternative solution is proposed, which relies on the use of a deep-learning model trained for video-frame interpolation. This is a process, where the frame per second of a video is increased by generating additional frames between the existing ones, thus creating a more visually fluid motion [64]. As described in Chapter 2, among several deep-learning frameworks developed for this task, one of the most advanced is the Real-Time Intermediate Flow Estimation (RIFE) [67], which will be extensively used in this research project. Compared to other methods, this model can achieve highly accurate results at a significant speed. This is a consequence of the coarse-to-fine approach with increased resolution employed for the inference of the intermediate flow between existing frames.

The main objective of this work is to show how (RIFE) can be used for purposes that are far from the original aim for which it was developed. To do so, datasets obtained with different imaging instruments, that generate data characterized by different length scales, are considered. The first investigated dataset is made of printed graphene-nanosheet images, obtained with FIB-SEM, where the milling direction is taken as equivalent to the video time direction. The resolution of this dataset is then improved by the application of RIFE, and quantitatively validated using several approaches. In particular, together with standard computer-vision metrics, physical quantities are evaluated, which can be extracted from the final 3D reconstructions, after appropriate image binarization with standard software such as FIJI [45] or DRAGONFLY [46]. These are the porosity, tortuosity, and effective diffusivity, and their precise evaluation facilitates an understanding of what information content is preserved/improved during the interpolation. Furthermore, the proposed scheme is benchmarked against another flow-based deep-learning algorithm, DAIN [69], and against non-deep-learning methods. In particular, the examined approaches include the simple but widely used linear interpolation and the IsoFlow algorithm [156], a novel interpolation technique that takes into consideration the variation among slices by using optical flow [157]. Then, the same scheme is applied, at a completely different length scale, to both MRI and CT scans. In the first case, the 3D mapping is already isotropic, so that the reconstructed images can be compared to an available ground truth, as in the case of FIB-SEM. Instead, for CT scans, no ground truth can be used to validate the results. Therefore, a different approach, based on noise power spectrum evaluation, is used to show a significant enhancement of the picture quality. This result may enable to reduce the scanning rate and therefore the radiation dose for the patient. Additionally, the application of RIFE on coronary angiography data is investigated. In this instance, the data comprise actual videos rather than 3D tomographic reconstructions. The aim of this section is to explore a potential integration of video frame interpolation techniques in this medical procedure, live. This could allow for a reduction of radiation exposure during the assessment of the cardiovascular system, both for the patient and the medical practitioner.

It is worth mentioning that, in the context of 3D tomographic reconstruction, some neural networks have been developed to generate detailed 3D representations from a reduced number of 2D projections [160, 161]. However, for the applications investigated in this project, the goal is not only to obtain an enhanced 3D reconstruction but also to generate additional frames that can be analysed to retrieve important information about the examined systems. For instance, in the case of FIB-SEM-generated data, the individual 2D frames are used to study material-related properties, such as network porosity. For this reason, this work will focus on approaches capable of generating additional frames that can be used

to improve 3D reconstructions rather than algorithms that facilitate and enhance the reconstruction process.

5.1.1 Neural Network

RIFE [67] is the neural network chosen to augment the resolution of the tomographic reconstruction described in the introductory section of this chapter. This is a video-frame interpolation technique belonging to the flow-based category, as detailed in Chapter 2. It allows determining the flow between corresponding elements in consecutive frames with remarkably high accuracy and speed. For the purpose of this research project, some adjustments were made to the original code, mainly regarding the data upload and saving process, in order to make it more suitable for the examined dataset. Apart from this, the original architecture was not altered.

Like any extensive machine-learning model, RIFE undergoes regular updates. As of now, the HD model v4.6, referred to as RIFE HD [72], is the most advanced version available and it is trained on the Vimeo90K dataset. This dataset covers a wide range of scenarios, including various activities involving people, animals, and objects [73]. Despite being trained on data from significantly different contexts, RIFE HD allows for achieving valuable performances on the datasets studied in this research project, as will be detailed in this chapter. Nonetheless, some fine-tuning of the pre-trained models was also implemented. It is worth noting that fine-tuning the RIFE HD model is currently not feasible. Consequently, the fine-tuning was performed on the second-best model, namely RIFE$_m$ [74]. To avoid overfitting[1], the fine-tuning procedure should be realized on data that is not used for testing. This is available only for the application to printed graphene networks, described in the next sections. Therefore, fine-tuning was avoided for the medical applications. Specifically, the fine-tuning is performed on a subset of the graphene dataset, made of 1,000 portions of the original images, cropped to a 510×510-pixel size, and not used for testing. The results of both the original and the fine-tuned model are presented in the following section and compared throughout. It is worth mentioning that, despite being commonly used to enhance the model performance for specific tasks and datasets, fine-tuning can be counterproductive from a practical point of view, especially in the medical context. In fact, it would imply the need for relatively large datasets that must adhere to ethical and privacy regulations, in terms of training and sharing usage. Moreover, data might present different features depending on the manufacturer of the instruments used to generate them, hindering the possibility of developing a comprehensive and

[1]The term overfitting refers to a phenomenon that occurs when a machine-learning model nearly memorizes the training data, without recognizing underlying patterns. As a consequence, the model will have satisfactory performance on the training set but it will not be able to generalize to unseen data.

exhaustive model.

5.2 Application to Printed Graphene Network dataset

The main dataset used in this work is made of 801 images, generated with FIB-SEM, of printed nanostructured graphene networks, with a nanosheet length of approximately 700 nm. Each image, made of 4041×510 pixels, has a 5 nm resolution in the cross-section, while the slice thickness is 15 nm. Therefore, the voxel size in the resulting reconstructed volume is $5 \times 5 \times 15 = 375$ nm^3. Note that the voxel size achievable with conventional micro CT scanners is $10 - 1000$ times larger [162, 163]. Therefore, FIB-SEM nanotomography is more suitable than CT for the quantitative characterization of the graphene network morphology, which highly affects the material's properties, such as the network connectivity [164]. As previously mentioned, these graphene networks belong to the class of networks of solution-processed nanomaterials, which are currently being investigated for applications in many domains, including electronics, energy, catalysis, and sensing [150, 151, 165, 166]. The FIB-SEM nanotomography was performed by Cian Gabbett and Luke Doolan at Trinity College Dublin, on a ZEISS ATLAS 5 software (Version 5.3.3.31). The FIB was operated with a Gallium ion beam, at a voltage of 30 kV and current of 600 pA. Further details on sample preparation and data acquisition can be found in reference [44].

In order to prove the method's efficacy, some frames are removed from the dataset and used as ground truth for results assessment. Different scenarios are considered, where one, three, and seven consecutive frames are removed, respectively, although the seven-frame error suggests that it is not advisable to reduce so drastically the image density along the milling direction. A fraction of the original dataset is considered for the majority of the analysis, to reduce the computational effort and to inspect the images in more detail. Specifically, for the computer-vision metrics and porosity analysis, 100 images of the original dataset are considered. Each image of this subset is cut to a 510×510 pixels size. In contrast, for the tortuosity and effective diffusivity study, ten randomly selected volumes are considered, ranging from 55 % to 60 % of the original volume. It should be noted that in all cases the resolution is not altered.

5.2.1 Qualitative assessment of the results

A simple visual comparison offers a qualitative overview of the efficacy of the various interpolation methods examined. This is shown in Fig. 5.1 for the case where three consecutive frames are removed from the FIB-SEM sequence and then reconstructed by the different models. The first column shows the ground-truth

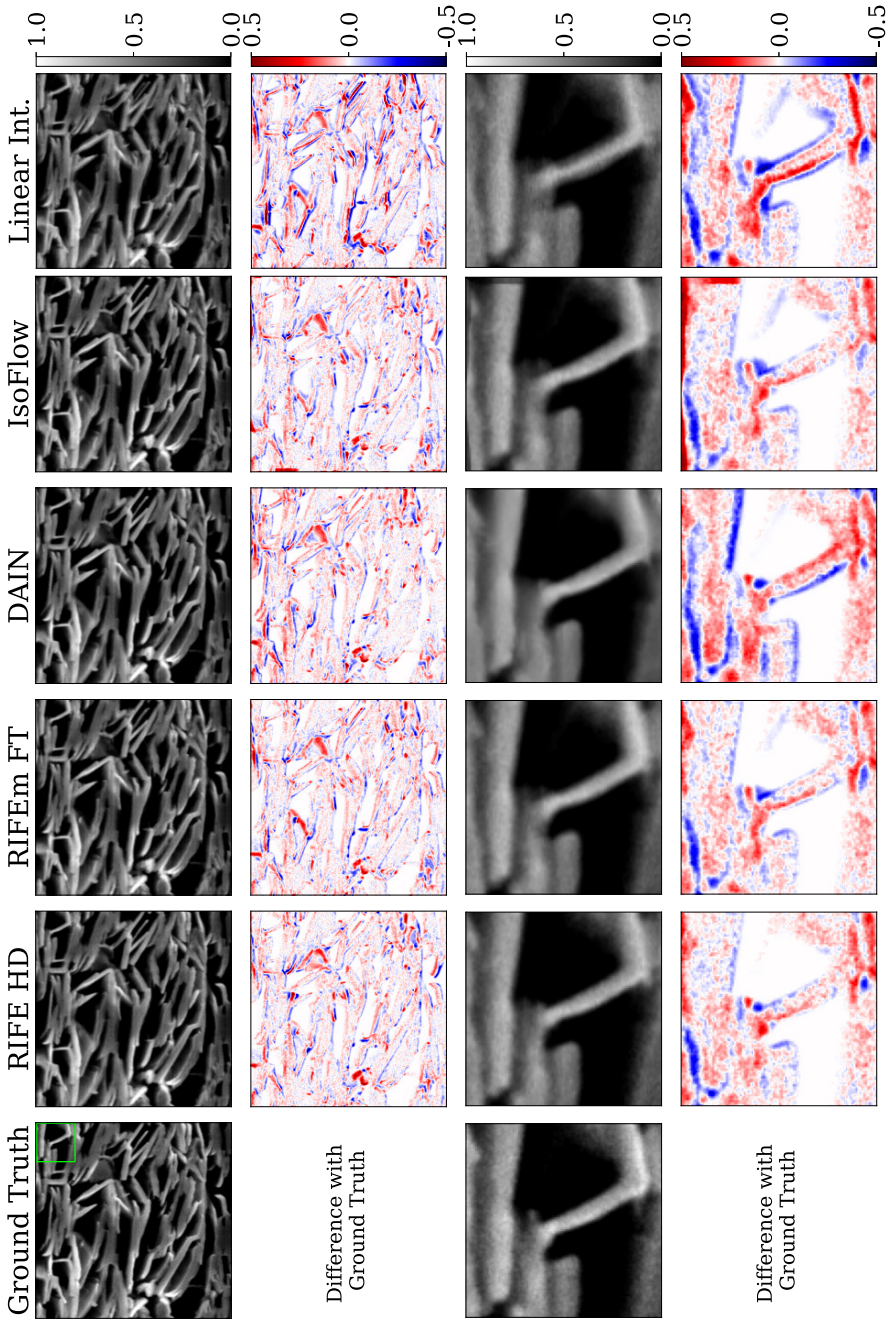

Figure 5.1 – Visual comparison of the frame reconstruction in the case where three frames are removed from the FIB-SEM sequence. From left to right it is shown: the ground truth image (original FIB-SEM), and those reconstructed by RIFE HD, the fine-tuned RIFE$_m$, DAIN, IsoFlow, and linear interpolation. The second row displays the difference between the ground truth and the reconstructions. A 100×100-pixel portion of each image (see green box in the upper left panel) is magnified and shown in the third row, while the differences from the original image are in the fourth row.

image, namely that removed from the original dataset, while the remaining ones contain the pictures reconstructed with the various methods. In order to better appreciate the quality of the reconstructions, additional pieces of evidence are provided, such as the difference between the ground truth and the reconstructed images (second row), the magnification of a 100×100-pixel portion of each picture (third row), and again the difference from their ground truth (fourth row). The differences are obtained by simply subtracting the grayscale bitmap of each pixel. Blue (red) regions mean that the reconstructed image appears lighter (darker) than the original one.

The inspection of the figure leads to some qualitative considerations on the different methods, and the comparison is particularly clear for the magnified images. The most notable feature is the loss of sharpness brought by the linear interpolation, which is not motion-aware. In fact, instead of tracing the motion of the border between a graphene nanosheet and a pore, namely the border region between dark and bright pixels, linear interpolation simply fills the space with an average grayscale. As a result, the image difference (e.g. see the rightmost lower panel) presents some dipolar distribution, which, as will shown below, causes information loss. A similar, although less pronounced, drawback is found for images reconstructed by DAIN, which also tends to over-smooth the graphene features. In contrast, IsoFlow appears to generate generally good-quality pictures, in particular in the middle of the frame. However, one can clearly notice a significant error appearing at the image border, which is not well reproduced and whose information thus needs to be discarded. Finally, the two RIFE models are clearly the best-performing ones. Of similar quality, they are able to maintain the original image sharpness across the entire field of view and do not seem to show any systematic failure.

Analogous illustrations can be found in Fig. 5.2 and Fig. 5.3, with examples of results for the one- and seven-replaced-frames cases. It is worth mentioning that the colourbars for the difference plots are not the same for the three proposed figures, due to different levels of inaccuracy of the various cases. A visual inspection of the results confirms the considerations made for the three-replaced-frames case, with the RIFE schemes being the better performing. Moreover, these figures seem to demonstrate that the capabilities of each model deteriorate when increasing the number of replaced frames. In fact, the difference plots of Fig. 5.3 present higher error values compared to the other cases, which indicated the presence of more inaccurate results when seven frames are replaced. This is an expected behaviour, which will be further investigated in the following section. In fact, although instructive, visual inspection just provides a qualitative understanding, and more quantitative metrics need to be evaluated in order to determine what image content is preserved by the various reconstructions.

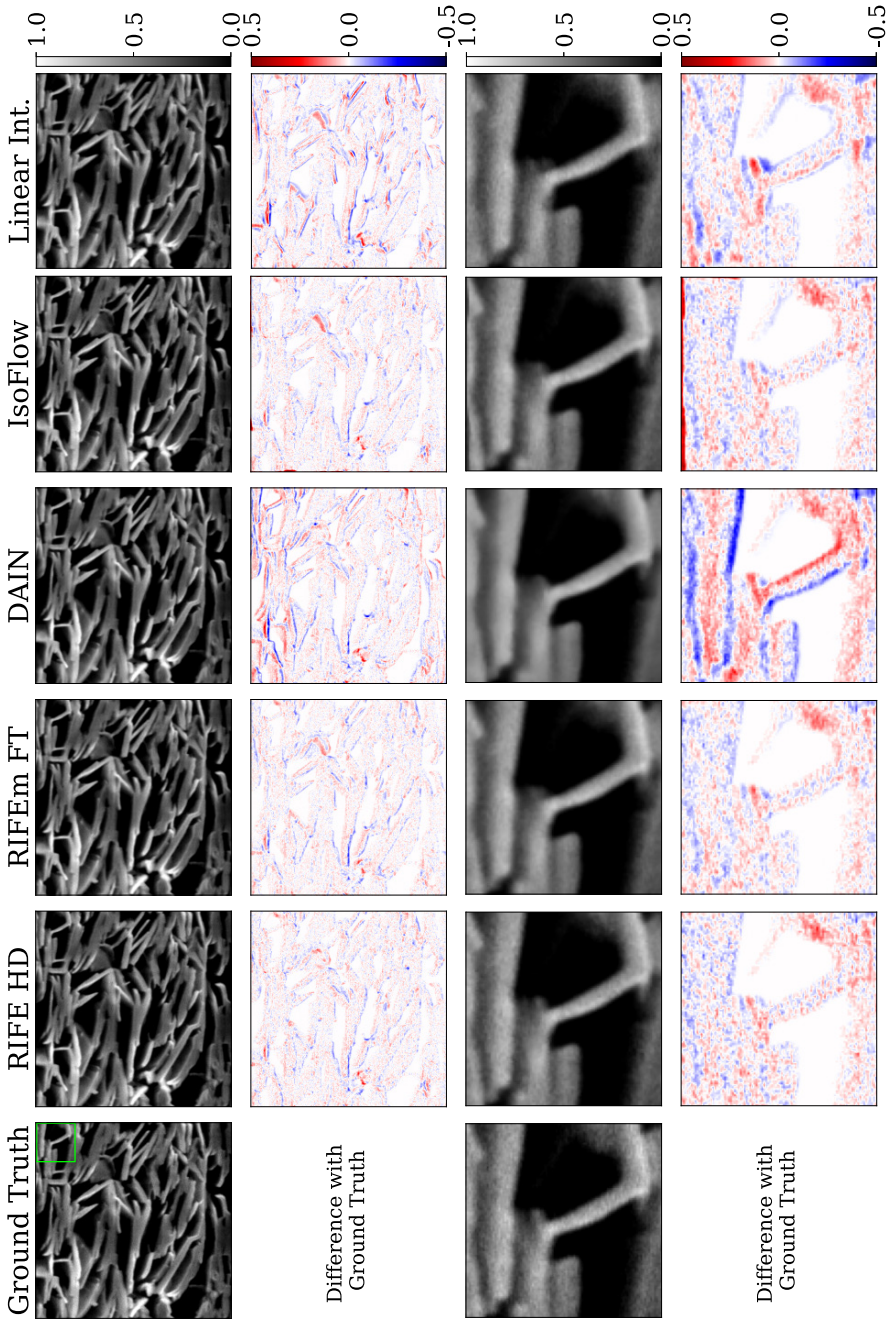

Figure 5.2 – Visual comparison of the frame reconstruction in the case where one frame is removed from the FIB-SEM sequence. From left to right it is shown: the ground truth image (original FIB-SEM), and those reconstructed by RIFE HD, the fine-tuned RIFE$_m$, DAIN, IsoFlow, and linear interpolation. The second row displays the difference between the ground truth and the reconstructions. A 100×100-pixel portion of each image (see green box in the upper left panel) is magnified and shown in the third row, while the differences from the original image are in the fourth row.

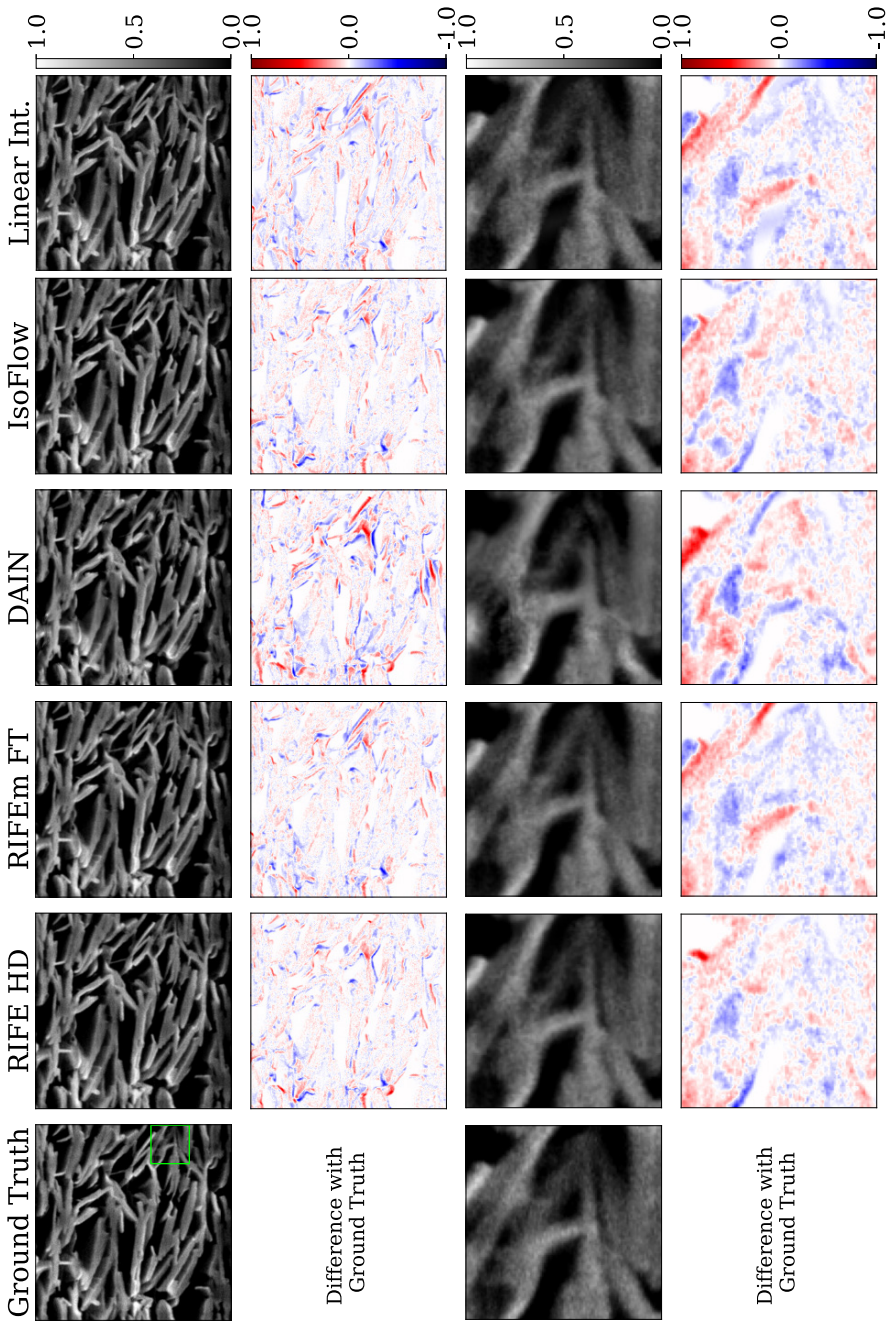

Figure 5.3 – Visual comparison of the frame reconstruction in the case where seven frames are removed from the FIB-SEM sequence. From left to right it is shown: the ground truth image (original FIB-SEM), and those reconstructed by RIFE HD, the fine-tuned RIFE$_m$, DAIN, IsoFlow, and linear interpolation. The second row displays the difference between the ground truth and the reconstructions. A 100×100-pixel portion of each image (see green box in the upper left panel) is magnified and shown in the third row, while the differences from the original image are in the fourth row.

