

Christian-Albrechts-Universität zu Kiel

DEPARTMENT OF COMPUTER SCIENCE

Master Thesis

**Denoising 3D TEM tomography via Advanced
Neural Radiance Fields**

Mithun Das

M.Sc in Computer Science

SUPERVISORS

Prof. Dr. Peer Kröger
Dr.-Ing. Claudius Zelenka

Kiel, Germany

09.01.2024

Abstract

This thesis is dedicated to improving the reconstruction of 3D images from 2D Transmission Electron Microscopy (TEM) data, utilizing Advanced Neural Radiance Fields (NeRF). The main goal is to enhance image quality and denoising capabilities, tackling the inherent challenges in TEM image processing. The research identifies and addresses a significant limitation of NeRF's application in this context, which can potentially make NeRF an ineffective approach. To counter this, we have adopted an architectural framework that integrates multiple NeRF models with a blend of traditional and advanced denoising techniques to improve image quality. Our findings show marked improvements in the quality of TEM images, supported by enhanced evaluation metrics.

Acknowledgments

I would like to express my deepest gratitude to **Prof. Dr. Peer Kröger** and **Dr.-Ing. Claudius Zelenka** for their invaluable guidance and unwavering support throughout the course of this thesis work. Their expertise and insights have been instrumental in shaping both the direction and success of this endeavor. I also extend special thanks to **Dr.-Ing. Niklas Wolff** for providing valuable datasets crucial for this thesis.

Additionally, my profound thanks go to my family, friends, and my girlfriend **Smruthi** for the unending love and support. Their constant encouragement and faith in my abilities have been a foundational source of strength and inspiration throughout this journey.

Declaration

I declare that I have produced the Master's thesis **Denoising 3D TEM tomography via Advanced Neural Radiance Fields** independently and without improper external assistance and that I have identified all word-for-word quotations of other authors, as well as comments based closely on other authors' ideas, and I have listed the relevant sources.

I am aware that, unless agreed otherwise, the Master's thesis produced under supervision represents a group achievement and forms part of the overall research of the supervising institution. As a result, none of the co-authors (e.g., authors of text, creative project staff, co-supervisors) may use passages from the thesis for commercial purposes or make them accessible to third parties without the (written) approval of all those involved due to reasons of copyright. Particular note must be taken of the Arbeitnehmerfindergesetz (German Employee Invention Act), according to which pre-publication of patent-related content is prohibited.

Place, Date

Name of candidate

Contents

Abstract	iii
Acknowledgments	v
Declaration	vii
1 Introduction	1
2 Background Information	5
2.0.1 Transmission electron microscopy(TEM)	5
2.0.2 Neural 3D shape representations	7
2.0.3 Novel view Synthesis	8
2.0.4 Camera Parameters	9
2.0.5 COLMAP	10
2.0.6 NeRF (Neural Radiance Field)	12
2.0.7 Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN)	14
3 Related Work	15
3.0.1 Inverse problem	15
3.0.2 Atomic Resolution	17
3.0.3 Noise Modeling	18
3.0.4 3D Convolutional Neural Networks	19
3.0.5 Denoising	20
4 Datasets	21
4.0.1 TEM Datasets	21
4.0.2 STEM Datasets	26
4.0.3 Synthetic Dataset	28
5 Methodology	29
5.0.1 Original Neural Radiance Fields (NeRF)	29
5.0.2 Extending NeRF to Handle Transmission Electron Microscopy (TEM) Data	30
5.0.3 Post-Reconstruction Processing with ESRGAN	33
5.0.4 Integrating Traditional Denoising Techniques	34
5.0.5 Adapted Framework for TEM Data Analysis	35
6 Implementation	37
6.0.1 Setup Overview	37