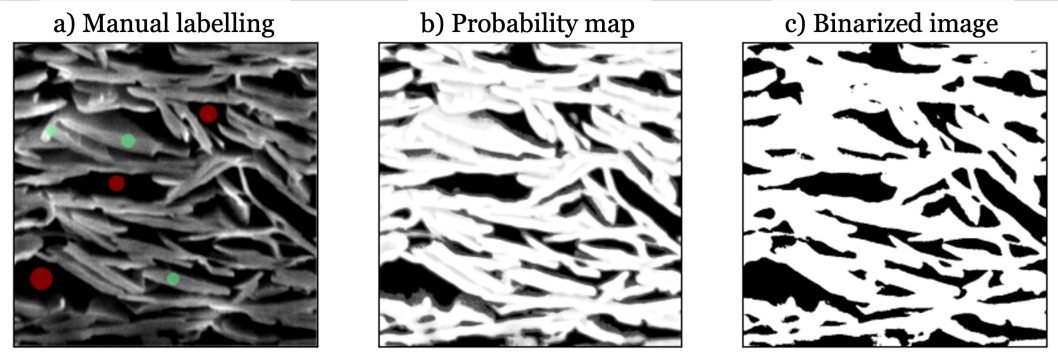

Figure 5.4 – Different phases of the procedure used to segment each frame into pore and nanosheet components. This generates binarized data, by using the trainable WEKA segmentation plugin of FIJI [167, 44]. In panel (a) the user first manually assigns the label of pore (red circles) and nanosheet (green circles) to some areas of the original image. This information is used to build the dataset for training a classifier, whose output is the probability map displayed in panel (b). Here the probability of each pixel being either pore or nanosheet is displayed using pixel intensity values. Panel (c) shows the final output of the procedure, namely the binarized image, obtained by applying a threshold on the probability map.

5.2.2 Quantitative assessment of the results

A full quantitative analysis is better performed on segmented images, where the pore and nanosheet components are well separated [44]. This can be obtained by using the trainable WEKA segmentation tool [167] available in FIJI [45]. The procedure to produce binarized data is demonstrated in Fig. 5.4. A set of images from the original dataset is used to train a model, whose goal is to classify each pixel of the image either as pore or as nanosheet. The training set is automatically built by WEKA following the manual identification of pore and nanosheet areas from the user. This is shown in Fig. 5.4(a), where the red circles identify pixels that are labelled as pores and the green circles represent pixels labelled as nanosheets. Panel (b) of the same figure displays the outcome of the application of the trained model, where the grayscale expresses the probability of each pixel being labelled as pore or nanosheet. This is called a probability map. Finally, a threshold is applied to obtain a binary classification, as shown in Fig. 5.4(c). In particular, here the ISODATA algorithm [168], available in FIJI, is used to select an appropriate threshold. Once the classifier is trained and the threshold is established, all the datasets obtained from the different interpolation strategies can be binarized. It should be noted that the performance of the classifier depends on the manual selection performed by the user. However, using the same classifier for all the analyzed datasets guarantees consistency in the segmentation process and, consequently, in the quantitative assessment of the various reconstruction methods.

The Mean Square Error (MSE) and the Structure Similarity Index Method (SSIM) are some of the standard metrics used in computer vision to evaluate results [169],

as discussed in Chapter 2. Both are full-reference metrics, meaning that the ground truth is required to assess their value. The MSE focuses on the pixel-by-pixel comparison and not on the structure of the image, while SSIM performs better in discriminating the structural information of the frames. Here the MSE is calculated between each of the generated frames and the corresponding image removed from the original dataset. The average of these values over 100 frames is then computed for each case of study (one, three, and seven replaced frames) and for each technique (RIFE HD, RIFE$_m$, DAIN, ISOFLOW and linear interpolation). The same procedure is followed for the evaluation of the SSIM and the results are available in Fig. 5.5. As expected, all models perform better when the number of removed frames remains limited, and in general there is a significant loss in performance for the case of seven replaced frames. In more detail, RIFE-type schemes are always the top performer, with linear interpolation, and also DAIN, remaining the most problematic. Interestingly, ISOFLOW appears quite accurate according to these computer-vision metrics, which clearly do not emphasize much the loss of resolution at the frame boundary. However, it will now be demonstrated that this does not necessarily translate into the ability to preserve information.

Ultimately, the quality of a reconstruction procedure must be measured with the quality of the information that is able to transfer/retrieve. In the case of printed graphene-nanosheet ensembles, some morphological properties can be measured and compared. The so-called network porosity, P, defined as the percentage of pores contained in each frame, is one of the most important features measured in 2D networks and it affects the material electrical properties [170]. This can be evaluated from the binarized images by the conventional image-processing software FIJI [45], and such analysis is performed here for each case of study and technique. The results are then expressed in terms of delta porosity, ΔP, which is the fractional difference between the porosity computed from images reconstructed with a particular method m, P_m, and that of the ground truth, P_{GT}, namely

$$\Delta P = 100 * \frac{\mid P_m - P_{\mathrm{GT}} \mid}{P_{\mathrm{GT}}} \,. \tag{5.1}$$

For this study, the ΔP of each image is computed and the average of these values is presented in Fig. 5.6. From the figure, the advantage of using RIFE is quite clear. In fact, although all methods, except for the linear interpolation, give a faithful approximation of P when one frame is removed from the sequence, differences start to emerge already at three replaced frames, where RIFE significantly outperforms all other schemes. The difference becomes even more evident for seven replaced frames, for which the RIFE error remains below 2 %. Also, it is interesting to note that, in contrast to what is suggested by the computer-vision metrics, ISOFLOW is not capable of accurately returning a precise porosity, mainly because of the

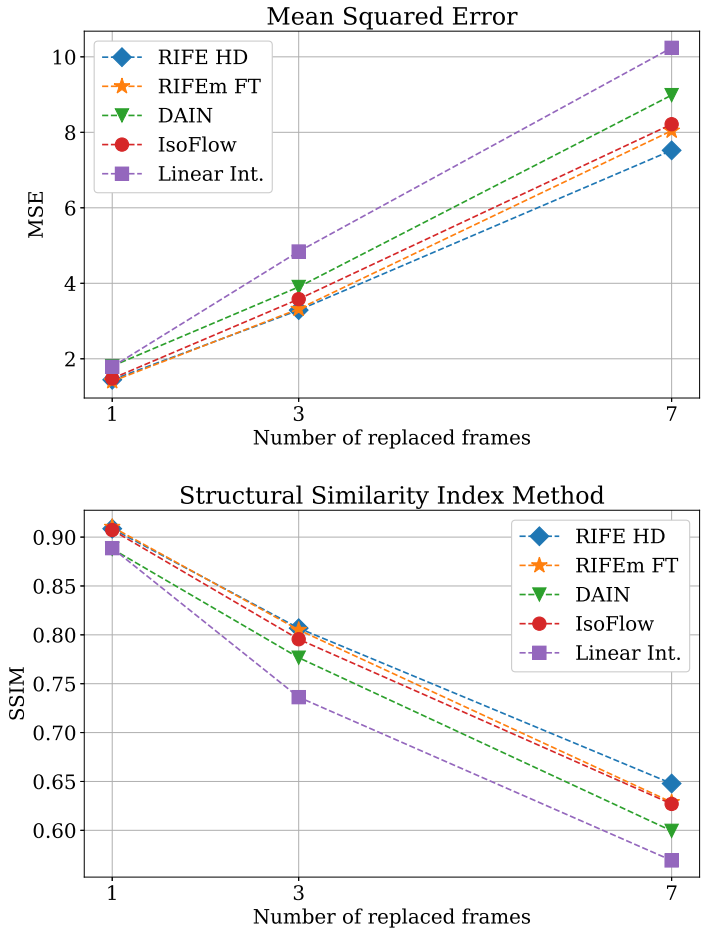

Figure 5.5 – Mean Squared Error (MSE - top panel) and Structural Similarity Index Method (SSIM - bottom panel) evaluated for each test case (one, three, and seven replaced frames) and each interpolation method. The MSE and SSIM are evaluated for each frame against the ground truth and they are expressed as an average over 100 frames.

poor description of the image borders. In contrast, the weak performance of linear interpolation has to be attributed to its inability to describe sharp borders between nanosheets and pores. In fact, using this method corresponds to performing a simple averaging between frames. It is worth emphasizing that using the proposed expression for the porosity, it is not possible to evince biases in the deviation from the ground truth values. In fact, the absolute value prevents the identification of systematically higher or smaller results for the investigated methods. However, as no discernible trends in the results were found, the absolute value was chosen for representation to achieve clearer visualization. This consideration applies also to the other metrics investigated in this section.

A second important structural feature that can be retrieved from nanostructured networks is the network tortuosity, τ, which can be evaluated using the TauFactor software [171]. This quantity describes the effect that a convolution in the geometry

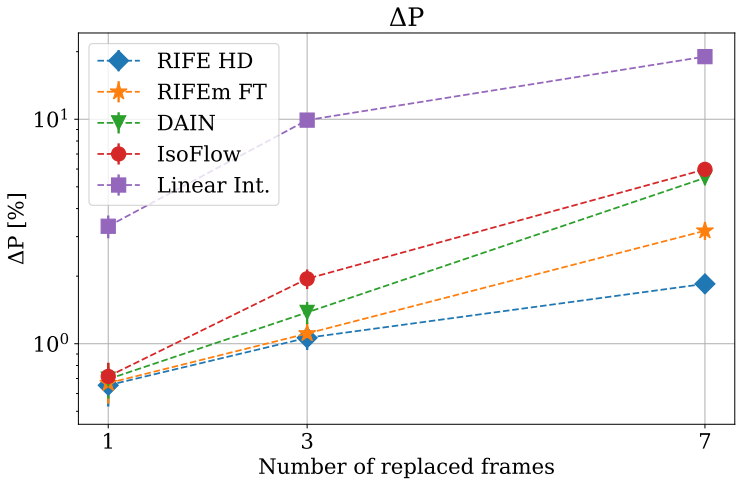

Figure 5.6 – Porosity, P, evaluated for each test case (one, three, and seven replaced frames) and each interpolation method. The porosity is evaluated for each frame against the ground truth and it is expressed in terms of ΔP [see Eq. (5.1)], namely as a percentage deviation from the ground-truth value. The image displays the average ΔP over the test set and the associated variance.

of heterogeneous media has on diffusive transport and can be measured for both the nanosheet and pore volumes. The nanosheet network tortuosity factor influences charge transport through the film. Pore tortuosity affects performance in nanosheet-based battery electrodes, while in gas sensing applications the pore tortuosity is directly linked to gas diffusion. The tortuosity, τ, and volume fraction, ε, of a phase are used to relate the reduction in the diffusive flux through that phase by comparing its effective diffusivity, D^{eff}, to the intrinsic diffusivity, D:

$$D^{\text{eff}} = D\frac{\varepsilon}{\tau}\,. \tag{5.2}$$

It has been proved that for the evaluation of the tortuosity and diffusivity, the sample volume needs to be adequately large, in order to be representative of the bulk and to reduce the effect of microscopic heterogeneities [172]. For this reason, ten randomly selected volumes are considered, ranging from 55 % to 60 % of the original one. As the size of the input sample highly affects the computation speed and memory requirement, not all methods are considered for this comparison. In particular, only RIFE HD and DAIN results are used as input for the tortuosity and diffusivity study. These two methods are chosen since they provide the best evaluation of the porosity. Then, the Python version of TauFactor [173] is run on Quadro RTX 8000 GPUs.

Also for this analysis, the fractional change of any given quantity from the ground truth is computed, and the results are displayed in Fig. 5.7. Confirming the results obtained for ΔP, also in this case RIFE is the best-performing method,

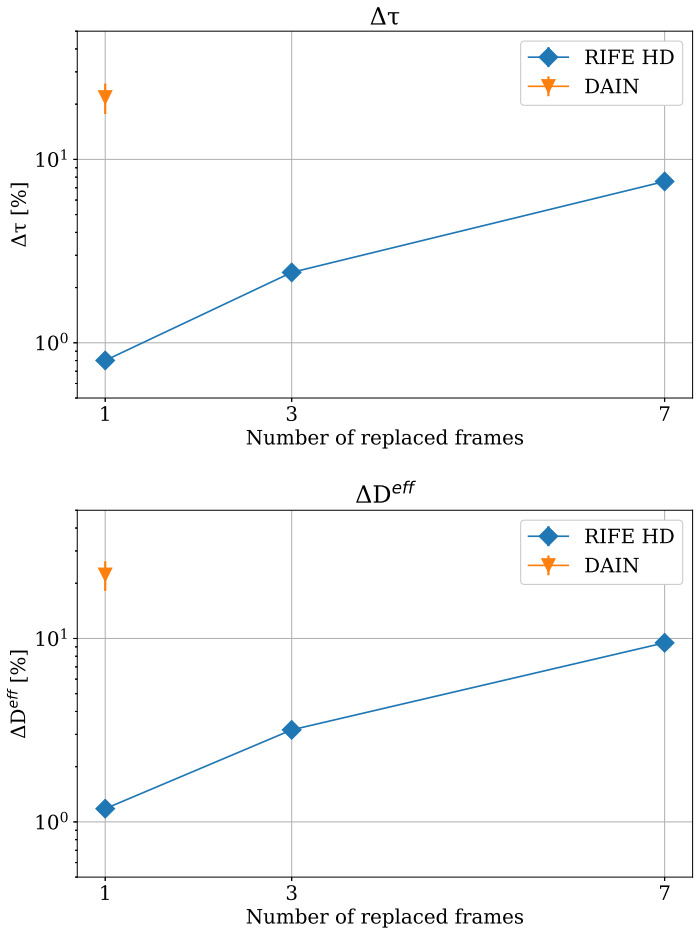

Figure 5.7 – Delta tortuosity, $\Delta\tau$ (top side panel), and effective diffusivity, ΔD^{eff} (bottom panel), evaluated for each test case (one, three, and seven replaced frames) for RIFE HD, and for one replaced frame for the DAIN interpolation. The metrics are evaluated for each frame against the ground truth and averaged over the sequence. The variance is also displayed, and in the case of RIFE it is smaller than the symbols' size.

with errors remaining below 2 % at three replaced frames for both τ and D^{eff}. In contrast, DAIN displays significant errors, exceeding 10 %, already for a single replaced frame, an error that suggested analysis at other replaced-frame rates was unnecessary.

Dependence on the features size

As demonstrated in the previous section, RIFE performs well at reconstructing the replaced frames of the FIB-SEM sequence, with errors on physically relevant quantities remaining below 10 % even for seven replaced frames. This is equivalent to having a milling thickness of about 100 nm, indeed a very favourable experimental condition. In this section, the limitations of the proposed approach in relation to the type of sample are discussed. A relevant problem with image interpolation

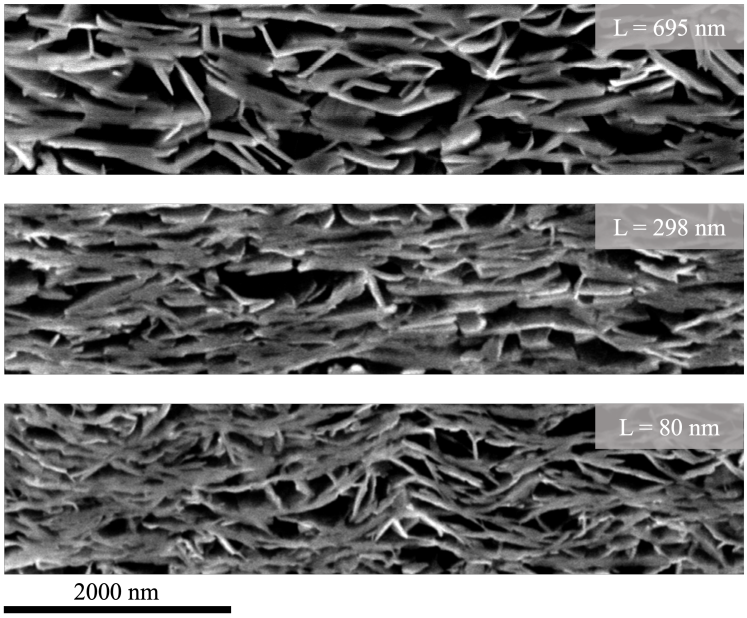

Figure 5.8 – Cross-sections of printed graphene nanosheets of different lengths (L). The nanosheet length decreases going from the top panel to the bottom one. In each case the image width shown is 6510 nm.

techniques concerns the level of continuity between consecutive frames, as detailed in Chapter 2. In fact, it is well understood that rapid changes between the images in a sequence can reduce the quality of the interpolated frames [64]. The same issue may arise when considering FIB-SEM measurements of graphene nanosheets of different lengths. In this case, shorter nanosheets will result in FIB-SEM images with more abrupt changes between consecutive frames. For instance, if the average nanosheet length is L and the milling distance L', for $L \sim L'$ one will encounter often the situation where a nanosheet present in one image will not appear in the next one.

In order to explore how the proposed model works with increasingly challenging datasets, the cases of networks made of shorter graphene nanosheets are investigated, specifically those with average lengths of 80 nm and 298 nm. Examples of such networks, together with the original one of 695 nm can be found in Fig. 5.8. It should be noted that all the images displayed in the figure have the same width, namely 6510 nm. This clarification is necessary to confirm that they are generated using different samples, and are not different magnifications of the same image. In this case, ΔP is used as an evaluation criterion for three replaced frames case, together with the original RIFE HD model. The results can be found in Fig. 5.9, where ΔP is shown against the average nanosheet length. For this comparison, the data are binarized following the procedure described in the previous section and ΔP is computed as an average over 100 images. It is evident from the figure, as expected, that the performance of the proposed model

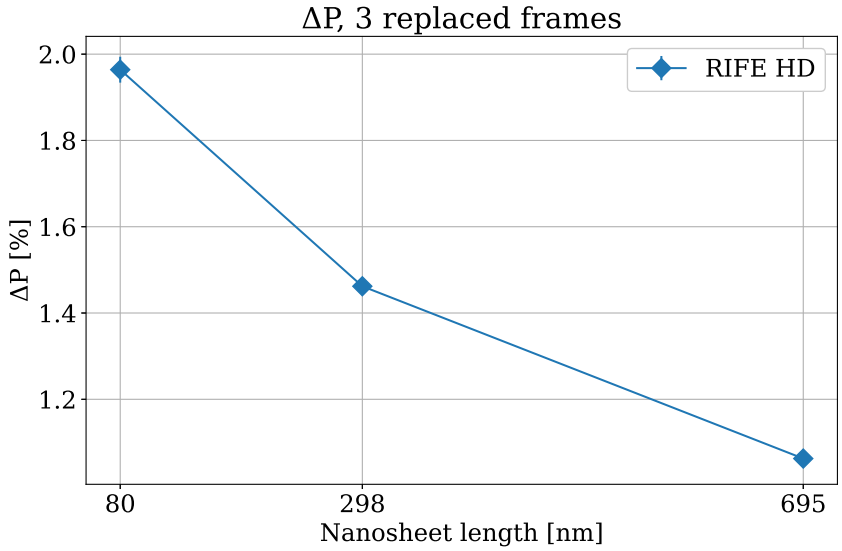

Figure 5.9 – Porosity evaluated for the three replaced frames case, for different nanosheet lengths (80 nm, 298 nm and 695 nm). The porosity is evaluated for each frame generated by RIFE HD against the ground truth and it is expressed in terms of ΔP [see Eq. (5.1)]. The average ΔP over 100 images and the associated variance is shown.

indeed deteriorates when reducing the nanosheet's length. However, even for the smallest sample, 80 nm, the error remains below a very acceptable 2 %. Note that in these conditions (nanosheet length 80 nm and three replaced frames, equivalent to ∼50 nm milling distance) the milling distance is about half of the average feature size of the sample. Since networks made of small flakes are certainly structurally more fragile than those made with larger ones, the fact that the milling frequency can be reduced significantly without a sensible loss in the accuracy of the morphology determination, establishes a possible new experimental condition, where the milling effects on the final morphology are strongly minimized.

5.3 Application to 3D medical datasets

The strategy proposed in this work is not limited to FIB-SEM generated data, but can be employed to increase the frame rate of datasets of materials at different scales, obtained by using different imaging instruments. The purpose of this section is to show such transfer across scales, and for this reason, no additional training or fine-tuning of the original model is performed. Being the original RIFE HD version the only one considered in this section, it will be referred to as RIFE from now on.

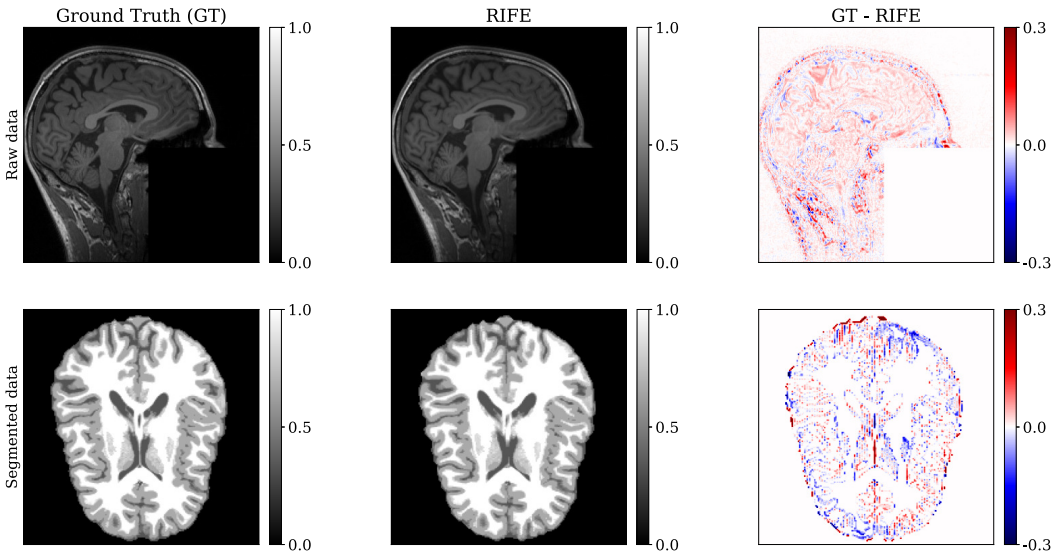

Figure 5.10 – Application of RIFE to a MRI dataset. An example of the original and the corresponding RIFE-generated frame in the sagittal view is shown in the first and second top panels, respectively. The third panel displays the intensity difference between the ground truth and the RIFE-generated images. The two datasets are segmented by using the Anatomical pipeline of the BRAINSUITE software, as displayed in the second row. The original and corresponding RIFE-generated data are shown in the first and second panels, respectively, while the third panel presents the difference between them. Although the results are here presented in axial view only, the segmentation is performed on the full volume.