6.0.2	Key Implementation Steps	37
6.0.3	Code Availability	39
7	Experiments & Results	41
7.1	Preliminary Evaluation of NeRFStudio	41
7.1.1	Utilization of Custom Datasets in NeRFStudio	41
7.1.2	Scaling Dataset Size	42
7.2	Exploration of TEM Image Experiments	43
7.3	Exploring Alternative Nerf Models for TEM Images	44
7.4	Exploration of COLMAP Alternatives for TEM Images	45
7.5	Experimental Analysis with TEM Dataset 3	46
7.5.1	Dataset 3: Training with 200 Epochs	46
7.5.2	Dataset 3: Training with 10,000 Epochs	47
7.5.3	Dataset 3: Camera Pose Optimization after 10,000 Epochs	48
7.5.4	Dataset 3: Exploring Different Activation Functions	49
7.5.5	Dataset 3 with different Network layer	51
7.5.6	Enhancing the Model with NeRF-Dark and NAN-NeRF Parameters	52
7.5.7	Postprocessing with Traditional Denoisers	53
7.5.8	Postprocessing with ESRGAN	56
7.6	Experimental Analysis with TEM Dataset 1	58
7.7	Experimental Analysis with TEM Dataset 2	59
7.8	Experimental Analysis with TEM Dataset 4	60
7.9	Experimental Analysis with STEM Dataset 1	61
7.10	Experimental Analysis with STEM Dataset 2	62
7.11	Experimental Analysis with Synthetic Dataset	63
7.12	Comparative PSNR & SSIM Analysis of Image Denoising	64
8	Conclusions	65
9	Future Scope	67
9.1	Advancements in NeRF Models	67
9.2	Extending Dataset Variety	67
9.3	Development of User-Friendly Software Tools	68
	Bibliography	69

CHAPTER 1

Introduction

Electron tomography, often known as ET, is now the method of choice for determining the three-dimensional ultra-structure of organelles and cells at nanoscale resolutions. The 3D volume of the specimen can be reproduced by first collecting a tilted set of 2D transmission electron microscopy (TEM) images over a wide-angle range (usually $\pm 60^\circ$ to 80°), and then computationally recombining the images. However, the low electron doses that are applicable to biological samples (which are typically less than $100 \text{ e}^-/\text{\AA}^2$) result in extremely poor signal-to-noise ratios (SNR) in the tomograms that are produced as a result [1]. The leading causes of the noise are the stochastic character of the events involving electron scattering and the constraints imposed by electron detection [2]. In addition, flaws in the alignment of the tilt axis, beam-induced specimen deformation, and distortions that are inherent to electron lenses all contribute to the contamination of the TEM data [2]. This leads to very tiny structural features, which are essential for interpreting the sophisticated cellular processes and chemical interactions, to get confused and distorted. Thus, one of the most important steps in the preprocessing stage that comes before extracting data with biological importance is lowering the level of noise in tomograms.

In order to improve 3D ET reconstructions, a number of different denoising algorithms have been implemented. To a certain extent, straightforward linear filters like median filtering, gaussian smoothing, and anisotropic diffusion filtration can reduce noise, but at the cost of a significant loss of high-resolution information [3]. More enhanced regularization approaches such as total variation (TV) minimization and sparse coding exploit image priors to preserve edges and the rigidity of an image. However, these methods frequently require extensive parameter tuning to strike a balance between the removal of noise and the absorption of detailed information. Deep learning models such as [4] have shown promise for 2D image denoising tasks. However, directly implementing such networks to tomograms slice-by-slice is unable to effectively exploit 3D contextual information and the spatial relationships between coordinates. Although several algorithms are capable of doing block-wise 3D denoising, they are restricted by computational restrictions [4]. Other methods involve training on simulated data, which might not translate well to tomograms taken from actual data.

While deep learning models like DnCNN [5] have shown promise for 2D image denoising tasks, directly applying such networks to tomograms slice-by-slice fails to fully utilize 3D contextual information and spatial relationships between voxels. Some methods perform block-wise 3D denoising but are limited by computational constraints when scaling

to large high-resolution volumes [6]. Other techniques pretrain on simulated data which may not generalize well to real experimental tomograms with complex noise textures [6]. Most existing deep learning approaches also lack interpretability into the learned features and struggle to denoise non-uniform noise distributions as encountered in practice.