5.3.1 MRI scans

The first example is an application of RIFE to human brain MRI scans. For this dataset, the voxels of the reconstructed volume are already cubic with a 1 mm^3 resolution. However, this is a useful case study, since it is possible to remove frames from the scan sequence and use them as ground truth for the validation, as in the case of FIB-SEM. In particular, every other frame is removed from the original dataset, which has been downloaded from the Brainstorm repository [174, 175]. This is well documented and freely available online for download under the GNU general public license. For this study, brain scans in the sagittal view are considered as input for RIFE.

A visual comparison of one original and the corresponding RIFE-generated slice in the sagittal view is shown in the first row of Fig. 5.10, where the patient was defaced to fulfil privacy requirements. The third column of the mentioned figure shows the difference between the original and the generated image. Clearly, the visual inspection of the reconstructed image appears very positive with a difference from the ground truth (see top right panel), which presents an error similar to that of the FIB-SEM data, despite the rather different length scale, and little structure in its distribution. The BRAINSUITE software [176] is then used to quantitatively

compare the original and the RIFE-generated volumes, again trying to understand whether the information content is preserved. One of the main components of the BRAINSUITE software is the Anatomical pipeline, which allows one to retrieve cortical surface models from MRI data. Moreover, it conducts surface-constrained volumetric registration (i.e. the process of aligning or matching different images or datasets so that they are in a common coordinate system or spatial relationship), aligning the MRI with a labelled anatomical atlas. This alignment facilitates the delineation of anatomical regions of interest within the MRI brain volume and on the cortical surface models. In this context, the Anatomical pipeline is performed on both stacks to obtain the full brain segmentation, whose output is shown in the axial view in the first two panels of the second row of Fig. 5.10. Also in this case, the visual inspection is similar to that made for the original images, with similar error characteristics. These segmentations are then used to evaluate the gray-matter volume variation (GMV), a widely used metric for the investigation of brain disorders such as Alzheimer's disease [18, 177]. This is defined as,

$$\mathrm{GMV} = \frac{p_{\mathrm{gray}}}{p_{\mathrm{gray}} + p_{\mathrm{white}}}\,, \tag{5.3}$$

where p_{gray} and p_{white} indicate the number of gray and white pixels, respectively. The percentage error between the GMV of the two datasets is computed to be 0.5 %, again very low.

5.3.2 CT scans

The second medical application investigated here refers to CT scans. Note that this measurement technique is not limited to the medical space, but it is also widely used in industrial settings and, in general, as a research tool across materials science [88]. The use of interpolation methods for medical CT could be really transformative since a reduction of the collected frames translates into a reduction of the radiation dose delivered to the patient. As a consequence, the potential risk of radiation-induced cancer will diminish [19]. Alternatively, one may have the possibility to perform more frequent scans for close monitoring of particular diseases.

The dataset used for this investigation is downloaded from the Cancer Imaging Archive [178, 179] and is provided as a set of 152 frames in the axial view, with a pixel size of $0.74 \times 0.74 \times 2.49$ mm. For this example, the voxels in the reconstructed volume are not cubic and no ground truth is available. RIFE is then used to generate three additional frames between every two existing ones.

The result can be seen in Fig. 5.11, Fig. 5.12, and Fig. 5.13. Fig. 5.11 show the data in axial view. The panel on the left-hand side show one of the original frames of the CT dataset, while the right-hand-side panel presents a frame generated

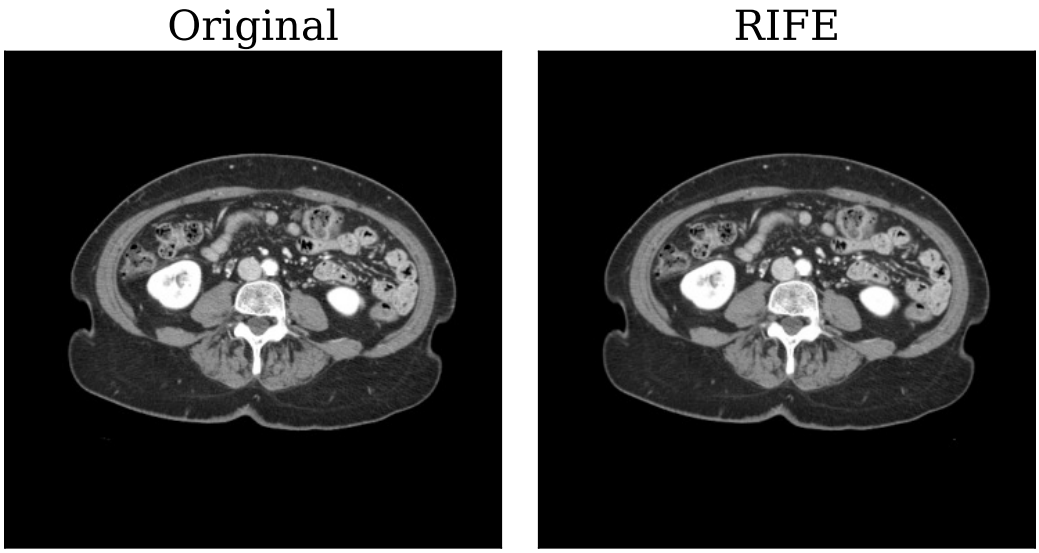

Figure 5.11 – Application of RIFE to an X-ray Computed Tomography dataset. Here the results are presented in the axial view, namely the view in which the original sequence, used as input for RIFE, is depicted. On the left-hand-side panel, there is one of the original frames from the CT dataset, while the right-hand-side panel displays a frame created using RIFE, with the former image serving as an input for the interpolation model. It is essential to emphasize that these two panels depict different sections of the body, allowing for only a qualitative comparison. This figure enables us to discern that RIFE effectively produces reasonable representations of CT scans in the axial view.

with RIFE, using the former image as an input for the interpolation model. It is important to note that in this case, the two panels do not represent the same section of the body, therefore they can only be compared from a qualitative point of view. From this figure, we can only observe that RIFE is able to generate realistic representations of CT scans in the axial view. Fig. 5.12 displays the original and RIFE-augmented datasets in coronal view, on the left- and right-hand-side panels, respectively. The original-sized data are shown in the first row, while the second row presents 150×150-pixels magnified portions of the data (indicated by green boxes in the first row). The red and blue boxes are used for the noise power spectrum analysis that will be described shortly. For the coronal view, the comparative figure helps assess the improvement introduced by RIFE. This visual comparison appears quite favourable, with the RIFE reconstruction being in general smoother than the original image. This is, of course, the result of having more frames added to the sequence. A similar comparison is proposed in Fig. 5.13, for the sagittal view. Also in this case, and especially from the magnified portion displayed in the second row, it is evident that the RIFE-augmented dataset presents a more natural appearance of the body features.

Since no ground truth is available in this case, a different approach is needed to provide a quantitative assessment of the interpolation procedure. Therefore,

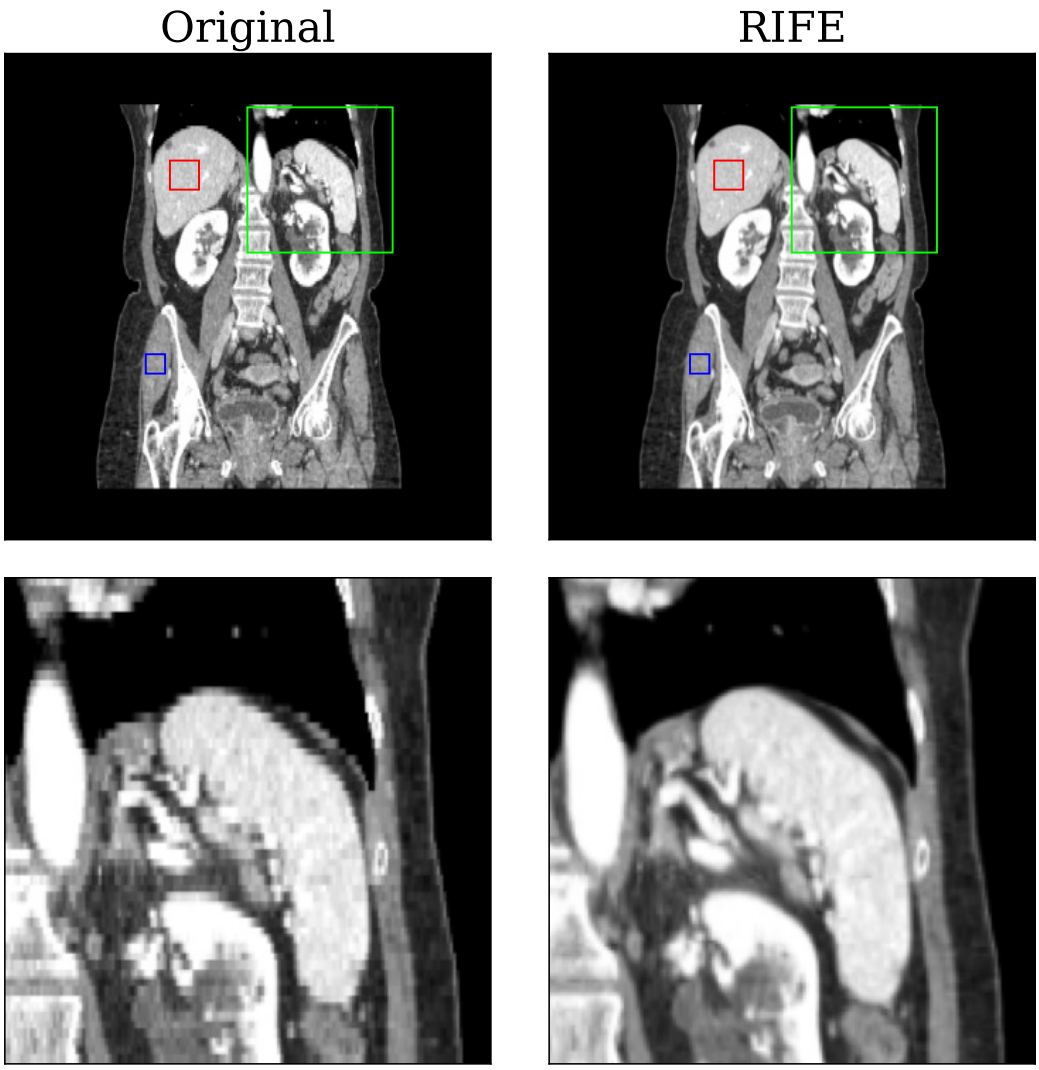

Figure 5.12 – Application of RIFE to an X-ray Computed Tomography dataset. Here the results are presented in the coronal view. The left-hand-side panel displays the original dataset, while the right-hand-side panel show the RIFE-augmented dataset, where three additional frames have been added between every two consecutive frames. The green boxes in the first row of the figure indicate a 150×150-pixels portion of the original-sized data that is magnified in the second row. The red (*Case 1*) and blue (*Case 2*) boxes represent the uniform regions used for the noise power spectrum analysis, whose results are presented in Fig. 5.14.

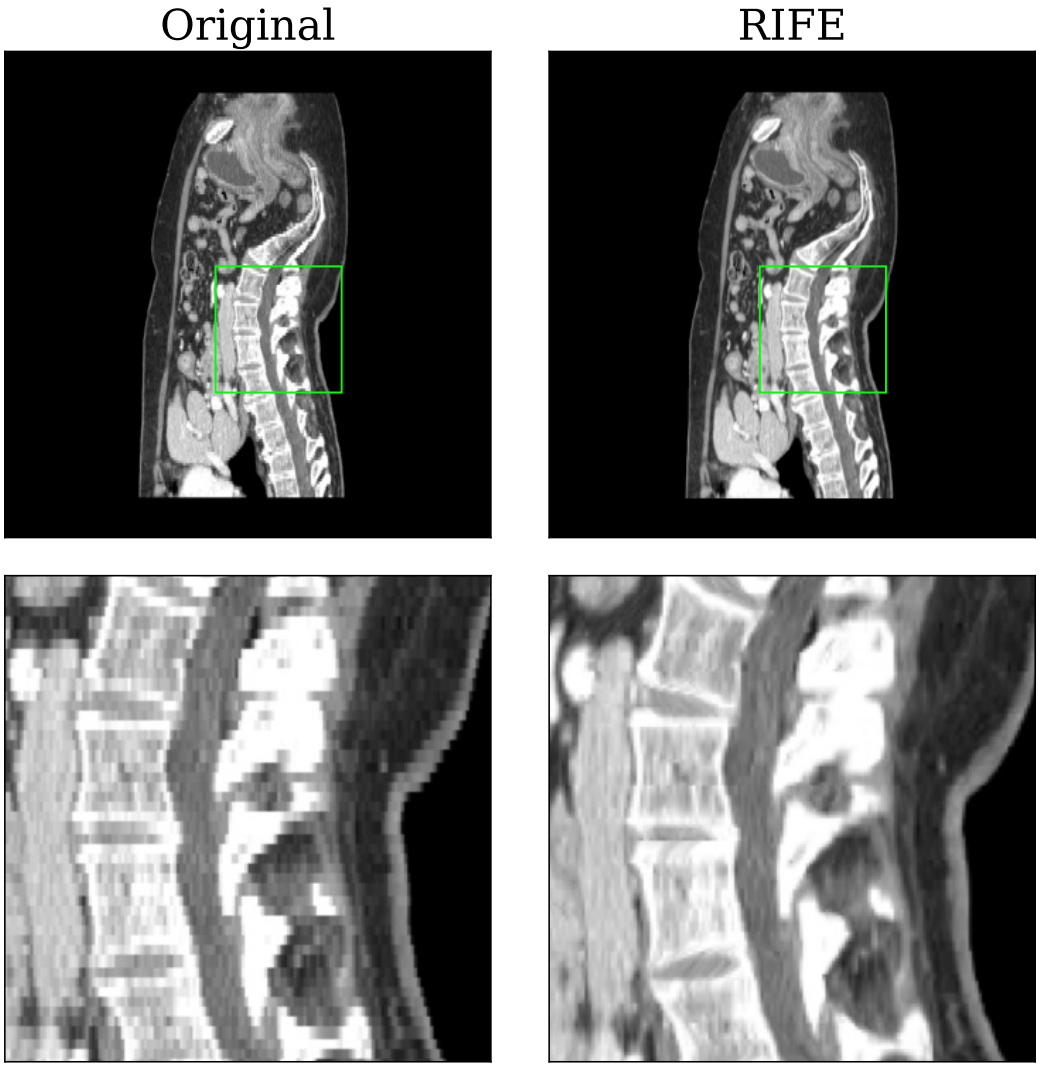

Figure 5.13 – Application of RIFE to an X-ray Computed Tomography dataset. Here the results are presented in the sagittal view. The left-hand-side panel displays the original dataset, while the right-hand-side panel show the RIFE-augmented dataset, where three additional frames have been added between every two consecutive frames. The green boxes in the first row of the figure indicate a 150×150-pixels portion of the original-sized data that is magnified in the second row.

the two image stacks, the original and the RIFE-reconstructed one, are compared by estimating the noise power spectrum on the coronal plane, computed over uniform regions of a sequence of ten images. This metric is generally used for the assessment of the image quality of CT scanners and it is evaluated over uniform regions of interest in water-filled phantoms [121]. As such, in order to adapt this metric to clinical datasets, it is necessary to select uniform areas of the image. For instance, the areas in Fig. 5.12 marked with the red and blue squares represent two possible regions of interest, here called *Case 1* and *Case 2*, respectively. The noise power spectrum of these areas, evaluated over a stack of ten frames, is shown in Fig. 5.14. From the analysis of both cases, it is evident that augmenting the number of frames using RIFE is associated with a reduction of the noise in the data, a reduction visible across the entire frequency range. This quantifies the original observation that RIFE-augmented images look smoother. The same consideration holds for other uniform regions that have been investigated with this approach.

It is worth mentioning that medical images undergo significant processing, which impacts how noise and resolution appear in the final results. This is influenced by several factors, such as image acquisition techniques, reconstruction methodologies, and any additional post-processing steps. Further improvements, achievable with different approaches, are not investigated in this work. Evaluating the effect that different enhancement strategies have on the performance of RIFE, in different contexts, will be the subject of future studies.

5.4 Application to coronary angiography videos

This section discusses an additional application of the video frame interpolation method RIFE in the medical context. In this case, the technique is not used for applications involving 3D tomography data, but for purposes closer to its original development objective, namely increasing videos' frame rates.

Notably, data in video format plays a crucial role in several medical procedures, such as coronary angiography [91]. This practice employs fluoroscopy [90], a real-time X-ray imaging technique, to assess the cardiovascular system. Coronary angiography is an invasive, but relatively safe medical operation, where a catheter is inserted into the human body, in order to deliver a contrast dye into the coronary arteries. Live X-ray imaging helps to guide the catheter and visualize anomalies in the blood flowing through the blood vessels. As explained in Chapter 2, one of the drawbacks of this procedure is that both the patient and the medical practitioner are exposed to ionizing radiation, which is associated with cancer risks [19]. This aspect is particularly concerning for doctors, who perform several procedures during their lifetime and need to adhere to the guidelines for radiation protection [96].

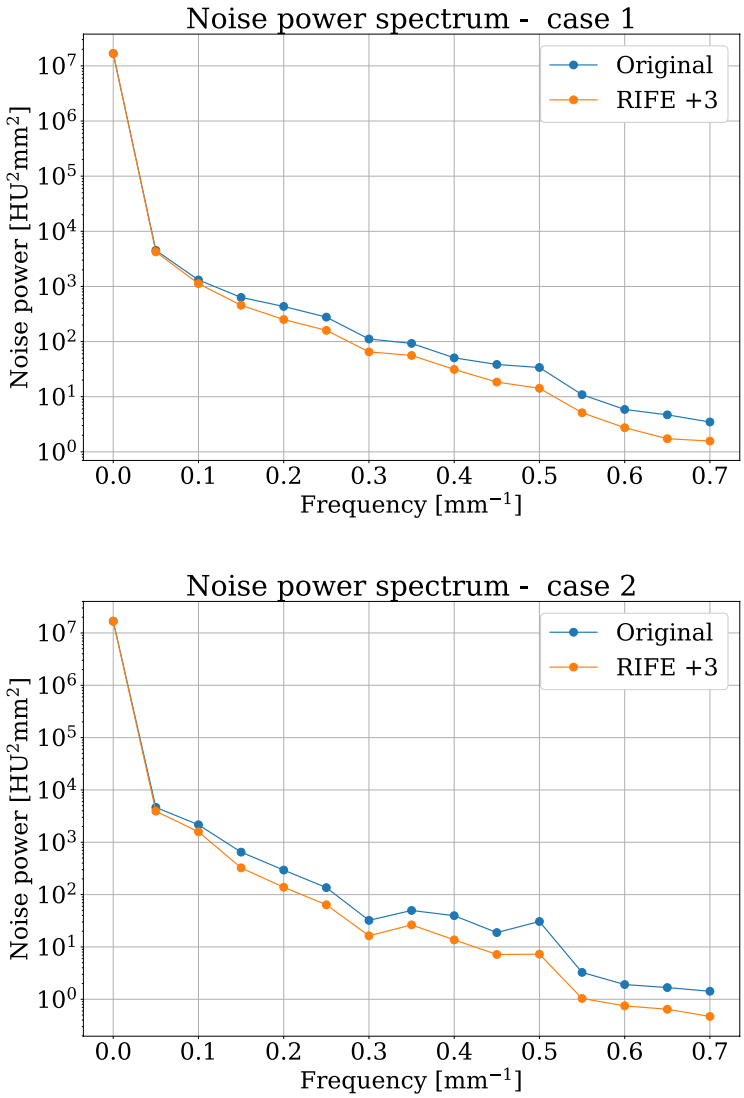

Figure 5.14 – Noise power spectrum analysis over two uniform regions in the coronal view (see Fig. 5.12 for definition). The comparison is conducted between the original and the RIFE-augmented dataset. The use of RIFE results in noise reduction for both the selected regions and across the entire frequency range. The figure on the top panel shows the results for the first case (red box in Fig. 5.12), while the bottom panel displays the results for the second case (blue box in Fig. 5.12).

One key factor in estimating the dose to which they are exposed is the frame rate of the generated videos. In fact, a higher frame rate enhances video quality, resulting in a more detailed and smoother representation of the procedure, though at the cost of increased X-ray exposure due to the greater number of frames captured. One possible solution to this problem is to generate synthetic frames to reduce the exposure to X-rays. The aim of this section is to demonstrate how RIFE can be used to obtain higher frame rates by adopting a fixed amount of ionizing radiation, or, equivalently, how to reduce the X-ray exposure while working at an established frame rate. To give an example, this technique could enable the transformation of a video initially captured at a rate of 15 fps into a video operating at a higher frame rate of 30 fps, allowing good quality visualization and low X-ray exposure at the same time.

Remarkably, some investigation on the topic has been already carried out [180]. According to this study, which is limited to models realized before the development of RIFE, the best-performing algorithm is RRIN [181], which stands for Residue Refinement Interpolation. This video frame interpolation neural network utilizes adaptive weighting and residue refinement to generate intermediate frames between existing ones. In the coronary angiography study of reference [180], RRIN is compared to five other interpolation approaches, including DAIN [69]. The comparison is realized in terms of PSNR and SSIM, after fine-tuning the models on a large dataset, which is not made available to the public. Regrettably, the time required for frame generation is not investigated. This is a crucial factor in contexts, where the algorithm should be applied in real-time, as in the case of coronary angiography procedures, and not as a post-processing optimization. Furthermore, as previously discussed, the estimation of metrics such as PSNR and SSIM alone might not be the best strategy to assess the quality of the information one can retrieve from the data.

The aim of this section is to propose a more complete investigation of this subject, by considering additional evaluation metrics. Three main approaches are being investigated, namely RIFE, RRIN, and linear interpolation. To be consistent with the mentioned study [180], fine-tuning is performed on the pre-trained RRIN model.

Two types of datasets are considered for testing the different strategies. The first dataset consists of quality assurance test objects. Secondly, clinically acquired video frames of human patients are analyzed.

5.4.1 Quality assurance test objects

In order to assess the model's performance, quality assurance test objects are considered, also known as phantoms. These are specialized tools developed to inspect the accuracy of medical imaging instruments. They are designed to replicate

Table 5.1 – Video Frame Rate, Length, and Number of Frames for the data acquired with the Leeds test object.

Frame Rate (fps)	*Video Length (s)*	*Number of Frames*
4	17.00	68
7.5	10.53	79
15	10.46	157
30	11.93	358

different human tissue densities and structures. For this research project, the Leeds test object set [182] for digital subtraction fluorography[2] [183] is considered. Specifically, the TO J3 component is used. This consists of two plates, each one divided into four sections, with the relevant test details. One of the plates is a fixed-base plate, while the second is a rotatable top plate. Videos of the Leeds test object were acquired at the Mater Misericordiae University Hospital, Dublin, on a Siemens Artis Zee angiography system, using the FL Service Mode, which is the standard dose mode. The energy of the X-ray was set to 70 kVp, while the field of view was 42 cm. To specify, kVp, or kilovoltage peak, refers to the peak voltage applied across the X-ray tube in an X-ray machine. This value allows radiologists to control the quality and penetration of X-rays used in medical imaging systems.

The videos were acquired at different frame rates, for different intervals of time. More details can be found in Table 5.1. The object was manually rotated with a stick by an operator. Therefore, the rotation speed is unknown. The software FIJI was used to convert the files from video format (*.IMA* extension) to sequences of images (size 960×960 pixels) in *.tiff* format. For this analysis, RIFE is compared to linear interpolation results. RRIN is excluded from the investigation due to the unfeasibility of performing fine-tuning on it. This is caused by the lack of additional data that could be used for training, which should be separated from the test set, in order to avoid overfitting. For this preliminary investigation, only the videos captured at 15 and 30 fps are considered. For these datasets, frames are removed from the original sequence and used as ground truth for comparison. Specifically, two cases are considered: one and three frames are discarded every two frames. The selected datasets will be referred to as 15 fps and 30 fps in this work, meaning that the original frame rate will be used to indicate them. However, it is worth observing that a 30 fps video from which every other frame is removed corresponds to a 15 fps video. Similarly, removing three frames between every two existing frames of a 30 fps video is equivalent to having a 7.5 fps video.

Firstly, the results are analyzed visually. Fig. 5.15 displays a visual comparison for the case of the 30 fps video and one replaced frame. RIFE and linear interpolation

[2]Digital subtraction fluorography is an imaging technique that combines real-time X-ray, namely fluoroscopy, with digital image processing, in order to enhance the visualization of anatomical features.

are applied to the full sequence, but only four frames are shown to examine the performance of the different augmentation methods. The first column displays the original ground-truth data. The second and third columns show the images generated using RIFE and linear interpolation, respectively. The difference between these results and the ground-truth image is depicted in the fourth and fifth columns, for the two analyzed methodologies. The red/blue colourbar refers only to the images in these last two columns. The first three columns show grayscale images, with intensity within the range of $0 - 1$.

A similar comparison is presented in Fig. 5.16, for the case of three removed frames. The same kind of investigation is performed for the case of videos acquired at 15 fps, whose results are displayed in Fig. 5.17 and 5.18, for the one- and three-replaced-frames cases, respectively.

From all the figures, it is evident that RIFE performs better than linear interpolation in the intermediate frame generation task. In fact, RIFE is able to correctly depict the objects' movement and presents errors mainly on the static background, which shows some noise. In contrast, the linear interpolation results show systematic failures around the borders of the moving parts of the phantom, in addition to small errors in the static areas. Undoubtedly, errors are more evident for both methods in the three-replaced-frames cases, for any frame rate of the original video. This feature is expected and was experienced also for the other datasets investigated in this chapter. It should be noted that the dark stick seen in various locations across all the images is not part of the Leeds test object but is the tool used by the operator to rotate the test equipment.

In order to provide a more quantitative interpretation of these results, several metrics are considered. Firstly, some standard computer vision metrics, such as MSE, PSNR, and SSIM. Multi-scale SSIM is also considered, referred to as MS SSIM. Additionally, FID (Fréchet Inception Distance) and WDS (Wasserstein Distance Score) are computed. More details about the listed metrics can be found in Chapter 2. To recap their meaning, the metrics that indicate good performance when they have higher values are PSNR, SSIM, and MS SSIM. In contrast, MSE, FID, and WDS show better outcomes when they are low. It should be noted that only the frames artificially generated are considered for the comparison, meaning that the identical frames (i.e. the ones not removed from the original dataset) are not included in the error evaluation, since there would be no associated error. In this section, the results are shown as violin plots (Fig. 5.19, 5.20, 5.21, 5.22, 5.23, 5.24), in order to highlight the distribution of the examined metric over the considered sequence of frames. This provides a summary of the metric across the frames, which can be insightful for understanding the overall performance. For each metric, the cases of one and three replaced frames are presented, both for the video collected at 30 fps (left-hand side panel) and the video collected at 15 fps (right-hand side panel). The RIFE and

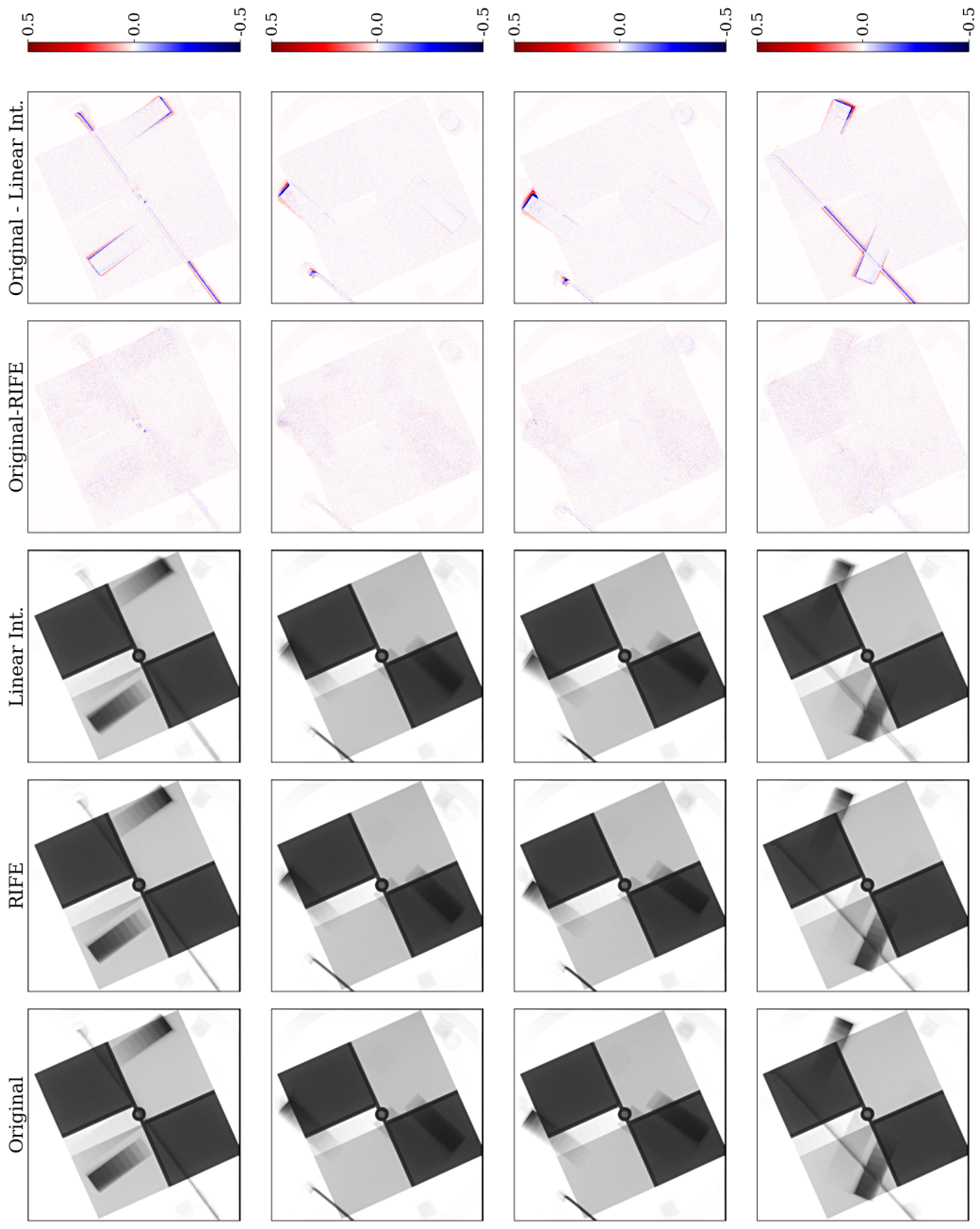

Figure 5.15 – Visual comparison of the frame reconstruction in the case where one frame is removed from the 30 fps Leeds test object sequence. Four examples of results are presented, one for each row. From left to right the figure displays: the ground truth image (original frame removed from the dataset), those generated by RIFE and linear interpolation, the difference between the original data and RIFE, and the difference between the original data and the linear interpolation result. Regarding the image intensities, the first three columns are grayscale images with values between 0 and 1, while the last two columns both follow the colourbar depicted on the right side of the figure.

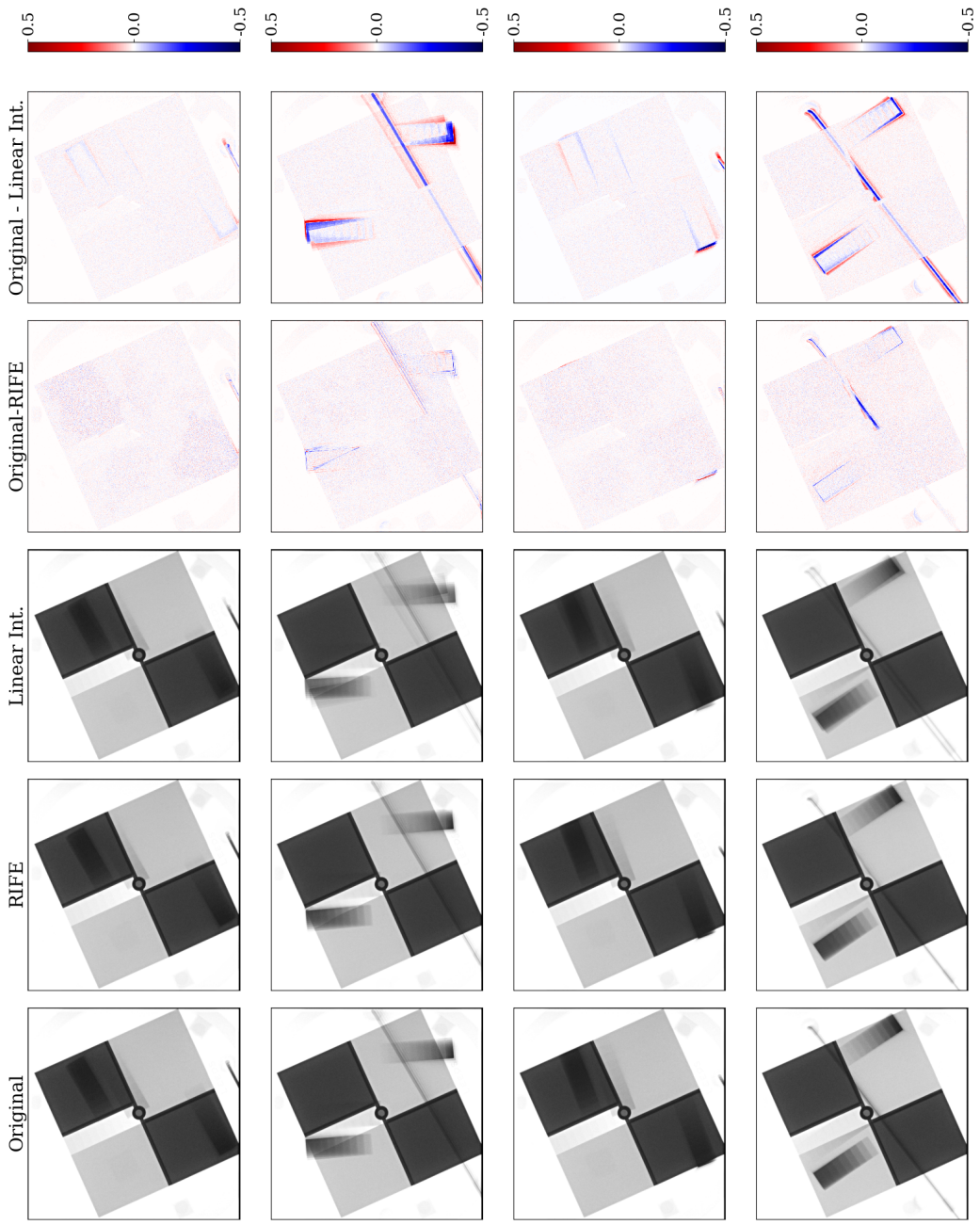

Figure 5.16 – Visual comparison of the frame reconstruction in the case where three frames are removed from the 30 fps Leeds test object sequence. Four examples of results are presented, one for each row. From left to right the figure displays: the ground truth image (original frame removed from the dataset), those generated by RIFE and linear interpolation, the difference between the original data and RIFE, and the difference between the original data and the linear interpolation result. Regarding the image intensities, the first three columns are grayscale images with values between 0 and 1, while the last two columns both follow the colourbar depicted on the right side of the figure.

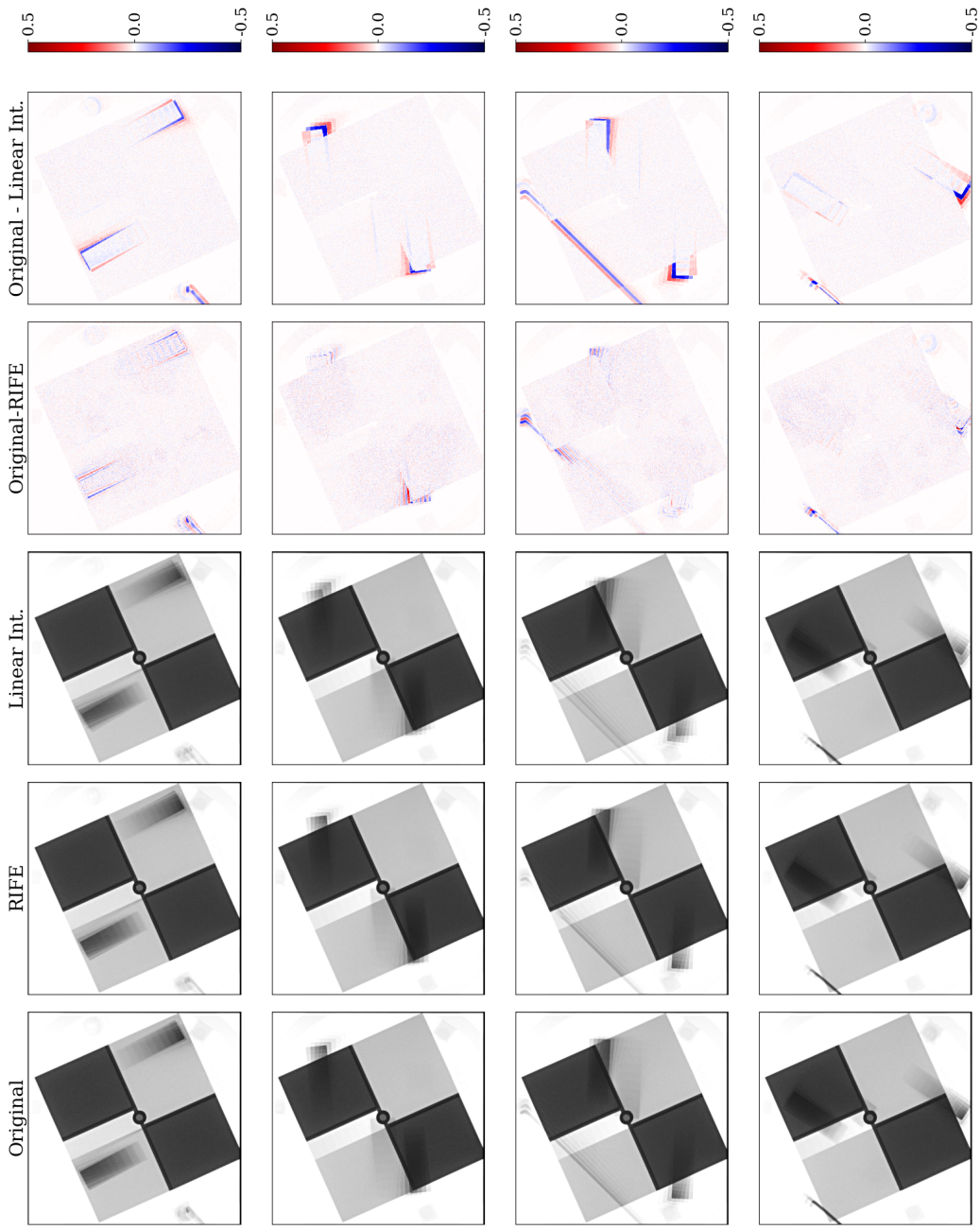

Figure 5.17 – Visual comparison of the frame reconstruction in the case where one frame is removed from the 15 fps Leeds test object sequence. Four examples of results are presented, one for each row. From left to right the figure displays: the ground truth image (original frame removed from the dataset), those generated by RIFE and linear interpolation, the difference between the original data and RIFE, and the difference between the original data and the linear interpolation result. Regarding the image intensities, the first three columns are grayscale images with values between 0 and 1, while the last two columns both follow the colourbar depicted on the right side of the figure.

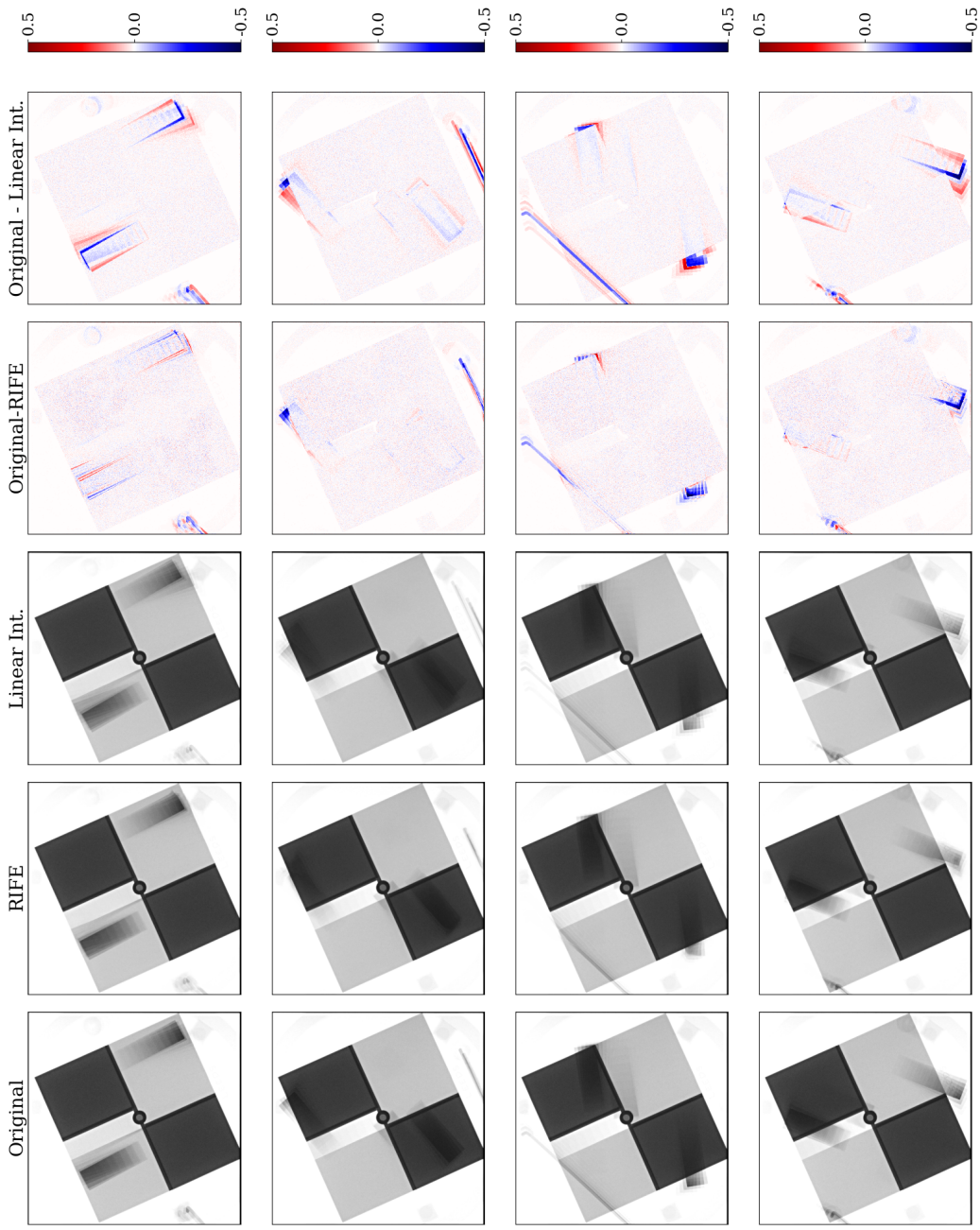

Figure 5.18 – Visual comparison of the frame reconstruction in the case where three frames are removed from the 15 fps Leeds test object sequence. Four examples of results are presented, one for each row. From left to right the figure displays: the ground truth image (original frame removed from the dataset), those generated by RIFE and linear interpolation, the difference between the original data and RIFE, and the difference between the original data and the linear interpolation result. Regarding the image intensities, the first three columns are grayscale images with values between 0 and 1, while the last two columns both follow the colourbar depicted on the right side of the figure.

Table 5.2 – Comparison of mean values of several metrics, detailed in the first column. The mean values are presented for a video acquired at 30 fps. The case of one and three replaced frames are considered, for both the investigated interpolation methods, namely RIFE and linear interpolation. The data refers to the Leeds test object.