More recently, neural radiance fields (NeRF) [7] have demonstrated unprecedented ability to synthesize photorealistic novel views of complex 3D scenes using a continuous volumetric representation. NeRFs learn a 5D radiance field where each 3D coordinate (X,Y,Z) is mapped to an emitted color (R,G,B) and volume density σ using a standard multilayer perceptron (MLP). The key advantages of NeRF over other 3D deep learning representations are:

- Coordinate-based MLPs better capture local spatial relationships compared to convolutional networks.
- Continuous scene modeling enables synthesizing views from arbitrary poses.

In this work, we propose adapting NeRF for directly denoising 3D volumes reconstructed from tilt series electron tomography (ET) data. By training on pairs of noisy input and clean target volumes, the NeRF model may learn to infer superior denoised outputs closely matching the ground truth. The continuous volumetric modeling could outperform other 3D networks that lack such inductive bias. This approach could significantly enhance interpretability of structural details from electron tomograms.

However, directly applying NeRF for reconstructing volumes from noisy TEM tilt series poses significant challenges. The COLMAP structure-from-motion algorithm is utilized by the standard NeRF pipeline in order to perform camera pose estimation for each input image. The high noise levels of TEM projections, on the other hand, can make it difficult for COLMAP to properly establish the viewing angles. Because of this, it is challenging to train NeRF directly on raw TEM pictures that contain noise. Within the scope of this research, we suggest alterations to the NeRF framework that, if implemented, will make it possible to obtain a more accurate camera pose estimation from noisy TEM tilt series, where COLMAP fails to provide any camera poses.

This study modifications to the NeRF framework to enable more robust camera pose estimation from noisy TEM tilt series. We also investigate training strategies and loss formulations to better condition the model on the noise characteristics of real ET data. This includes using robust loss functions that focus on structure rather than pixel intensities, as well as adaptive sampling and conditioning schemes. By adapting NeRF to handle noisy inputs in this manner, we aim to overcome the limitations of standard NeRF applied directly out-of-the-box to electron tomography volumes. Our noise-aware NeRF model could open new possibilities for high-fidelity 3D denoising and analysis of ET reconstructions.

A visual comparison between the original TEM images and the outcomes produced by our suggested method in Figure 1.1. This contrast draws attention to the difficulties caused by the noise in the original data and emphasizes how our method can improve image quality and lead to more precise analysis. Our innovative denoising technique was motivated by the significant reduction in noise and improvement in clarity, as illustrated

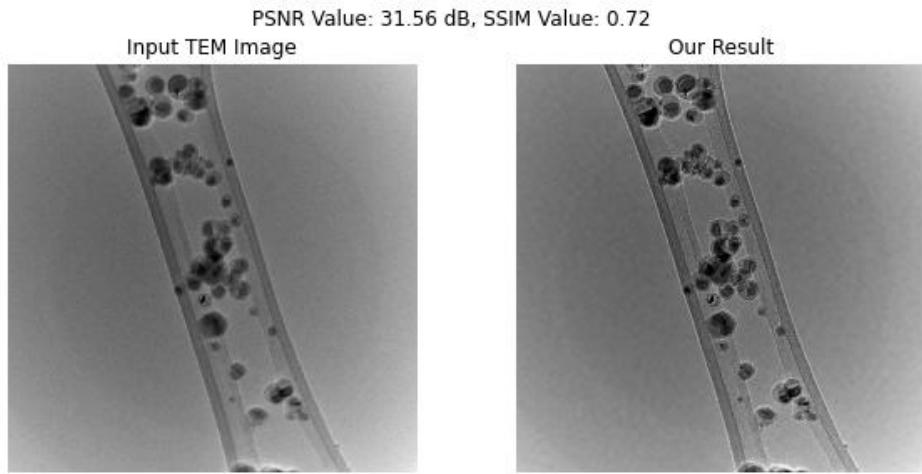


Figure 1.1: Comparative Analysis of Original TEM Images and Enhanced Outputs.

in the picture. It demonstrates the significant influence that efficient noise reduction can have on TEM picture interpretability, which is necessary for precise 3D reconstruction and later biological investigation.