Video 30 fps				
	1 replaced		*3 replaced*	
	RIFE	Linear Int.	RIFE	Linear Int.
MSE	0.00012	0.00027	0.00019	0.00054
PSNR	87.5151	84.2797	85.4859	81.2677
SSIM	0.97682	0.98260	0.96896	0.97566
MS SSIM	0.98416	0.98119	0.97690	0.96909
FID	0.05102	0.13586	0.08285	0.26514
WDS	0.00147	0.00154	0.00166	0.00236

Table 5.3 – Comparison of mean values of several metrics, detailed in the first column. The mean values are presented for a video acquired at 15 fps. The case of one and three replaced frames are considered, for both the investigated interpolation methods, namely RIFE and linear interpolation. The data refers to the Leeds test object.

Video 15 fps				
	1 replaced		*3 replaced*	
	RIFE	Linear Int.	RIFE	Linear Int.
MSE	0.00020	0.00034	0.00044	0.00065
PSNR	85.4398	83.3016	82.1740	80.4832
SSIM	0.97194	0.98312	0.96320	0.97688
MS SSIM	0.97550	0.97631	0.96451	0.96496
FID	0.09638	0.17980	0.21860	0.34800
WDS	0.00155	0.00197	0.00214	0.00310

linear interpolation methods are considered. For each method-case combination, the mean value of the analyzed metric is displayed in the relevant violin plot, as a black dot. The exact numeric values of the mean for each metric distribution are detailed in Table 5.2 and Table 5.3. As expected, and in agreement with the visual analysis of the images, the results are more inaccurate when increasing the number of replaced frames from one to three, for all the considered metrics. Overall, errors appear to be higher for all results obtained on the 15 fps sequence, as foreseeable. The comparison between RIFE and linear interpolation is consistent with what can be evinced for the visual analysis of the results for all metrics except SSIM, for both the 30 fps and 15 fps cases. In this latter case, also the MS SSIM metric appears to be slightly better for the linear interpolation method than RIFE; indeed, the MS SSIM mean value of linear interpolation is 0.0831 % higher than the one of RIFE. For all the other metrics, error reduction is experienced when using RIFE instead of linear interpolation.

Additional studies are needed to expand and complete this investigation. Firstly

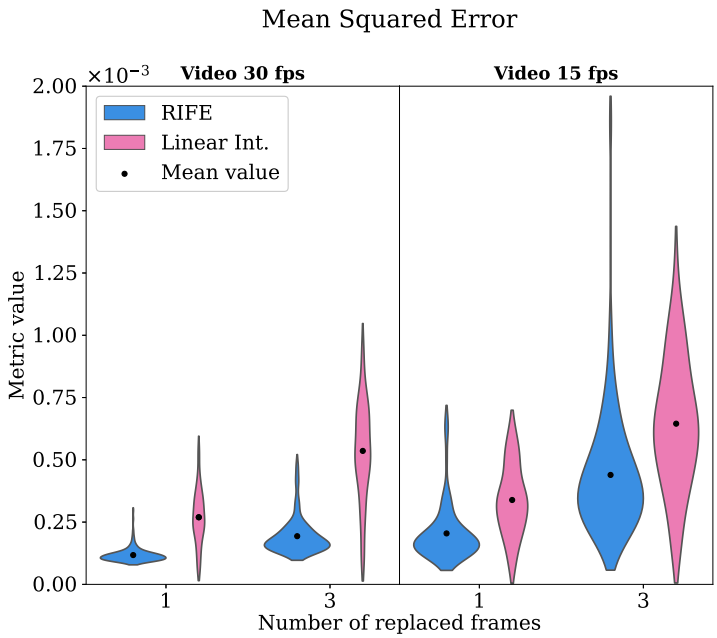

Figure 5.19 – Comparison of Mean Squared Error (MSE) distributions for RIFE and linear interpolation methods across two test cases (one and three replaced frames). The results for video acquired at 30 fps are presented on the left-hand side panel, while the results for the 15 fps video are on the right-hand side panel. Violin plots illustrate the distribution of metric values, with the black dot representing the mean value for each method-case combination.

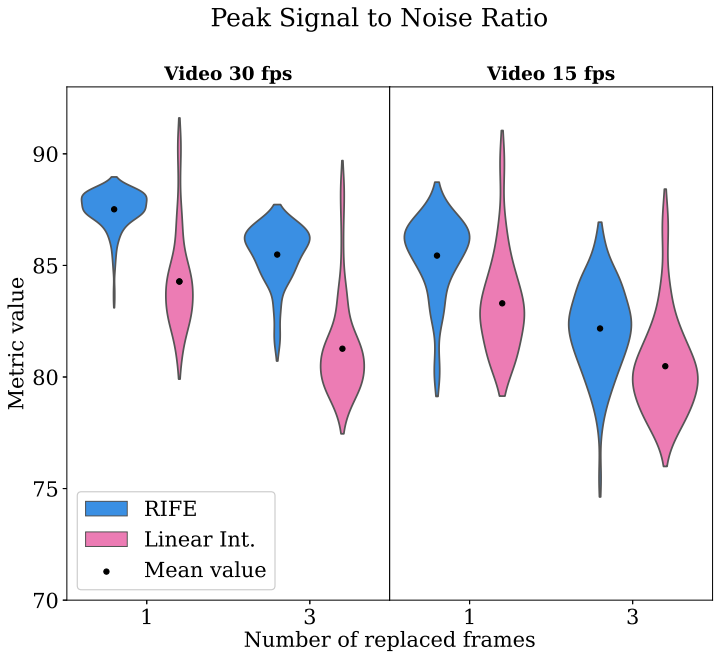

Figure 5.20 – Comparison of Peak Signal to Noise Ratio (PSNR) distributions for RIFE and linear interpolation methods across two test cases (one and three replaced frames). The results for video acquired at 30 fps are presented on the left-hand side panel, while the results for the 15 fps video are on the right-hand side panel. Violin plots illustrate the distribution of metric values, with the black dot representing the mean value for each method-case combination. The results refer to the Leeds test object.

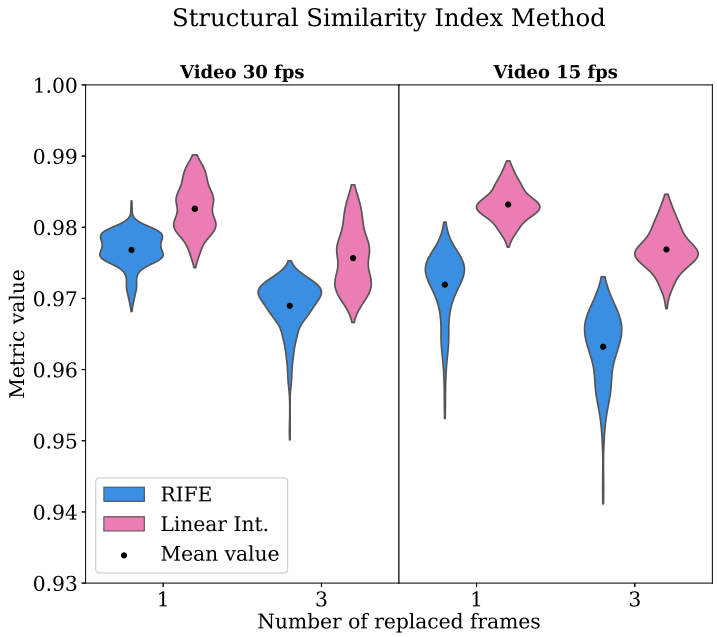

Figure 5.21 – Comparison of Structural Similarity Index Method (SSIM) distributions for RIFE and linear interpolation methods across two test cases (one and three replaced frames). The results for video acquired at 30 fps are presented on the left-hand side panel, while the results for the 15 fps video are on the right-hand side panel. Violin plots illustrate the distribution of metric values, with the black dot representing the mean value for each method-case combination. The results refer to the Leeds test object.

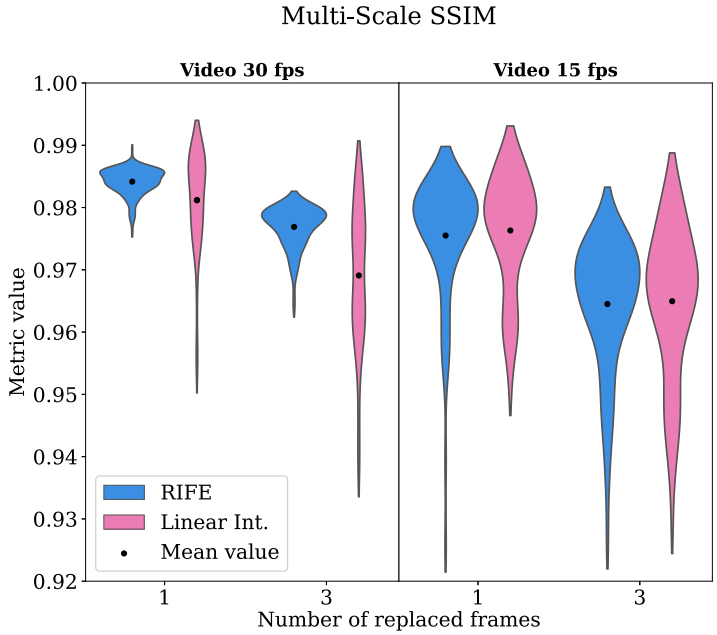

Figure 5.22 – Comparison of Multi-Scale SSIM (MS SSIM) distributions for RIFE and linear interpolation methods across two test cases (one and three replaced frames). The results for video acquired at 30 fps are presented on the left-hand side panel, while the results for the 15 fps video are on the right-hand side panel. Violin plots illustrate the distribution of metric values, with the black dot representing the mean value for each method-case combination. The results refer to the Leeds test object.

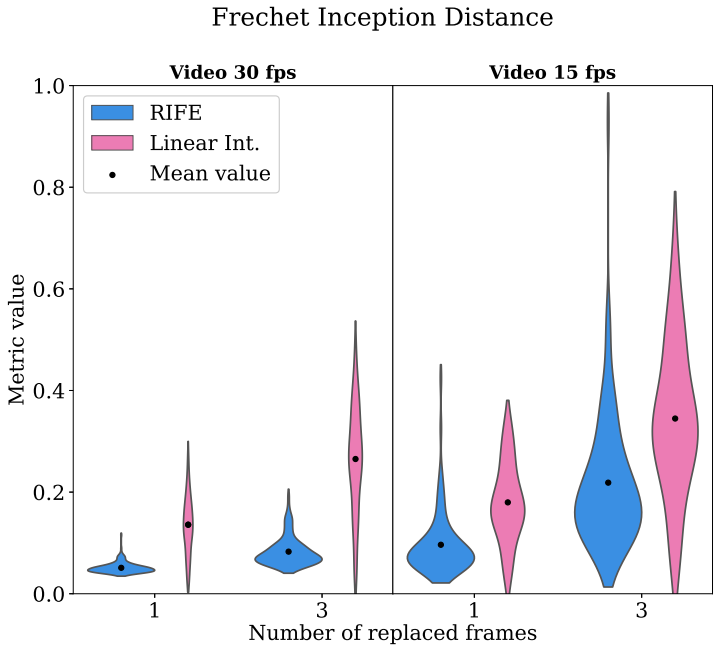

Figure 5.23 – Comparison of Frechet Inception Distance (FID) distributions for RIFE and linear interpolation methods across two test cases (one and three replaced frames). The results for video acquired at 30 fps are presented on the left-hand side panel, while the results for the 15 fps video are on the right-hand side panel. Violin plots illustrate the distribution of metric values, with the black dot representing the mean value for each method-case combination. The results refer to the Leeds test object.

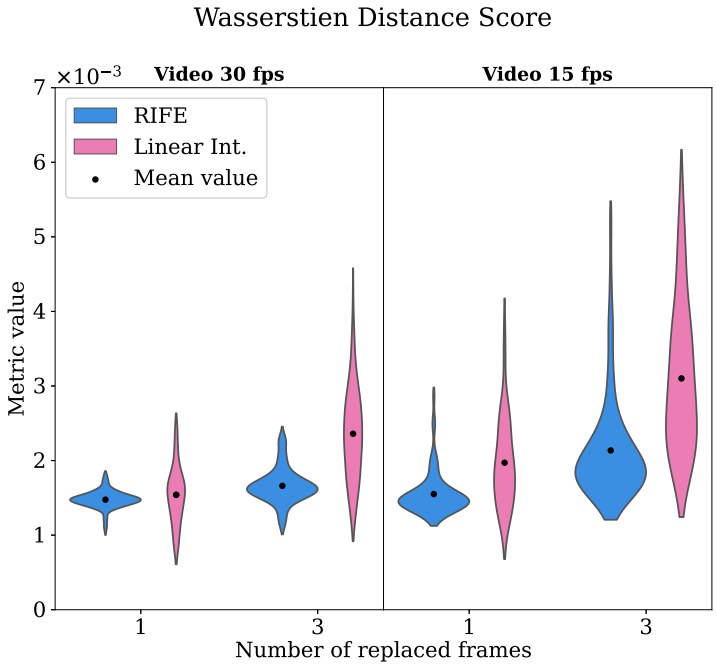

Figure 5.24 – Comparison of Wasserstein Distance Score (WDS) distributions for RIFE and linear interpolation methods across two test cases (one and three replaced frames). The results for video acquired at 30 fps are presented on the left-hand side panel, while the results for the 15 fps video are on the right-hand side panel. Violin plots illustrate the distribution of metric values, with the black dot representing the mean value for each method-case combination. The results refer to the Leeds test object.

videos acquired at different frame rates should be examined with the proposed approaches. In line with the results obtained for the 30 fps and 15 fps cases, a deterioration of the performance of each model is expected when reducing the video frame rate. Secondly, other test objects should be taken into consideration, such as the NEMA phantom [184], widely used to benchmark cardiac fluoroscopy performance.

Furthermore, it would be beneficial to examine the metrics values by focusing only on the moving parts of the images, without considering errors related to the background. In fact, these regions are affected by static noise, which is not too relevant to the aim of fluoroscopy procedures. This alternative analysis would help to focus on the key areas of interest for fluoroscopy and could potentially lead to more clinically relevant results. It must be noted that the background-related noise would be removed only for evaluation purposes and not for the frame generation, for instance with some smoothing filter. Indeed, this would cause flickering in the video, due to different noise levels between the original and the artificially added frames.

Lastly, the acquisition of fluoroscopy videos at different X-ray energy levels should be addressed. The kVp is a critical parameter in X-ray imaging that determines the energy and penetration of the X-ray beam, with impact on image contrast, quality, and radiation dose to the patient. Having images with different levels of contrast could affect the interpolation performance, for all the examined methods. Therefore, interesting results could be deduced from this analysis, which could have an effect on the kVp settings during clinical acquisition too.

5.4.2 Clinical data

Clinical data from coronary angiography procedures are also considered for this study. The aim of this section is to describe some preliminary results, obtained on a dataset made of 124 frames (size 512×512 pixels), extracted from a coronary angiography video, with a frame rate of 30 fps. Also for this case, one frame every two frames is removed and used as ground truth, for performance assessment. For this investigation, three techniques are compared: RIFE, RRIN, and linear interpolation. In order to make the study consistent with the one conducted by Yin *et al.* [180], fine-tuning is performed on the pre-trained RRIN model. Specifically, the algorithm is fine-tuned for 100 epochs on 233 triplets from a clinical dataset acquired at 30 fps, different from the dataset used for testing.

As always, the first comparison is visual. Fig. 5.25 and 5.26 show the results for two different frames, randomly selected from the clinical dataset. In both representations, the image in the first column shows the original frame removed from the dataset, namely the ground truth. The same frame, generated with different interpolation methods, is presented in the following columns. Going

from left to right, the figure displays the results obtained with RIFE, RRIN, and linear interpolation. On the second row, the difference between the ground truth and the mentioned methods can be found. In order to investigate the results more in detail, a portion of size 100×100-pixel of the images in the first row is magnified and displayed in the third row. The exact location of this portion is specified by the green box in the upper left panel. Finally, the differences between the ground truth and the analyzed interpolation method, for the reduced images, are presented in the fourth row. From both examples, it is evident that linear interpolation should be avoided, since it provides inaccurate results, especially in areas separating the background from the moving objects (i.e. the blood vessels in this case). RIFE and RRIN seem to provide similar results in terms of vessel localization. However, the frame generated by using RRIN presents smoothed features and, in general, higher brightness, as it can be evinced by the prevalence of blue pixels in the panels in the second and fourth rows. Blue pixels correspond to negative values in the difference, meaning that the RRIN-generated image is made of pixels with systematically higher values, compared to the ground-truth data. In contrast, RIFE difference images present a more uniform error distribution in the background areas, which reflects the expected static noise effect. This is not necessarily a problem when the goal of the analysis is to examine the blood flowing through the vessels per se, but it can make the video visualization unpleasant, due to flickering between the experimentally acquired and the artificially generated images. Therefore, real-time use during the medical procedure would be hindered by this aspect.

Another important investigation concerns the time required to generate one additional frame, using two different methods, namely RIFE and RRIN. The comparison is carried out both with and without the use of a GPU, which can have a significant impact on the model speed. This investigation is performed considering clinical images of size 512×512 pixels. Without a GPU, the average time to generate an image using RRIN is about 6.14 s, which can be reduced to 1.39 s with RIFE. The performance improves for both models when a GPU is employed. Specifically, RRIN requires 0.03 s on average to produce a 512×512 pixels image, while RIFE only takes 0.01 s. In order to be suitable for real-time applications, the model should be able to perform the interpolation task in less than the human reaction time, that is in the range of 0.20 to 0.25 s. Therefore, the solutions without GPU need to be excluded from live applications. When employing a GPU, RIFE appears to be more advantageous from the speed perspective. It is worth mentioning that linear interpolation was excluded from this speed analysis, due to the numerous inaccuracies found in the generated frames.

Another aspect that makes RIFE a better candidate for the purpose of this investigation, is that there is no requirement for fine-tuning. In fact, this would

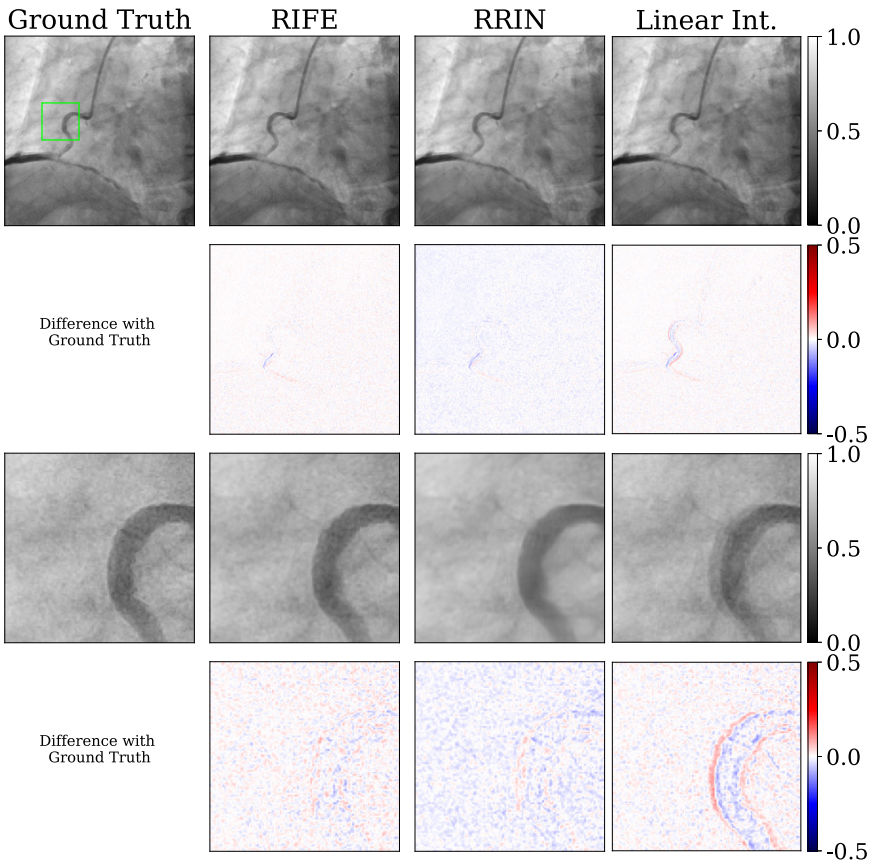

Figure 5.25 – First example of visual comparison of the image reconstruction in the case where one frame is removed from the clinical coronary angiography sequence. From left to right the figure displays: the ground truth image (original frame removed from the dataset), those generated by RIFE, RRIN, and linear interpolation. The second row displays the difference between the ground truth and the reconstructions. A 100×100-pixel portion of each image (see green box in the upper left panel) is magnified and shown in the third row, while the differences from the original image are in the fourth row.

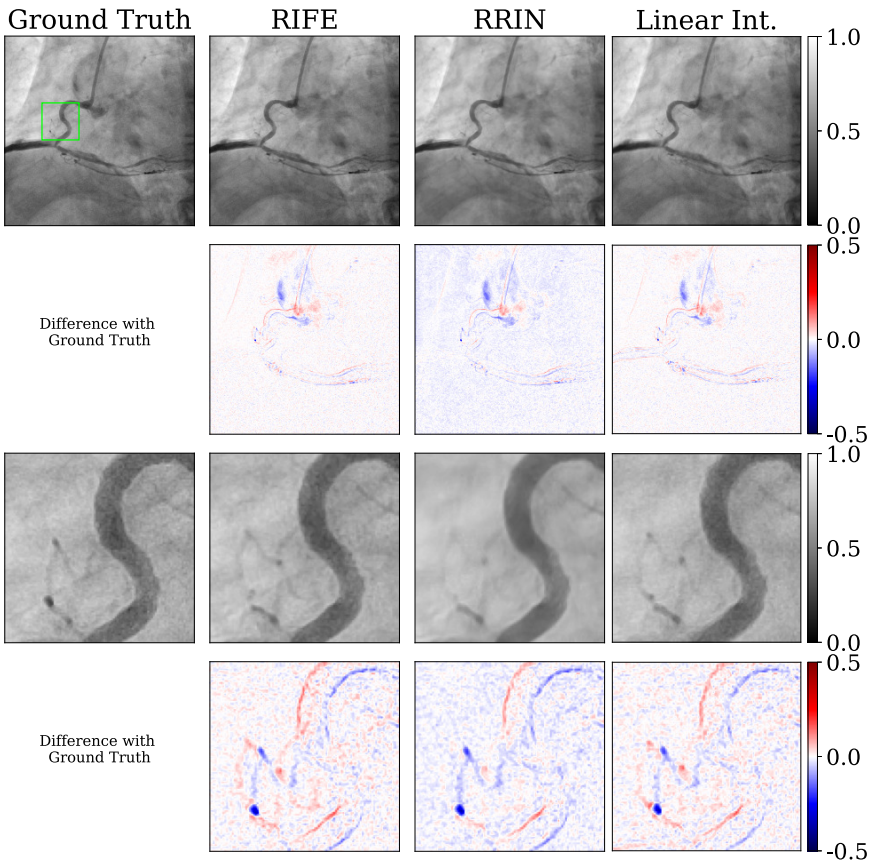

Figure 5.26 – Second example of visual comparison of the image reconstruction in the case where one frame is removed from the clinical coronary angiography sequence. From left to right the figure displays: the ground truth image (original frame removed from the dataset), those generated by RIFE, RRIN, and linear interpolation. The second row displays the difference between the ground truth and the reconstructions. A 100×100-pixel portion of each image (see green box in the upper left panel) is magnified and shown in the third row, while the differences from the original image are in the fourth row.