CHAPTER 2

Background Information

2.0.1 Transmission electron microscopy(TEM)

A beam of electrons is used in transmission electron microscopy (TEM), which generates images of specimens with a resolution far higher than that of optical microscopes [8]. In transmission electron microscopy, electrons are emitted by a tungsten filament or field emission source and then accelerated under high voltage (typically 100-300 kV) [9]. Electromagnetic lenses concentrate the electron beam such that it is directed toward the extremely thin sample. When they go through the sample, electrons have a variety of interactions with the sample, depending on the density and the thickness of the material. This produces an electron diffraction pattern, which may be interpreted to reveal information about the structure of the material [10].

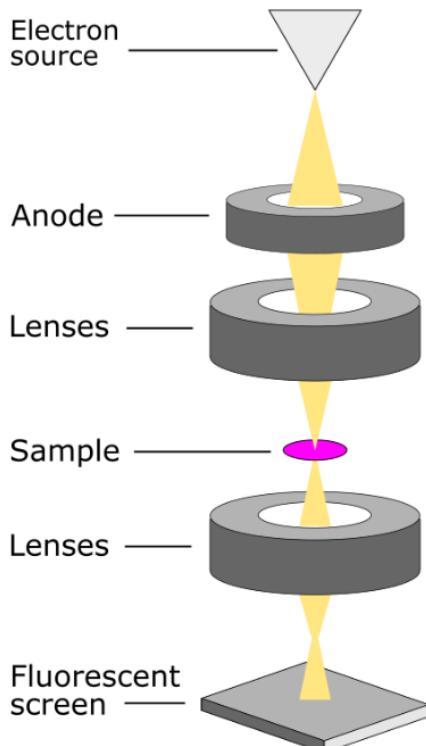


Figure 2.1: Schematic diagram of Transmission Electron Microscope
Source: <https://anapath.ch/electron-microscopy-2/>

Additional lenses concentrate the transmitted electrons so they may be captured as an image on a detector or camera [9]. The transmission electron microscope (TEM) may provide magnifications of up to 2 million times [9], which enables the viewing of structures and details on a scale as tiny as a nanometer or an angstrom. Because of this, it is an extremely useful instrument for study in materials science, cell biology, molecular structure analysis, and semiconductors [9]. Imaging mode and diffraction mode are the major modes of operation for the transmission electron microscope (TEM) [11]. The image that is created by the transmitted electrons is used by the imaging mode. It is possible to examine either the diffraction pattern or the image depending on how the magnetic lenses are adjusted. The electron diffraction patterns are the primary focus of the diffraction mode, which focuses on the crystal structure [11]. The preparation of samples is an essential part of TEM. To facilitate electron transmission, specimens must have a thickness of between 50 and 100 nanometers (nm)[11]. Staining with substantial amounts of heavy metal salts is required for biological and polymer materials to produce contrast. Imaging of hydrated materials is possible because of specialized methods such as cryo-TEM, which vitrifies the samples [11].

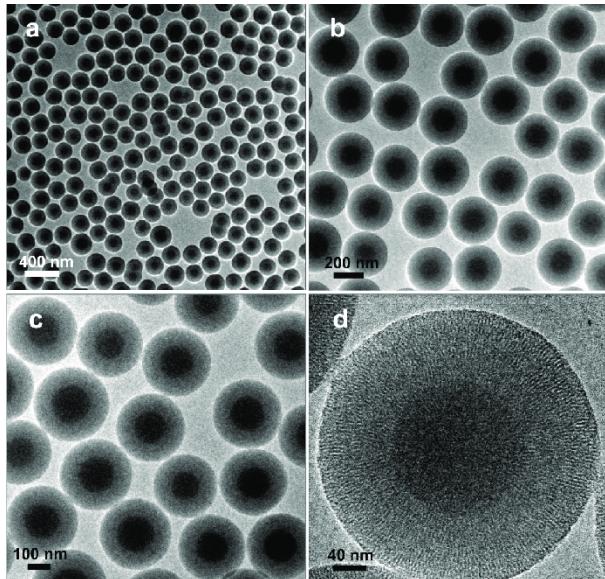


Figure 2.2: TEM images of a mesoporous material(MCM-41) at different magnifications [12]