Table 5.4 – Comparison of mean values of several metrics, detailed in the first column. The mean values are presented for a clinical video acquired at 30 fps. The case of one replaced frame is considered, for all the investigated interpolation methods, namely RIFE, linear interpolation, and RRIN.

Video 30 fps			
	1 replaced		
	RIFE	Linear Int.	RRIN
MSE	0.00025	0.00034	0.00019
PSNR	84.0746	82.7833	85.3158
SSIM	0.94925	0.93639	0.97158
MS SSIM	0.94635	0.92827	0.96208
FID	0.10755	0.14205	0.08813
WDS	0.00250	0.00276	0.00171

require additional data, ideally coming from other acquisition sessions, carried out with different equipment, to avoid the risk of overfitting. Moreover, any data belonging to the medical context should undergo some ethical approval procedures, which might limit the model distribution. It should be noted that the original pre-trained RRIN model has been considered for an initial visual performance assessment. However, the images were more inaccurate and, therefore, only the fine-tuned version of RRIN is presented in this work.

A quantitative analysis, similar to the one performed for the Leeds test object, was conducted also in this case. The results can be found in Fig. 5.27, 5.28, and 5.29, which displays the distribution of several metrics (MSE, PSNR, SSIM, MS SSIM, FID, and WDS), for the 30 fps video and one replaced frame. The mean value of each distribution is detailed in Table 5.4, for each method-case combination. For all the analyzed metrics, the results are not consistent with what can be concluded from the visual comparison. In fact, as it can be seen from Fig. 5.27, 5.28, and 5.29, RRIN appears to be the best-performing approach, according to all metrics, followed by RIFE and lastly by the linear interpolation. The reason behind these outcomes could be related to the smoothing that characterizes the RRIN images, which contributes to lowering the error, especially in the background areas, which is the greater part of the image.

This controversy between the visual evaluation and the computed results seems to indicate that the presented metrics are not adequate for performance assessments, in this context. The same analysis was also conducted on two other datasets, also retrieved from 30 fps video sequences, with similar results. Therefore, these metrics might not fully capture the nuances of clinical data or the preferences of medical experts. For this reason, the described evaluation was not extended to other videos, obtained at different frame rates.

Different approaches are currently being investigated to analyze the results related to clinical data. One possible solution is to adopt a strategy similar to the

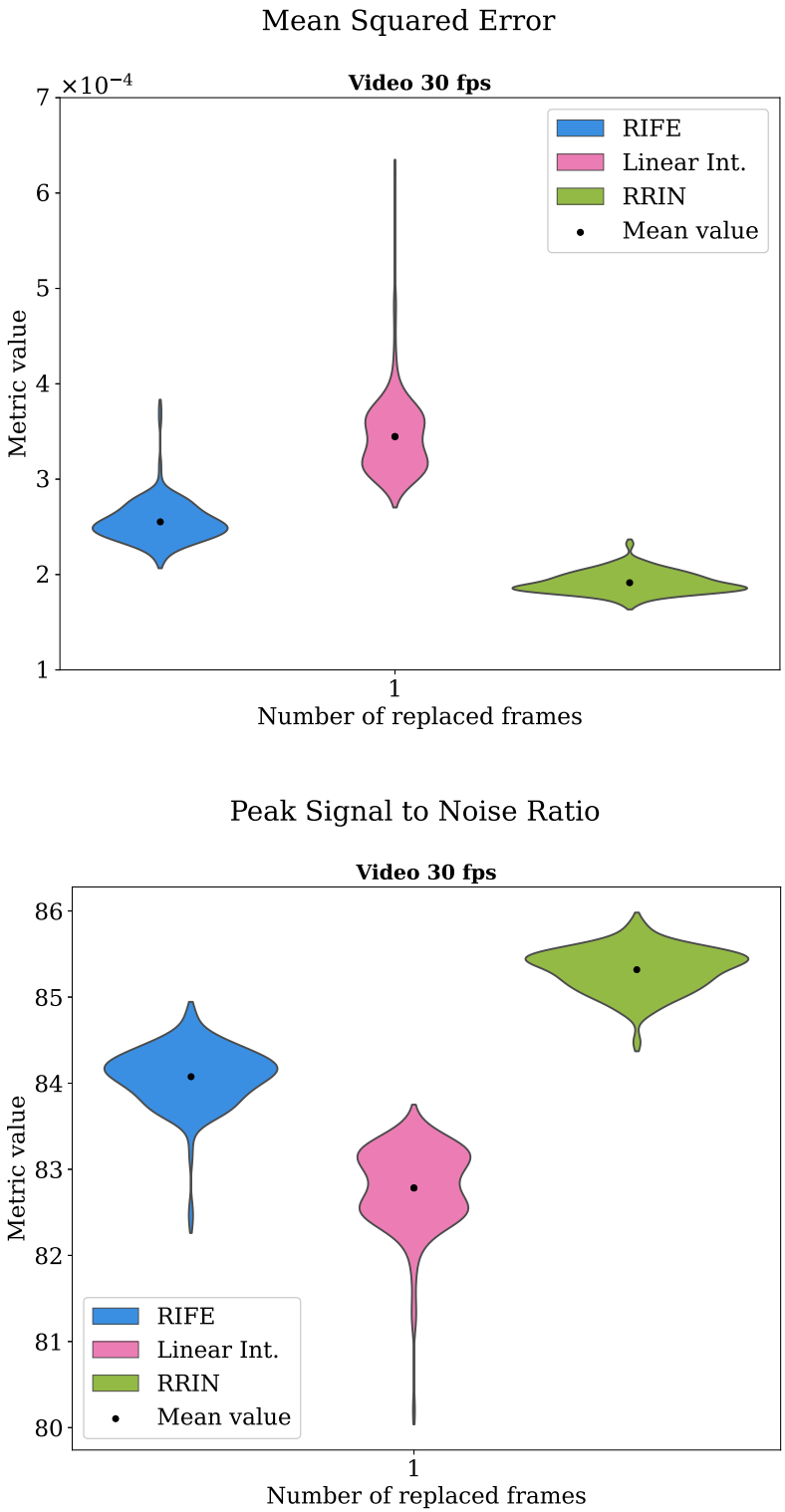

Figure 5.27 – Comparison of Mean Squared Error (MSE) distributions (top panel) and Peak Signal to Noise Ratio (PSNR) distributions (bottom panel) for RIFE, linear interpolation, and RRIN methods. The results for video acquired at 30 fps are presented, for the case of one replaced frame. The black dot represents the mean value for each method-case combination.

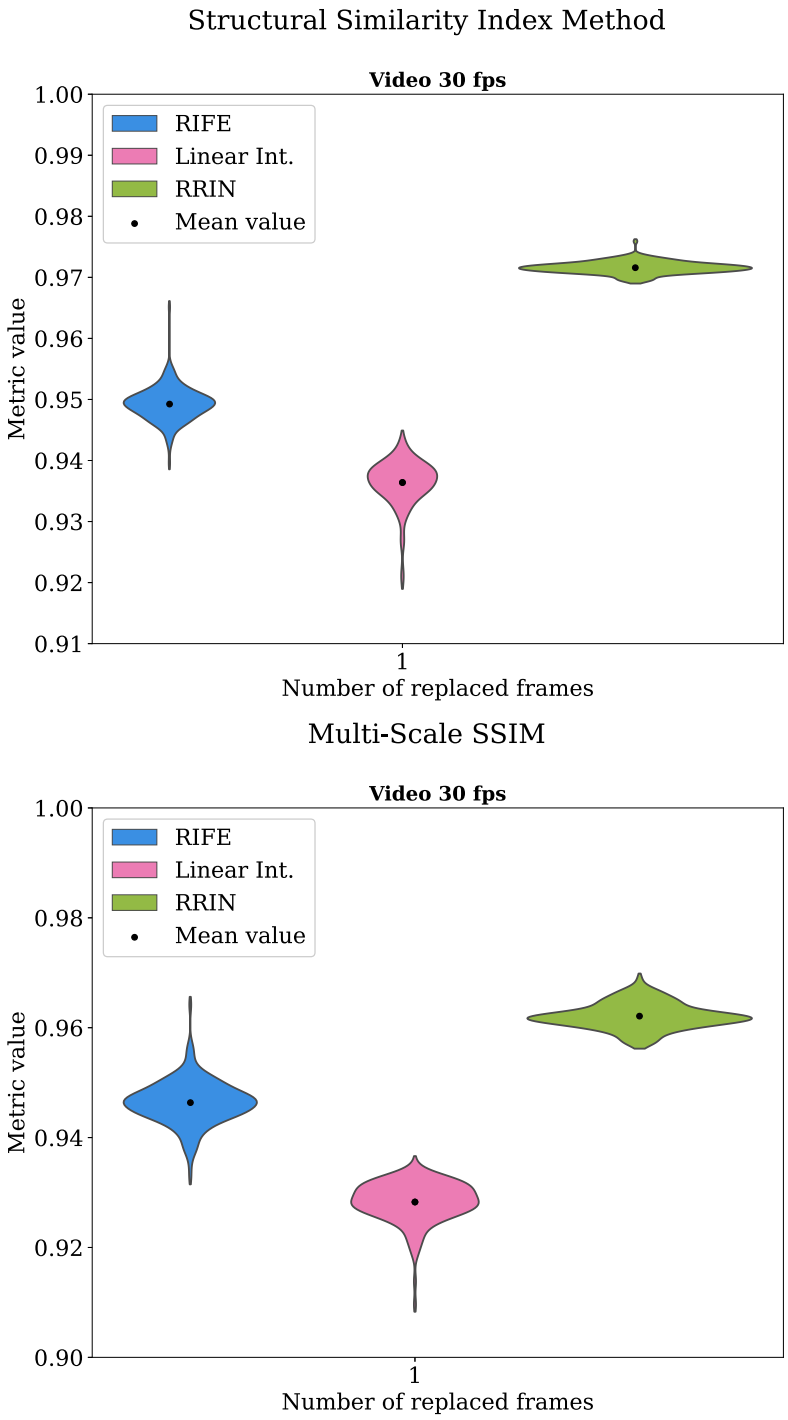

Figure 5.28 – Comparison of Structural Similarity Index Method (SSIM) distributions (top panel) and Multi-Scale SSIM (MS SSIM) distributions (bottom panel) for RIFE, linear interpolation, and RRIN methods. The results for video acquired at 30 fps are presented, for the case of one replaced frame. The black dot represents the mean value for each method-case combination.

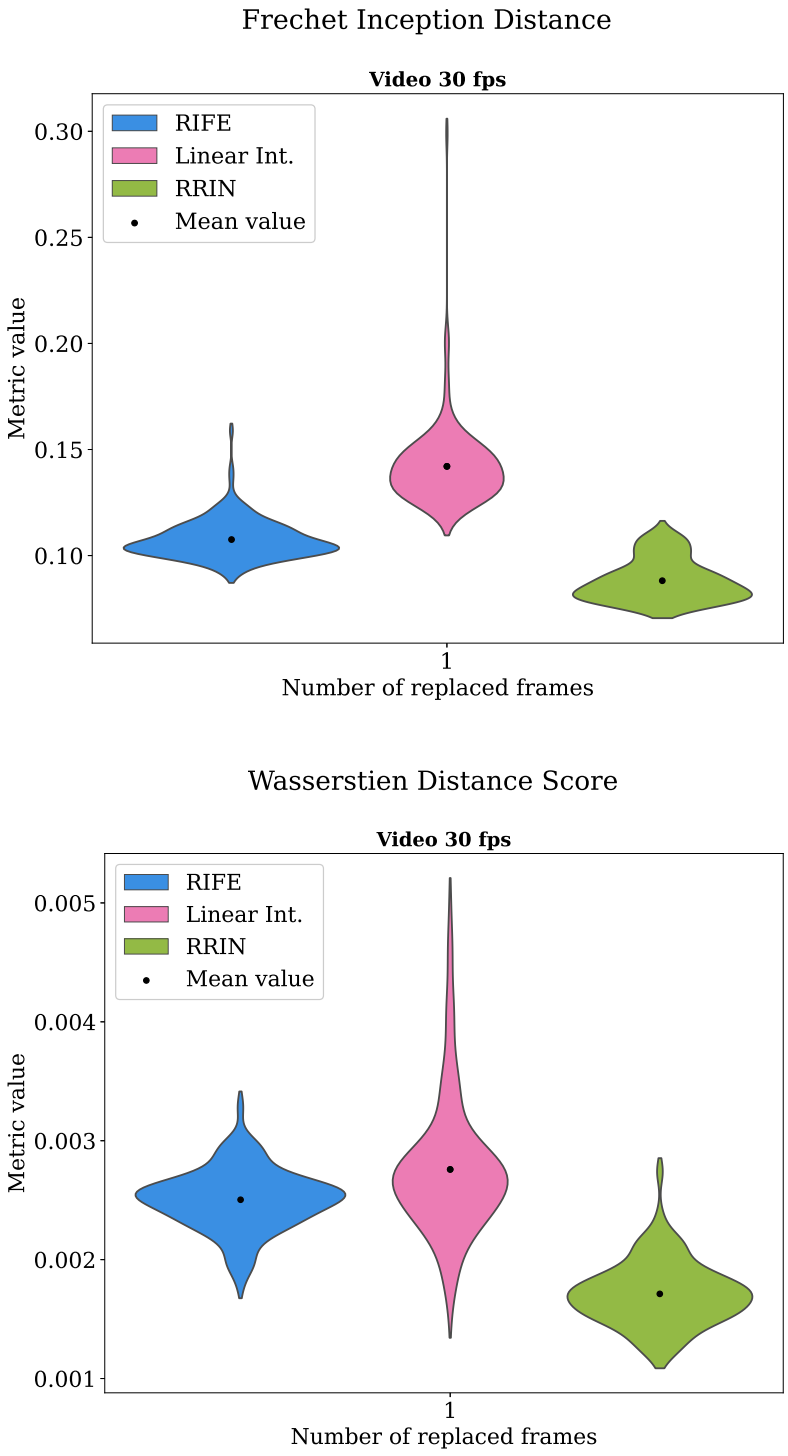

Figure 5.29 – Comparison of Frechet Inception Distance (FID) distributions (top panel) and Wasserstein Distance Score (WDS) (bottom panel) for RIFE, linear interpolation, and RRIN methods. The results for video acquired at 30 fps are presented, for the case of one replaced frame. The black dot represents the mean value for each method-case combination.

one followed for the FIB-SEM dataset, meaning the segmentation of images before evaluating specific metrics. However, in this case, being the nature of the images more convoluted, a more sophisticated approach would be necessary, compared to the binarization performed with the WEKA segmentation tool [167] available in FIJI [45]. In the clinical case, a neural network developed for image segmentation should be used. Notably, several algorithms are available for image segmentation, also for medical data [185, 186]. However, the images and the correspondent segmentation masks needed to train or fine-tune these models are not easily accessible. Interestingly, segmentation methods developed for retinal vessels datasets could be good candidates for the segmentation of coronary angiography data, since the images present similar features [187]. The use of these pre-trained models could potentially overcome the limitation posed by the scarcity of labelled coronary angiography data.

In addition to the factors discussed earlier, there are other crucial aspects that demand consideration when assessing medical images. Among these, two noteworthy features are the level of noise and the image sharpness, both of which can substantially influence the comprehensibility and utility of clinical data. Specifically, the presence of noise corresponds to variations in pixel intensity, which can obscure details and structures in the image, affecting the data interpretability. Image sharpness refers to the clarity and definition of edges and fine details within an image. In the medical context, images with poor sharpness can lead to misinterpretations, with severe consequences for the patient care. Therefore, this aspect should be an integral part of image quality assessment in the clinical context.

Furthermore, it is important to verify that the artificially generated data have a positive effect on the diagnostic process. A possible solution to solve this ambiguity is the involvement of medical experts in the development of the evaluation procedure. Specifically, doctors can provide their opinion on the generated data by assigning quality scores to the different video frames. This approach is also useful for the identification of possible inconsistencies and biases that could affect the algorithms. It is worth mentioning that, for clinical applications, the identification of suitable assessment metrics is not as straightforward as it is for the FIB-SEM example. Indeed, for the dataset analyzed in Section 5.2.2, features such as network porosity and tortuosity can be easily computed to achieve an exhaustive description of information retrievable from the data, which affects the materials' properties. In the healthcare context, the desirable image characteristics depend on the specific purpose of the procedures. Notably, coronary angiography can serve different purposes [188].

5.5 Conclusions and outlook

This chapter demonstrates that a state-of-the-art neural network, developed for video-frame interpolation, can be used to increase the resolution of image sequences in 3D tomography. This can be applied, without further training, across different length scales, going from a few nanometers to millimeters, and to the most diverse types of samples. As the main benchmark, a dataset of images of printed graphene nanostructured networks has been considered, obtained with the destructive FIB-SEM-NT technique. For this, computer-vision metrics have been carefully evaluated, but most importantly the quality of the information content that can be extracted from the 3D reconstruction has been investigated. In particular, the porosity, tortuosity, and effective diffusivity of the original dataset have been computed. The outcomes of this evaluation were then compared to datasets where an increasing number of images were removed and replaced with computer-generated ones.

In general, this study demonstrates that motion-aware video-frame interpolation outperforms any other interpolation strategies. In particular, it has been shown that it is not prone to image blurring, typical of simple linear interpolation, or to resolution loss at the image boundaries, as shown by some hybrid optical-flow algorithms. Most importantly, the error on the determination of morphological observables, such as the porosity, remains below 2 % as long as the milling thickness is less than approximately half of the nanosheet length. This suggests a very favourable experimental condition, where the effects of the milling on the measured morphology are significantly mitigated. Further investigations in this context will involve the application of RIFE to other FIB-SEM-generated datasets, made of different materials, such as printed WS_2 nanosheets and Silver nanosheets. The results presented in this work suggest that the performance should not change when considering materials with similar nanosheet lengths. In fact, the RIFE HD model shows excellent performance on datasets never seen before, without being fine-tuned. Additionally, more complex network features should be investigated to compare the different interpolation methods. Some examples are nanosheet alignment and connectivity [170].

The analysis was also extended to datasets taken from the medical field. These include a 3D tomography of the human brain volume, acquired with magnetic resonance imaging, an X-ray computed tomography of the human torso, and coronary angiography videos. In the first case, a ground truth was available and allowed us to show that the estimate of the gray-matter volume variation is only affected by 0.5 %, when half of the images in a scan are replaced by video-frame interpolated ones. This suggests that the scan rate can be actually reduced during the experimental procedure, saving acquisition time. In contrast,

for the CT scan, no ground truth was available. Therefore, other metrics have been estimated. In particular, it has been shown that data augmentation with computer-interpolated images, improves the power spectrum of the tomographic reconstruction, confirming the visual impression of smoother images. This result may be potentially transformative, since it can pave the way for reduced scan rates, with the consequent reduction in the radiation dose to be administrated to the patient. For both the described healthcare applications, further studies are necessary to quantify the improvement provided by the synthetically generated data. For instance, organ segmentation, using specialized tools, might be performed on the original and augmented datasets and compared through metrics such as the Dice similarity coefficient [114].

The last investigated application, in the medical context, concerns the use of interpolation techniques to increase the frame rate of coronary angiography videos, aiming to reduce the ionizing radiation exposure. Comparative analyses involving the methods RIFE, RRIN, and linear interpolation are conducted on quality assurance test objects and clinically acquired video frames. For both cases, visual evaluations show that RIFE maintains better object movement representation. RRIN results also demonstrate good interpolation capability. However, it is also associated with systematic image smoothing, which can cause flickering in the videos and interfere with the clinical interpretation of the data. Importantly, the interpolation speed is also analyzed, with outcomes that indicate RIFE as the most suitable method for real-time applications with a GPU. In the clinical dataset, discrepancies between visual evaluation and quantitative metrics (MSE, PSNR, SSIM, MS SSIM, FID, and WDS) suggest that the chosen metrics might not fully capture clinical preferences or nuances. Future investigations are planned, including the examination of different frame rates and alternative test objects like the NEMA phantom [184]. The analysis will also focus on evaluating specific image features in clinical data, considering factors such as noise and image sharpness.

In general, for all the explored medical applications, the aim of this research project is only to be a proof of concept, to demonstrate the feasibility of the proposed approach. Further investigations are needed before claiming a practical use in the clinical context. Additional studies should involve applications to data generated with different instrumentation, in order to ensure the model's ability to generalize well on data that might have different characteristics.

Moreover, the opinion of medical professionals is crucial to assess the information content retrievable from the data. This can be achieved through surveys, where the medical experts are exposed to a set of artificially generated images and are asked to grade their quality according to a specified scale. Despite being a subjective evaluation, this approach is widely employed in the healthcare field, to guarantee that the model results are in accordance with human experts. In

fact, medical practitioners could be able to identify artefacts and biases affecting the algorithms, not easily recognizable by computer vision metrics. Lastly, for all medical applications, in order to have an accurate analysis and identify appropriate metrics, it is advisable to define a specific purpose for which the data is used. This allows us to assess the model's performance in generating information that is useful for a precise purpose.