There is a possibility that radiation will destroy sensitive specimens, which is one of the TEM's limitations [8]. Imaging of living biological samples is likewise not possible due to the vacuum environment. Nevertheless, TEM continues to be an essential instrument for high-resolution structural characterization in both the physical and biological sciences [8]. In this study, transmission electron microscopy was used to examine Janus-like particles that were created from block copolymers. TEM gives the resolution and contrast necessary to clearly examine the nanostructure morphology and surface topology of the Janus particles [10].

2.0.2 Neural 3D shape representations

Major innovations in deep learning have enabled neural networks to automatically represent and render complex 3D shapes 2.3, which was not previously feasible [13]. Neural implicit models can effectively represent shapes by mapping 3D coordinates to occupancy probabilities, signed distance values, or view-dependent radiance [14]. This contrasts with traditional explicit surface and volumetric representations like meshes and voxels.

Early works focused on learning continuous signed distance functions for representing 3D surfaces on synthetic shape datasets [15]. Subsequent techniques aimed to relax the dependence on 3D supervision by formulating differentiable rendering losses that could be optimized using only 2D images [16]. However, these approaches were limited to simplistic and smooth shapes.



Figure 2.3: DeepSDF’s 3D bunny shape representation [13]

More recently, coordinate-based neural radiance fields have achieved significant improvements in modeling complex real-world 3D scenes [7]. By representing scene properties like volume density and view-dependent emitted radiance as continuous 5D functions, novel photorealistic views of intricate scenes can be rendered.

In this work, we investigate leveraging the capabilities of modern neural 3D representations to reconstruct and denoise volumes from transmission electron microscopy tilt series. By training these networks to map from noisy TEM observations to cleaner target volumes in a self-supervised manner, they may learn specialized priors relevant to electron microscopy. Coordinate-based modeling may also better capture critical local context compared to other 3D approaches. This could significantly enhance the interpretability of fine structural details from TEM tomograms.

2.0.3 Novel view Synthesis

The act of creating fresh photographic perspectives on a subject from one or more input photos is called **View synthesis** 2.4. This may be done with either a single image or many images. This allows the create unique synthetic viewpoints using only a small amount of photographic data. View synthesis is useful in a variety of contexts, including virtual reality, augmented reality, and the reconstruction of three-dimensional models[17]. Many different techniques have been used for view synthesis. The multi-view stereo approach builds a three-dimensional reconstruction of a scene by piecing together a few photographs obtained with a variety of cameras [18][17]. Then, this model may be displayed from any perspective. Image-based rendering distorts and interpolates pixels depending on the original inputs to infer new viewpoints [19]. These methods concentrate on identifying correspondences between different pictures.

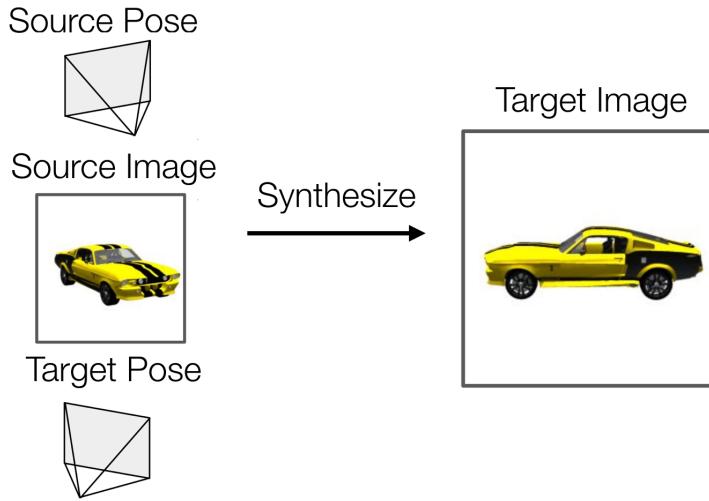


Figure 2.4: Multi-view to Novel view synthesis[20]

The most recent deep learning algorithms develop an implicit representation of the image generation process using neural networks. The neural rendering algorithms directly produce unique views by making predictions about the values of pixels based on the attributes of the scene that they have learned. Neural radiance fields (NeRF) [7] is a method for efficiently encoding a scene as a continuous five-dimensional function that maps three-dimensional coordinates to volume density and view-dependent brightness[7]. The continuous volumetric scene representation that NeRF provides has made it possible to do photorealistic view synthesis with only a few photos. The capacity to implicitly infer a three-dimensional structure and appearance from just two-dimensional supervision is the primary benefit offered by neural view synthesis systems. Because of this, formal three-dimensional modeling or estimate is not required. These learning-based systems continue to increase the realism and flexibility of new view creation across a wide variety of applications, including augmented reality, virtual tourism, and 3D photography [21].

2.0.4 Camera Parameters

The geometric and optical properties of a camera are referred to as its camera parameters. These parameters define how a camera constructs a picture from the 3D world [20]. Understanding the process of picture generation as well as the tasks involved in 3D computer vision, relies heavily on an accurate representation of these factors.

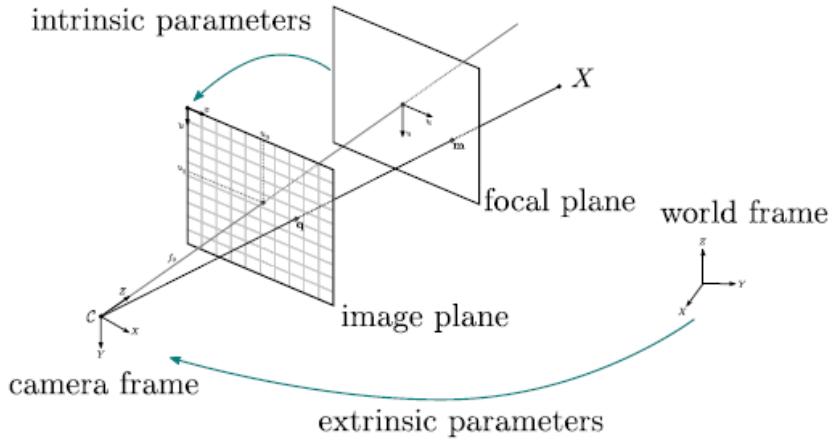


Figure 2.5: Schematic Representation of Camera Parameters. Source: OpenMVG Documentation.

Intrinsic Parameters are those that are unique to a camera and are not affected by the scene 2.5:

- **Focal Length (f)**: The distance from the optical center to the image plane when the image is sharp. A primary component that determines both the field of view and the magnification [22]. When dealing with non-square pixels, the x and y axes may have unique values.
- **Principal Point (c_x, c_y)**: The coordinates (c_x, c_y) of the image center on the sensor plane. It accounts for lenses that are not perfectly aligned [22].
- **Skew Coefficient (α)**: A rotation of the axis between the pixel grid and the sensor that considers non-rectangular pixel shapes [22]. Produces a shearing transformation when applied.
- **Distortion Coefficients**: This model simulates optical distortions such as radial, tangential, and narrow prism effects. Radial is the most noticeable and gives an impression like a barrel or pincushion [22].

The intrinsic camera matrix K can be represented as:

$$K = \begin{bmatrix} f_x & \alpha & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.1)$$

where f_x and f_y are the focal lengths expressed in pixel units.

Extrinsic Parameters are determined by the position of the camera in relation to the world 2.5:

- **Rotation Matrix (R):** A 3x3 matrix describing the camera's orientation in a world coordinate system [23]. Represented by a sequence of rotations (Euler angles).
- **Translation Vector (T):** A vector defining the position of the camera's origin in world coordinates [23].

The extrinsic parameters are combined into a 3x4 matrix $[R \mid T]$, defining the transformation from world coordinates to camera coordinates.