In addition to all the proposed investigations, the effect that different imaging conditions have on the performance of the RIFE algorithm should be explored, for all the applications. For instance, the input data could present different brightness, contrast, and resolution levels, all factors that could affect the interpolation output.

In summary, this chapter has shown that video-frame interpolation techniques can be successfully applied to 3D tomography regardless of the acquisition experimental technique and of the nature of the specimen to image. This can improve practices when radiation-dose damage or the acquisition time are issues limiting the applicability of the method.

CHAPTER 6

CONCLUSIONS

ELECTRON microscopy plays a decisive role in a multitude of research fields and industrial applications, yielding invaluable insights and discoveries. However, despite its remarkable achievements, electron microscopy still faces challenges rooted in instrumentation limitations and specimen fragility, which can affect the quality of the retrieved information. Overcoming these limitations traditionally required substantial investments in hardware advancements.

This research project explored an alternative strategy, that is steadily gaining more recognition in the field of electron microscopy and that does not necessitate hardware modifications. This approach is based on the use of machine-learning techniques to enhance imaging capabilities. Furthermore, this project extended its scope beyond materials science and electron microscopy, with applications in the domain of medical imaging instruments.

Throughout the entire work, particular effort was devoted to the development of accurate and relevant validation methods, to assess the results objectively. This can be particularly challenging when dealing with visual data enhancement, especially in the medical context.

The first electron-microscopy-related problem tackled in this work concerns one of the main limitations of Scanning Transmission Electron Microscopes (STEMs). These are powerful imaging instruments, that allow achieving the highest resolution of all electron microscopes, below 0.1 nm. However, this capability typically necessitates the use of high electron doses, a practice that can induce specimen damage and compromise observation quality, mainly due to knock-on and radiolysis damage mechanisms. These detrimental factors can be reduced when the beam intensity is decreased. However, this approach hampers the prospect of extracting

valuable insights from the data. This is due to the presence of Poisson noise, an effect, related to the quantized nature of the electron beam, that cannot be corrected at the instrumentation level and that increases when the number of incident electrons is reduced. In the context of modern imaging instruments and digital acquisition, other types of noise can be disregarded. To overcome this challenge and be able to examine beam-sensitive materials, this work proposed a machine learning-based strategy designed to noticeably enhance the quality of STEM data acquired at low electron doses. Trained on a dataset of simulated images, which reproduce realistic data acquisition scenarios, the model was tested on both synthetic and experimental data. The results of these tests clearly demonstrated an enhanced image quality and the ability to extract valuable information from the data without any bias. Several dose levels were available for the experimental images, representing a Gold nanoparticle deposited on an amorphous Carbon substrate, acquired in digital mode. This allowed an investigation of the model's performance at different levels of complexity. The primary distinction among the displayed reconstructions lies in the morphology of the individual atoms, which becomes increasingly more spherical with higher doses. This serves as evidence of the algorithm's impartiality in generating spherical atoms. Moreover, the residual (i.e. the difference between the original noisy and the reconstructed images) and the Fast Fourier Transform of the residual are presented, to demonstrate the features of the removed Poisson noise. Quantitive assessments of the results have been performed on simulated data, which allows the comparison of the investigated metrics with ground truth results. Firstly, the line profile analysis was performed on noise-free, noisy, and reconstructed images of TePb, simulated at a dose of 1,000 $e^-/\text{Å}^2$. The comparison of the three intensity profiles along a defined scan line demonstrates that the outcome of the proposed denoising approach maintains the same content of information of the ground truth (namely the original noise-free image), in terms of locations and discernment between distinct atomic species. Additionally, a workflow was proposed to compare the precision of atomic column localization in datasets obtained with different strategies, which are validated against the ground truth values. Atomic column localization is crucial for extracting structural information and quantifying lattice strain, which impacts material properties. The investigation was performed on simulated images of Tellurene, at different dose levels, within the range 500-10,000 $e^-/\text{Å}^2$. The results of this study show that the strain error, both in the horizontal and vertical directions, is dose-dependent, with noisy images exhibiting higher errors. The denoising autoencoder significantly improves strain error, offering adaptability to different dose levels. In contrast, Gaussian filtering, a commonly used solution for denoising, fails to enhance column localization except at very high doses.

Future developments of this project involve a potential integration of the

algorithm into the live data acquisition. This would be useful to assist microscopy users during the experimental procedure. In fact, based on the denoised data, the user would be able to adjust the implemented electron dose. Several aspects of the developed method would make it suitable for this pursuit. Firstly, the denoising process for a 128×128-pixel image can be completed in just about one second. Additionally, the proposed scheme operates autonomously and does not rely on human input or specific knowledge of electron dose.

Perspective studies will also focus on the development of quantitative validation methods to further assess the results achieved on experimental images. In this case, full-reference metrics cannot be used, due to the lack of ground-truth data. Among others, one possible solution is to involve expert microscopists to grade the quality of the artificially generated images, according to specified quality scores.

A comprehensive exploration of the autoencoder's latent space is crucial to obtain a deep understanding of the model's performance and to gain valuable insights into the denoising process. Analyzing the latent space can reveal the clustering and distribution of encoded features, disclosing the model's ability to capture essential information while filtering out noise. The study of the relationships between encoded representations and their corresponding denoised images has the potential to offer a clear understanding of how the model converts noisy inputs into cleaner outputs.

In summary, utilizing this model could lead to a significant reduction in the required electron dose for experimental acquisitions, enabling the analysis of highly beam-sensitive materials that would otherwise be challenging to study.

The application of the developed model to analog data, also affected by Gaussian noise led to the development of the next project. The main goal of this next project was to facilitate the defects quantification of STEM-acquired images of transition metal dichalcogenides (TMDs) by using a denoising autoencoder, whose capabilities have been confirmed by the results presented in Chapter 3. The model is here applied to experimental analog data, acquired with two different microscopes (Nion and Titan), which can achieve different resolutions. The presented preliminary results are obtained on MoS_2 samples, produced following the liquid phase exfoliation preparation method. The denoised images were subsequently processed using the software Atomap [144] to locate atomic columns belonging to the different sublattices. The percentage of vacancies was then calculated based on the deviation from the expected number of chalcogen atoms in an ideal lattice. Results indicated an average vacancy percentage of approximately 4 % for the Nion dataset and about 6 % for the Titan dataset. However, these findings are influenced by various factors impacting vacancy counting. The neural network-based denoising algorithm, while effective, could potentially introduce inaccuracies when applied to noisy images. To assess the impact of this source of error, future investi-

gations should include a comparison of vacancy identification between the original noise-free images and the denoised version, for simulated data. Furthermore, the precision of Atomap's atom localization could be evaluated by comparing its results with those obtained from alternative software like StatSTEM [135]. The utilization of the wraparound lattice approximation, designed to account for edge effects, may also introduce inaccuracies.

Future research will extend the study to other TMDs such as WS_2 and $PtSe_2$ and explore differences in vacancy counts between samples prepared using different methods, including mechanical exfoliation. It's important to note that the current methodology does not distinguish vacancies in different layers of the material. To address this, a threshold-based approach will be considered. Lastly, digitally acquired data should be investigated to simplify vacancy identification by eliminating Gaussian noise as an obstacle to atom localization.

Another electron microscopy-related problem was addressed in this work, namely, the anisotropic resolution of FIB-SEM-generated 3D tomography. FIB-SEM imaging technique is instrumental in the investigation of nanostructured networks. However, the lack of isotropic resolution in the reconstructed volumes can hinder the analysis of the morphology of these materials. For this project, an interdisciplinary approach was followed. In fact, a neural network developed for video frame interpolation was used to generate additional frames between existing ones, which allows for achieving cubic voxels in the volume reconstruction. The application of this strategy on a dataset made of networks of printed graphene nanosheets was presented, together with a meticulous analysis of the results, according to different strategies. Comprehensive assessments included computer-vision metrics and a focus on the quality of information extracted from 3D reconstructions, such as porosity, tortuosity, and effective diffusivity. These features were evaluated on datasets generated using different interpolation methods, which are compared throughout the analysis. It should be noted that all the used metrics are full-reference. In fact, frames were removed from the original dataset to be used as ground truth.

The results indicate that motion-aware video-frame interpolation, like the chosen RIFE model, outperforms other interpolation methods. It effectively mitigates issues like image blurring and resolution loss at image boundaries, making it particularly advantageous for improving morphological measurements. Even when milling thickness is a significant factor, this technique maintains an error below 2 % for the porosity evaluation, indicating its suitability for challenging experimental conditions.

Importantly, the exploration of a diversified set of validation metrics led to the conclusion that standard computer-vision metrics can lead to imprecise results. For instance, the interpolation method ISOFLOW appears to be better performing than

another video frame interpolation neural network, namely DAIN, according to the evaluation of both MSE and SSIM metrics. However, an examination of properties specific to the analyzed system, such as the network porosity, demonstrated that the results are actually reversed for these two approaches. As a consequence, an inspection of other sample-related features should be performed, in order to sustain the proposed approach. Some examples include the study of the alignment and connectivity of the nanostructured networks, which highly affect the properties of the investigated materials. Future research endeavours will also extend the application of this approach to other FIB-SEM-generated datasets, acquired with samples made of different materials.

Notably, the investigation was not limited to FIB-SEM data, and the method's versatility was demonstrated across various length scales, from nanometers to millimeters, and diverse sample types. In fact, the described approach was further extended to healthcare applications, including Magnetic Resonance Imaging (MRI) of the human brain, X-ray Computed Tomography (CT) of the torso, and coronary angiography videos. In the case of MRI, the study revealed a 0.5 % impact on the estimate of gray-matter volume variation when replacing half of the scan images with interpolated ones, potentially leading to shorter scan times. This test was made possible by the availability of ground-truth data, namely the frames removed every two existing frames from the original dataset, described by cubic voxels. In contrast, in the case of X-ray CT, the ground truth was not available. Therefore, the assessment of the results was achieved by comparing the noise power spectrum of the original and the artificially-augmented datasets, evaluated after 3D reconstruction. This comparison indicated a noise reduction in the latter case, which supports the perceived smoother image appearance. As a consequence, the patient's radiation exposure could be potentially reduced.

Likewise, in the case of the application of RIFE on coronary angiography videos, the ultimate goal is to limit the ionizing radiation delivered not only to the patient but also to the medical practitioner performing the procedure in the same room. For this study, two different types of datasets were investigated, videos acquired on a quality assurance test object and clinical data, both obtained at various frame rates. Comparative analyses have been conducted, involving methods such as RIFE, RRIN, and linear interpolation. Visual evaluations have demonstrated that RIFE excels in preserving object movement representation. RRIN also exhibits good interpolation capabilities but introduces systematic image smoothing, potentially causing flickering in the resulting video and interfering with clinical data interpretation. Additionally, the study includes an assessment of interpolation speed, which suggests that RIFE is the most suitable method for real-time applications with GPU support. For both datasets, the visual comparison was followed by a more quantitative evaluation, comprising full-reference metrics such as MSE, PSNR,

SSIM, MS SSIM, FID, and WDS. In the case of the test object videos, the results are consistent with the visual assessment of the generated video frames, depicting RIFE as the best-performing model. This outcome was not experienced in the case of clinically acquired videos of human patients. Indeed, according to this analysis, RRIN is the most advantageous of all methods, despite the defects identified from visual comparison. Therefore, different evaluation metrics should be addressed, in order to fully capture the features of this kind of data. As discussed throughout the entire work, establishing appropriate metrics for the evaluation of image quality is a challenging task, especially in the medical context. For instance, image sharpness and noise should be considered, characteristics that could highly affect the clinical interpretation.

In general, for all the described medical applications, additional assessment procedures should be explored. Specifically, the input of medical experts should be considered, through surveys and quality grading systems. Indeed, medical practitioners may have the ability to identify artefacts and biases that affect algorithms, which might not be easily recognizable through computer vision metrics. Furthermore, the assessment procedures should be defined according to a specific medical purpose, which could facilitate the identification of meaningful metrics. Despite the need for further inspection in the medical application, this part of the research work demonstrated the successful implementation of video-frame interpolation techniques to enhance 3D tomography across diverse scales and sample types.

To conclude, this research project has effectively demonstrated the successful application of neural networks to enhance today's capabilities in the field of electron microscopy, also extended to the medical imaging areas. Thanks to these tools, imaging efficiency can be improved, in terms of the range of material that can be analyzed, the speed of the acquisition process, and the quality of the retrieved information. Importantly, once the models are trained, they operate autonomously, meaning that they do not require input from the users. This makes them accessible to individuals from diverse backgrounds, even those without training in data science or machine learning.

The ultimate goal of this work is to make a contribution toward the integration of machine-learning techniques into imaging systems, with a particular focus on electron microscopy instruments, in order to assist and advance the imaging power. This integration has the potential to open doors to exciting new discoveries in various domains.

BIBLIOGRAPHY

[1] Ruska, E. The development of the electron microscope and of electron microscopy. *Biosci. Rep.* **7**, 607–629 (1987).

[2] Brydson, R. *Aberration-corrected analytical transmission electron microscopy* (Wiley Online Library, 2011).

[3] Valdrè, U. *Electron microscopy in material science* (Elsevier, 2012).

[4] Richert-Pöggeler, K. R., Franzke, K., Hipp, K. & Kleespies, R. G. Electron microscopy methods for virus diagnosis and high resolution analysis of viruses. *Front. Microbiol.* **9**, 3255 (2019).

[5] Klang, V., Valenta, C. & Matsko, N. B. Electron microscopy of pharmaceutical systems. *Micron* **44**, 45–74 (2013).

[6] Haider, M. *et al.* Electron microscopy image enhanced. *Nature* **392**, 768–769 (1998).

[7] Egelman, E. H. The current revolution in cryo-EM. *Biophys. J.* **110**, 1008–1012 (2016).

[8] Quigley, F., McBean, P., O'Donovan, P., Peters, J. J. & Jones, L. Cost and capability compromises in STEM instrumentation for low-voltage imaging. *Microsc. Microanal.* **28**, 1437–1443 (2022).

[9] Treder, K. P., Huang, C., Kim, J. S. & Kirkland, A. I. Applications of deep learning in electron microscopy. *Microscopy* **71**, i100–i115 (2022).

[10] Goodfellow, I., Bengio, Y., Courville, A. & Bengio, Y. *Deep learning*, vol. 1 (MIT press Cambridge, 2016).

[11] Cheng, H.-T. *et al.* Wide & deep learning for recommender systems. In *Proceedings of the 1st workshop on deep learning for recommender systems*, 7–10 (2016).

[12] Suta, P., Lan, X., Wu, B., Mongkolnam, P. & Chan, J. H. An overview of machine learning in chatbots. *Int. J. Mech. Eng. Robot. Res.* **9**, 502–510 (2020).

[13] Chauhan, K. Virtual assistant: A review. *Int. J. Res. Eng. Sci. Manag.* **3**, 138–140 (2020).

[14] Gangavarapu, T., Jaidhar, C. & Chanduka, B. Applicability of machine learning in spam and phishing email filtering: review and approaches. *Artif. Intell. Rev.* **53**, 5019–5081 (2020).

[15] Khan, A. A., Laghari, A. A. & Awan, S. A. Machine learning in computer vision: a review. *EAI Endorsed Trans. Scalable Inf.* **8**, e4–e4 (2021).

[16] Ede, J. M. Deep learning in electron microscopy. *Mach. Learn.: Sci. Technol.* (2020).

[17] Miotto, R., Wang, F., Wang, S., Jiang, X. & Dudley, J. T. Deep learning for healthcare: review, opportunities and challenges. *Brief. Bioinformatics* **19**, 1236–1246 (2018).

[18] Thompson, P. M. *et al.* Dynamics of gray matter loss in Alzheimer's disease. *J. Neurosci.* **23**, 994–1005 (2003).

[19] Yu, L. *et al.* Radiation dose reduction in computed tomography: techniques and future perspective. *Imaging Med.* **1**, 65 (2009).

[20] Davidson, M. W. Pioneers in optics: Zacharias Janssen and Johannes Kepler. *Microscopy Today* **17**, 44–47 (2009).

[21] Carpenter, W. B. & Dallinger, W. H. *The microscope and its revelations*, vol. 1 (J. & A. Churchill, 1891).

[22] Abbe, E. Beiträge zur theorie des mikroskops und der mikroskopischen wahrnehmung. *Archiv für mikroskopische Anatomie* **9**, 413–468 (1873).

[23] Vangindertael, J. *et al.* An introduction to optical super-resolution microscopy for the adventurous biologist. *Methods Appl. Fluoresc.* **6**, 022003 (2018).

[24] Stephens, D. J. & Allan, V. J. Light microscopy techniques for live cell imaging. *Science* **300**, 82–86 (2003).

[25] Wang, Z. L. New developments in transmission electron microscopy for nanotechnology. *Adv. Mater.* **15**, 1497–1514 (2003).

[26] Vladár, A. E., Postek, M. T. & Ming, B. On the sub-nanometer resolution of scanning electron and helium ion microscopes. *Microscopy Today* **17**, 6–13 (2009).

[27] Murphy, D. B. & Davidson, M. W. *Fundamentals of light microscopy and electronic imaging* (John Wiley & Sons, 2012).

[28] Rosenauer, A., Krause, F. F., Müller, K., Schowalter, M. & Mehrtens, T. Conventional transmission electron microscopy imaging beyond the diffraction and information limits. *Phys. Rev. Lett.* **113**, 096101 (2014).

[29] Kirkland, E. J. *Advanced computing in electron microscopy* (Springer, 1998).

[30] Pryor, A., Ophus, C. & Miao, J. A streaming multi-GPU implementation of image simulation algorithms for scanning transmission electron microscopy. *Adv. Struct. Chem. Imaging* **3**, 1–14 (2017).

[31] Ophus, C. A fast image simulation algorithm for scanning transmission electron microscopy. *Adv. Struct. Chem. Imaging* **3**, 1–11 (2017).

[32] Ponce, A., Mejía-Rosales, S. & José-Yacamán, M. Scanning transmission electron microscopy methods for the analysis of nanoparticles. In *Nanoparticles in Biology and Medicine*, 453–471 (Springer, 2012).

[33] Pennycook, S. J. & Nellist, P. D. *Scanning transmission electron microscopy: imaging and analysis* (Springer Science & Business Media, 2011).

[34] Jones, L. Quantitative ADF STEM: acquisition, analysis and interpretation. *IOP Conf. Ser.: Mater. Sci. Eng.* **109**, 012008 (2016).

[35] Egerton, R., Li, P. & Malac, M. Radiation damage in the TEM and SEM. *Micron* **35**, 399–409 (2004).

[36] Egerton, R. Control of radiation damage in the TEM. *Ultramicroscopy* **127**, 100–108 (2013).

[37] Lee, Z., Rose, H., Lehtinen, O., Biskupek, J. & Kaiser, U. Electron dose dependence of signal-to-noise ratio, atom contrast and resolution in transmission electron microscope images. *Ultramicroscopy* **145**, 3–12 (2014).

[38] Jones, L. & Nellist, P. D. Identifying and correcting scan noise and drift in the scanning transmission electron microscope. *Microsc. Microanal.* **19**, 1050 (2013).

[39] Mullarkey, T., Downing, C. & Jones, L. Development of a practicable digital pulse read-out for dark-field STEM. *Microsc. Microanal.* **27**, 99–108 (2021).

[40] Lin, R., Zhang, R., Wang, C., Yang, X.-Q. & Xin, H. L. TEMImageNet and AtomSegNet deep learning training library and models for high-precision atom segmentation, localization, denoising, and super-resolution processing of atom-resolution scanning TEM images. *arXiv preprint arXiv:2012.09093* (2020).

[41] Cowley, J. M. & Moodie, A. F. The scattering of electrons by atoms and crystals. i. a new theoretical approach. *Acta Crystallogr.* **10**, 609–619 (1957).

[42] Ashcroft, N. W. & Mermin, N. D. *Solid state physics* (Cengage Learning, 2022).

[43] Van Der Hoeven, J. E. *et al.* Bridging the gap: 3D real-space characterization of colloidal assemblies via FIB-SEM tomography. *Nanoscale* **11**, 5304–5316 (2019).

[44] Gabbett, C. *et al.* 3D-imaging of printed nanostructured networks using high-resolution FIB-SEM nanotomography (2023). URL `https://arxiv.org/abs/2301.11046`.

[45] Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. methods* **9**, 676–682 (2012).

[46] Object Research Systems (ORS) Inc., M. Dragonfly 3.1 (computer software) (2016). Available online: `http://www.theobjects.com/dragonfly`.

[47] Wirth, R. Focused Ion Beam (FIB) combined with SEM and TEM: Advanced analytical tools for studies of chemical composition, microstructure and crystal structure in geomaterials on a nanometre scale. *Chem. Geol.* **261**, 217–229 (2009).

[48] Reyntjens, S. & Puers, R. A review of focused ion beam applications in microsystem technology. *J. Micromech. Microeng.* **11**, 287 (2001).

[49] Uchic, M. D., Groeber, M. A., Dimiduk, D. M. & Simmons, J. 3D microstructural characterization of nickel superalloys via serial-sectioning using a dual beam FIB-SEM. *Scr. Mater.* **55**, 23–28 (2006).

[50] Gu, L., Wang, N., Tang, X., Changela, H. *et al.* Application of FIB-SEM techniques for the advanced characterization of earth and planetary materials. *Scanning* **2020** (2020).

[51] Munroe, P. R. The application of focused ion beam microscopy in the material sciences. *Mater. Charact.* **60**, 2–13 (2009).

[52] Grandfield, K., Engqvist, H. *et al.* Focused ion beam in the study of biomaterials and biological matter. *Adv. Mater. Sci. Eng.* **2012** (2012).

[53] Xu, C. S. *et al.* Enhanced FIB-SEM systems for large-volume 3D imaging. *elife* **6**, e25916 (2017).

[54] Berlinski, D. *The advent of the algorithm: the 300-year journey from an idea to the computer* (Houghton Mifflin Harcourt, 2001).

[55] Buchanan, B. G. A (very) brief history of artificial intelligence. *Ai Magazine* **26**, 53–53 (2005).

[56] Nilsson, N. J. *The quest for artificial intelligence* (Cambridge University Press, 2009).

[57] Nadikattu, A. K. R. Influence of artificial intelligence on robotics industry. *Int. J. Creat. Res. Thoughts* 2320–2882 (2021).

[58] Yu, K.-H., Beam, A. L. & Kohane, I. S. Artificial intelligence in healthcare. *Nat. Biomed. Eng.* **2**, 719–731 (2018).

[59] Cao, L. Ai in finance: A review. *Available at SSRN 3647625* (2020).

[60] Tahiru, F. AI in education: A systematic literature review. *J. Cases Inf. Technol.* **23**, 1–20 (2021).

[61] Schmidt, R. M., Schneider, F. & Hennig, P. Descending through a crowded valley-benchmarking deep learning optimizers. In *International Conference on Machine Learning*, 9367–9376 (PMLR, 2021).

[62] Vincent, P. *et al.* Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11** (2010).

[63] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process.* **25** (2012).

[64] Parihar, A. S., Varshney, D., Pandya, K. & Aggarwal, A. A comprehensive survey on video frame interpolation techniques. *Vis. Comput.* 1–25 (2021).

[65] Jin, M., Hu, Z. & Favaro, P. Learning to extract flawless slow motion from blurry videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8112–8121 (2019).

[66] Wu, Y., Wen, Q. & Chen, Q. Optimizing video prediction via video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17814–17823 (2022).

[67] Huang, Z., Zhang, T., Heng, W., Shi, B. & Zhou, S. Real-time intermediate flow estimation for video frame interpolation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (2022).

[68] Parihar, A. S., Varshney, D., Pandya, K. & Aggarwal, A. A comprehensive survey on video frame interpolation techniques. *Vis. Comput.* 1–25 (2022).

[69] Bao, W. *et al.* Depth-aware video frame interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3703–3712 (2019).

[70] Jiang, H. *et al.* Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 9000–9008 (2018).

[71] Park, J., Lee, C. & Kim, C.-S. Asymmetric bilateral motion estimation for video frame interpolation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14539–14548 (2021).

[72] Practical-RIFE. `https://github.com/hzwer/Practical-RIFE`. Accessed: 2023-06-30.

[73] Xue, T., Chen, B., Wu, J., Wei, D. & Freeman, W. T. Video enhancement with task-oriented flow. *Int. J. Comput. Vis.* **127**, 1106–1125 (2019).

[74] Real-Time Intermediate Flow Estimation for Video Frame Interpolation. `https://github.com/megvii-research/ECCV2022-RIFE`. Accessed: 2023-06-30.

[75] Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).

[76] Hou, Q., Ghildyal, A. & Liu, F. A perceptual quality metric for video frame interpolation. In *European Conference on Computer Vision*, 234–253 (Springer, 2022).

[77] Dost, S. *et al.* Reduced reference image and video quality assessments: review of methods. *Eurasip J. Image Video Process.* **2022**, 1–31 (2022).

[78] Kourtis, M.-A., Koumaras, H. & Liberal, F. Reduced-reference video quality assessment using a static video pattern. *J. Electron. Imaging* **25**, 043011–043011 (2016).

[79] Shahid, M., Rossholm, A., Lövström, B. & Zepernick, H.-J. No-reference image and video quality assessment: a classification and review of recent approaches. *Eurasip J. Image Video Process.* **2014**, 1–32 (2014).

[80] Saad, M. A., Bovik, A. C. & Charrier, C. Blind prediction of natural video quality. *IEEE Trans. Image Process.* **23**, 1352–1365 (2014).

[81] You, J. & Korhonen, J. Deep neural networks for no-reference video quality assessment. In *2019 IEEE International Conference on Image Processing (ICIP)*, 2349–2353 (IEEE, 2019).

[82] Le Callet, P., Viard-Gaudin, C. & Barba, D. A convolutional neural network approach for objective video quality assessment. *IEEE Trans. Neural Netw. Learn. Syst.* **17**, 1316–1327 (2006).

[83] Hasegawa, B. H. *The physics of medical x-ray imaging* (Medical Physics Pub. Co., 1990).

[84] Costello, J. E., Cecava, N. D., Tucker, J. E. & Bau, J. L. CT radiation dose: current controversies and dose reduction strategies. *AJR Am. J. Roentgenol.* **201**, 1283–1290 (2013).

[85] Johnson, T., Fink, C., Schönberg, S. O. & Reiser, M. F. *Dual energy CT in clinical practice*, vol. 201 (Springer, 2011).

[86] Beister, M., Kolditz, D. & Kalender, W. A. Iterative reconstruction methods in X-ray CT. *Phys. Med.* **28**, 94–108 (2012).

[87] Hara, A. K. *et al.* Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. *AJR Am. J. Roentgenol.* **193**, 764–771 (2009).

[88] Withers, P. J. *et al.* X-ray computed tomography. *Nat. Rev. Methods Primers* **1**, 18 (2021).

[89] Carmignato, S., Dewulf, W. & Leach, R. *Industrial X-ray computed tomography*, vol. 10 (Springer, 2018).

[90] Miller, D. L. Overview of contemporary interventional fluoroscopy procedures. *Health Phys.* **95**, 638–644 (2008).

[91] Baim, D. S. & Grossman, W. Coronary angiography. In *Cardiac catheterization and angiography. Third edition* (Lippincott Williams & Wilkins, 1986).

[92] Schoepf, U. J., Becker, C. R., Ohnesorge, B. M. & Yucel, E. K. CT of coronary artery disease. *Radiology* **232**, 18–37 (2004).

[93] Tavakol, M., Ashraf, S. & Brener, S. J. Risks and complications of coronary angiography: a comprehensive review. *Glob. J. Health Sci.* **4**, 65 (2012).

[94] Togni, M. *et al.* Percutaneous coronary interventions in Europe 1992–2001. *Eur. Heart J.* **25**, 1208–1213 (2004).

[95] Kim, K. P. & Miller, D. L. Minimising radiation exposure to physicians performing fluoroscopically guided cardiac catheterisation procedures: a review. *Radiat. Prot. Dosim.* **133**, 227–233 (2009).

[96] Protection, R. ICRP publication 103. *Ann ICRP* **37**, 2 (2007).

[97] Lee, J.-W. *et al.* Adaptively variable frame-rate fluoroscopy with an ultra-fast digital x-ray tube based on carbon nanotube field electron emitters. In *Medical Imaging 2020: Physics of Medical Imaging*, vol. 11312, 860–865 (SPIE, 2020).

[98] Van Geuns, R.-J. M. *et al.* Basic principles of magnetic resonance imaging. *Prog. Cardiovasc. Dis.* **42**, 149–156 (1999).

[99] Sartoretti, E. *et al.* Impact of acoustic noise reduction on patient experience in routine clinical magnetic resonance imaging. *Acad. Radiol.* **29**, 269–276 (2022).

[100] Paul, D. *et al.* Artificial intelligence in drug discovery and development. *Drug Discov. today* **26**, 80 (2021).

[101] Gourisaria, M. K., Das, S., Sharma, R., Rautaray, S. S. & Pandey, M. A deep learning model for malaria disease detection and analysis using deep convolutional neural networks. *Int. J. Emerg. Technol. Learn.* **11**, 699–704 (2020).

[102] Spyropoulos, C. D. AI planning and scheduling in the medical hospital environment (2000).

[103] Bekfani, T. *et al.* A current and future outlook on upcoming technologies in remote monitoring of patients with heart failure. *Eur. J. Heart Fail.* **23**, 175–185 (2021).

[104] Haidegger, T., Speidel, S., Stoyanov, D. & Satava, R. M. Robot-assisted minimally invasive surgery—surgical robotics in the data age. *Proc. IEEE* **110**, 835–846 (2022).

[105] Barragán-Montero, A. *et al.* Artificial intelligence and machine learning for medical imaging: A technology review. *Phys. Med.* **83**, 242–256 (2021).

[106] Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. Image reconstruction by domain-transform manifold learning. *Nature* **555**, 487–492 (2018).

[107] Maier, J., Sawall, S., Knaup, M. & Kachelrieß, M. Deep scatter estimation (DSE): Accurate real-time scatter estimation for X-ray CT using a deep convolutional neural network. *J. Nondestruct. Eval.* **37**, 1–9 (2018).

[108] Johnson, P. M., Recht, M. P. & Knoll, F. Improving the speed of MRI with artificial intelligence. In *Seminars in musculoskeletal radiology*, vol. 24, 012–020 (Thieme Medical Publishers, 2020).

[109] Han, C. *et al.* GAN-based synthetic brain MR image generation. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, 734–738 (IEEE, 2018).

[110] Timins, J. Radiation during pregnancy. *New Jersey medicine: the journal of the Medical Society of New Jersey* **98**, 29–33 (2001).

[111] Choudhury, A. & Asan, O. Impact of accountability, training, and human factors on the use of artificial intelligence in healthcare: Exploring the perceptions of healthcare practitioners in the US. *Human Factors in Healthcare* **2**, 100021 (2022).

[112] Palmisciano, P., Jamjoom, A. A., Taylor, D., Stoyanov, D. & Marcus, H. J. Attitudes of patients and their relatives toward artificial intelligence in neurosurgery. *World Neurosurg.* **138**, e627–e633 (2020).

[113] Šimundić, A.-M. Measures of diagnostic accuracy: basic definitions. *Electron. J. Int. Fed. Clin. Chem. Lab. Med.* **19**, 203 (2009).

[114] Zou, K. H. *et al.* Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Acad. Radiol.* **11**, 178–189 (2004).

[115] Ruthotto, L. & Haber, E. An introduction to deep generative modeling. *GAMM-Mitteilungen* **44**, e202100008 (2021).

[116] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Adv. Neural Inf. Process.* **30** (2017).

[117] Fréchet, M. Sur la distance de deux lois de probabilité. In *Annales de l'ISUP*, vol. 6, 183–198 (1957).

[118] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826 (2016).

[119] Vaserstein, L. N. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii* **5**, 64–72 (1969).

[120] Dolly, S., Chen, H.-C., Anastasio, M., Mutic, S. & Li, H. Practical considerations for noise power spectra estimation for clinical CT scanners. *J. Appl. Clin. Med. Phys.* **17**, 392–407 (2016).

[121] Verdun, F. *et al.* Image quality in CT: From physical measurements to model observers. *Phys. Med.* **31** (2015).

[122] Gambini, L., Mullarkey, T., Jones, L. & Sanvito, S. Machine-learning approach for quantified resolvability enhancement of low-dose STEM data. *Mach. Learn.: Sci. Technol.* **4**, 015025 (2023).

[123] Thakur, K. V., Damodare, O. H. & Sapkal, A. M. Poisson noise reducing bilateral filter. *Procedia Comput. Sci.* **79**, 861–865 (2016).

[124] Du, H. A nonlinear filtering algorithm for denoising HR (S) TEM micrographs. *Ultramicroscopy* **151**, 62–67 (2015).

[125] Meyer, J. C. *et al.* Experimental analysis of charge redistribution due to chemical bonding by high-resolution transmission electron microscopy. *Nat. Mater.* **10**, 209–215 (2011).

[126] Mevenkamp, N. *et al.* Poisson noise removal from high-resolution STEM images based on periodic block matching. *Adv. Struct. Chem. Imaging* **1**, 1–19 (2015).

[127] Lin, R., Zhang, R., Wang, C., Yang, X.-Q. & Xin, H. L. TEMImageNet training library and atomsegnet deep-learning models for high-precision atom segmentation, localization, denoising, and deblurring of atomic-resolution images. *sr* **11**, 1–15 (2021).

[128] Ziatdinov, M., Ghosh, A., Wong, C. Y. & Kalinin, S. V. AtomAI framework for deep learning analysis of image and spectroscopy data in electron and scanning probe microscopy. *Nat. Mach. Intell.* **4**, 1101–1112 (2022).

[129] Lobato, I., Friedrich, T. & Van Aert, S. Deep convolutional neural networks to restore single-shot electron microscopy images. *arXiv preprint arXiv:2303.17025* (2023).

[130] Kohl, H. A simple procedure for evaluating effective scattering cross-sections in stem. *Ultramicroscopy* **16**, 265–268 (1985).

[131] Chollet, F. *et al.* Keras. `https://keras.io` (2015).

[132] Abadi, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems (2015). URL `https://www.tensorflow.org/`. Software available from tensorflow.org.

[133] Krivanek, O. L. *et al.* Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy. *Nature* **464**, 571–574 (2010).

[134] Han, Y. *et al.* Strain mapping of two-dimensional heterostructures with subpicometer precision. *Nano Lett.* **18**, 3746–3751 (2018).

[135] De Backer, A., Van den Bos, K., Van den Broek, W., Sijbers, J. & Van Aert, S. StatSTEM: an efficient approach for accurate and precise model-based quantification of atomic resolution electron microscopy images. *Ultramicroscopy* **171**, 104–116 (2016).

[136] van Dyck, D. High-resolution electron microscopy. *Adv. Imaging Electron Phys.* **123**, 105–171 (2002).

[137] van der Walt, S. *et al.* the scikit-image contributors. *Scikit-image: image processing in Python. PeerJ* **2**, e453 (2014).

[138] Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).

[139] Choi, W. *et al.* Recent development of two-dimensional transition metal dichalcogenides and their applications. *Mater. Today* **20**, 116–130 (2017).

[140] Lee, J. *et al.* Electrical role of sulfur vacancies in MoS2: Transient current approach. *Appl. Surf. Sci.* **613**, 155900 (2023).

[141] Huang, Y. *et al.* Universal mechanical exfoliation of large-area 2D crystals. *Nat. Commun.* **11**, 2453 (2020).

[142] Hernandez, Y. *et al.* High-yield production of graphene by liquid-phase exfoliation of graphite. *Nat. Nanotechnol.* **3**, 563–568 (2008).

[143] Wu, Z. & Ni, Z. Spectroscopic investigation of defects in two-dimensional materials. *Nanophotonics* **6**, 1219–1237 (2017).

[144] Nord, M., Vullum, P. E., MacLaren, I., Tybell, T. & Holmestad, R. Atomap: a new software tool for the automated analysis of atomic resolution images using two-dimensional Gaussian fitting. *Adv. Struct. Chem. Imaging* **3**, 1–12 (2017).

[145] Madsen, J. *et al.* A deep learning approach to identify local structures in atomic-resolution transmission electron microscopy images. *Adv. Theory Simul.* **1**, 1800037 (2018).

[146] Prismatic. `http://https://prism-em.com/docs-params/`. Accessed: 2023-08-16.

[147] Atomap. `https://atomap.org/api_documentation.html`. Accessed: 2023-08-22.

[148] Chetih, N. & Messali, Z. Tomographic image reconstruction using filtered back projection (FBP) and algebraic reconstruction technique (ART). In *2015 3rd International Conference on Control, Engineering & Information Technology (CEIT)*, 1–6 (IEEE, 2015).

[149] Silva, A. C., Lawder, H. J., Hara, A., Kujak, J. & Pavlicek, W. Innovations in CT dose reduction strategy: application of the adaptive statistical iterative reconstruction algorithm. *AJR Am. J. Roentgenol.* **194**, 191–199 (2010).

[150] Tan, D., Jiang, C., Li, Q., Bi, S. & Song, J. Silver nanowire networks with preparations and applications: a review. *J. Mater. Sci.: Mater. Electron.* **31**, 15669–15696 (2020).

[151] Carey, T. *et al.* Inkjet printed circuits with 2D semiconductor inks for high-performance electronics. *Adv. Electron. Mater.* **7**, 2100112 (2021).

[152] Zhou, L., Fan, M., Hansen, C., Johnson, C. R. & Weiskopf, D. A review of three-dimensional medical image visualization. *Health Data Science* **2022**, 840519 (2022).

[153] Verdun, F. R. *et al.* Quality initiatives radiation risk: what you should know to tell your patient. *Radiographics* **28**, 1807–1816 (2008).

[154] González-Solares, E. A. *et al.* Imaging and molecular annotation of xenographs and tumours (IMAXT): High throughput data and analysis infrastructure. *Biological Imaging* **3**, e11 (2023).

[155] Roldán, D., Redenbach, C., Schladitz, K., Klingele, M. & Godehardt, M. Reconstructing porous structures from FIB-SEM image data: Optimizing sampling scheme and image processing. *Ultramicroscopy* **226**, 113291 (2021).

[156] González-Ruiz, V., García-Ortiz, J. P., Fernández-Fernández, M. & Fernández, J. J. Optical flow driven interpolation for isotropic FIB-SEM reconstructions. *Comput. Methods Programs. Biomed.* **221**, 106856 (2022).

[157] Nixon, M. & Aguado, A. *Feature extraction and image processing for computer vision* (Academic press, 2019).

[158] Hagita, K., Higuchi, T. & Jinnai, H. Super-resolution for asymmetric resolution of FIB-SEM 3D imaging using AI with deep learning. *sr* **8**, 1–8 (2018).

[159] Dahari, A., Kench, S., Squires, I. & Cooper, S. J. Fusion of complementary 2D and 3D mesostructural datasets using generative adversarial networks. *Adv. Energy Mater.* **13**, 2202407 (2023).

[160] Bladt, E., Pelt, D. M., Bals, S. & Batenburg, K. J. Electron tomography based on highly limited data using a neural network reconstruction technique. *Ultramicroscopy* **158**, 81–88 (2015).

[161] Ding, G., Liu, Y., Zhang, R. & Xin, H. L. A joint deep learning model to recover information and reduce artifacts in missing-wedge sinograms for electron tomography and beyond. *Sci. Rep.* **9**, 12803 (2019).

[162] Lavery, L., Harris, W., Bale, H. & Merkle, A. Recent advancements in 3D X-ray microscopes for additive manufacturing. *Microsc. Microanal.* **22**, 1762–1763 (2016).

[163] Lim, C., Yan, B., Yin, L. & Zhu, L. Geometric characteristics of three dimensional reconstructed anode electrodes of lithium ion batteries. *Energies* **7**, 2558–2572 (2014).

[164] Kelly, A. G., O'Suilleabhain, D., Gabbett, C. & Coleman, J. N. The electrical conductivity of solution-processed nanosheet networks. *Nat. Rev. Mater.* **7**, 217–234 (2022).

[165] Nicks, J., Sasitharan, K., Prasad, R. R., Ashworth, D. J. & Foster, J. A. Metal–organic framework nanosheets: programmable 2D materials for catalysis, sensing, electronics, and separation applications. *Adv. Funct. Mater.* **31**, 2103723 (2021).

[166] Konkena, B. *et al.* Liquid processing of interfacially grown Iron-Oxide flowers into 2D-platelets yields Lithium-Ion battery anodes with capacities of twice the theoretical value. *Small* **18**, 2203918 (2022).

[167] Arganda-Carreras, I. *et al.* Trainable Weka segmentation: a machine learning tool for microscopy pixel classification. *Bioinform.* **33**, 2424–2426 (2017).

[168] Ridler, T., Calvard, S. *et al.* Picture thresholding using an iterative selection method. *IEEE Trans. Syst. Man Cybern* **8**, 630–632 (1978).

[169] Sara, U., Akter, M. & Uddin, M. S. Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. *J. Comput. Commun.* **7**, 8–18 (2019).

[170] Gabbett, C. *Electrical, Mechanical & Morphological Characterisation of Nanosheet Networks*. Ph.D. thesis, Trinity College Dublin (2021).

[171] Cooper, S. J., Bertei, A., Shearing, P. R., Kilner, J. & Brandon, N. P. Taufactor: An open-source application for calculating tortuosity factors from tomographic data. *SoftwareX* **5**, 203–210 (2016).

[172] Tjaden, B., Brett, D. J. & Shearing, P. R. Tortuosity in electrochemical devices: a review of calculation approaches. *Int. Mater. Rev.* **63**, 47–67 (2018).

[173] TauFactor. `https://github.com/tldr-group/taufactor`. Accessed: 2023-06-30.

[174] Tadel, F. *et al.* MEG/EEG group analysis with brainstorm. *Front. Neurosci.* 76 (2019).

[175] Brainstorm. `http://neuroimage.usc.edu/brainstorm`. Accessed: 2023-06-30.

[176] Shattuck, D. W. & Leahy, R. M. Brainsuite: an automated cortical surface identification tool. *Med. Image Anal.* **6**, 129–142 (2002).

[177] Nakazawa, T. *et al.* Multiple-region grey matter atrophy as a predictor for the development of dementia in a community: the Hisayama study. *J. Neurol. Neurosurg. Psychiatry* **93**, 263–271 (2022).

[178] Clark, K. *et al.* The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* **26**, 1045–1057 (2013).

[179] Rutherford, M. *et al.* A DICOM dataset for evaluation of medical image de-identification (Pseudo-PHI-DICOM-Data) [data set]. The Cancer Imaging Archive (2021).

[180] Yin, X.-l. *et al.* Analysis of coronary angiography video interpolation methods to reduce x-ray exposure frequency based on deep learning. *Cardiovasc. Innov. App.* **6**, 17–24 (2021).

[181] Li, H., Yuan, Y. & Wang, Q. Video frame interpolation via residue refinement. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2613–2617 (IEEE, 2020).

[182] Cowen, A., Haywood, J., Workman, A. & Clarke, O. A set of x-ray test objects for image quality control in digital subtraction fluorography. i: design considerations. *Brit. J. Radiol.* **60**, 1001–1009 (1987).

[183] Crummy, A. B. *et al.* Computerized fluoroscopy: digital subtraction for intravenous angiocardiography and arteriography. *AJR Am. J. Roentgenol.* **135**, 1131–1140 (1980).

[184] Simon, R. *et al.* Criteria to optimise a dynamic flat detector system used for interventional radiology. *Radiat. Prot. Dosim.* **129**, 261–264 (2008).

[185] Ramesh, K., Kumar, G. K., Swapna, K., Datta, D. & Rajest, S. S. A review of medical image segmentation algorithms. *EAI Endorsed Trans.* **7**, e6–e6 (2021).

[186] Kaba, Ş., Haci, H., Isin, A., Ilhan, A. & Conkbayir, C. The application of deep learning for the segmentation and classification of coronary arteries. *Diagnostics* **13**, 2274 (2023).

[187] Shin, S. Y., Lee, S., Yun, I. D. & Lee, K. M. Deep vessel segmentation by learning graphical connectivity. *Med. Image Anal.* **58**, 101556 (2019).

[188] Garrone, P. *et al.* Quantitative coronary angiography in the current era: principles and applications. *J. Interv. Cardiol.* **22**, 527–536 (2009).