Together, intrinsic and extrinsic parameters form the camera projection matrix P , which is responsible for mapping 3D world points into 2D picture coordinates:

$$P = K \times [R|T] \quad (2.2)$$

For computer vision applications like posture estimation, 3D reconstruction, and unique view synthesis, accurate assessment of these parameters is essential.

Applications in augmented reality [24], autonomous navigation [24], and computational photography [24] heavily rely on precise camera calibration. Adapting camera models to new modalities, such as light field imaging, continues to be an active area of research.

2.0.5 COLMAP

COLMAP is an open-source pipeline that uses structure-from-motion (SfM) and multi-view stereo (MVS) to generate 3D models from 2D images [25]. Through solid correspondence construction, global optimization, and volumetric fusion, it features state-of-the-art reconstructions.

- **Feature Extraction and Matching**

First, appearance-based image features that can be paired between views are found and described 2.6. Based on local gradients, SIFT is frequently used to locate scale- and rotation-invariant key points [26]. Each key point has a high-dimensional descriptor vector that is insensitive to noise, perspective, and illumination [26]. Based on similarity measures like Euclidean or cosine distance, an effective closest

neighbor search matched characteristics between image pairings. Uncertain matches can be eliminated with the ratio test [26]. Outlier matches that are inconsistent with a single 3D point are eliminated by geometric verification using RANSAC [27].

- **Incremental Structure from Motion (SfM)**

In an incremental SfM method, the registered 2D-2D matches create initial sparse 3D point clouds [28]. An initial point cloud is plotted using an initial image pair. Which views to update next are efficiently chosen by robust visibility constraints [28]. With points recursively mapped from fresh views [29], camera poses are predicted using a Straight Linear Transform within a RANSAC cycle.

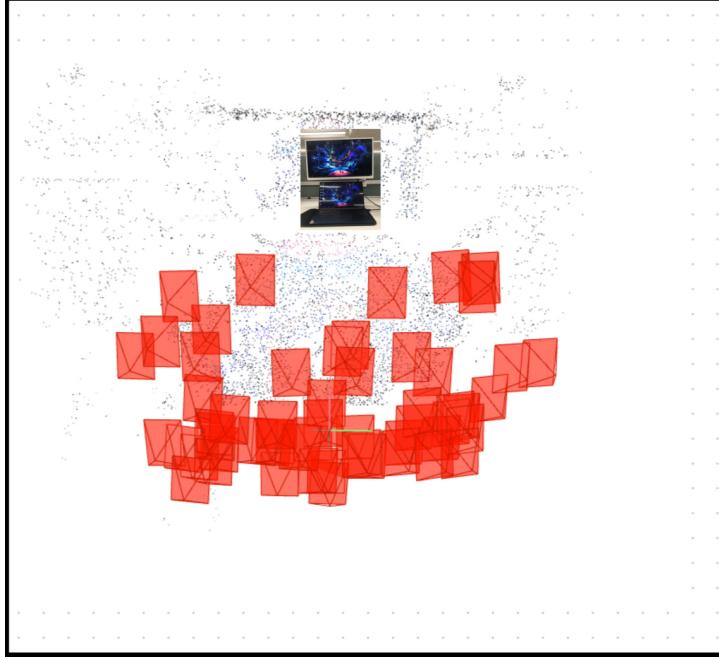


Figure 2.6: COLMAP camera position extracting from set of 2D images

- **Global SfM Optimization**

Utilizing bundle adjustment, the progressive reconstruction is globally improved to improve camera simultaneously poses and 3D point coordinates. Scale drift is reduced with regularization. Bundle adjustment reduces the top view error between the positions of anticipated and actual 2D features in all perspectives [30]. This enhances accuracy and comprehensiveness.

- **Multi-View Stereo (MVS) Depth Map Estimation**

The optimal cameras and points start the estimate of the multi-view stereo depth map. Using photo consistency metrics such as normalized cross correlation between distorted picture patches, dense correspondence is created in each view [31]. Accuracy is improved by regularization using filtering such as Gaussian smoothing [32]. The per-view depth maps that are constructed include geometric detail.

- **Surface Reconstruction**

When creating a final 3D surface mesh, volumetric fusion methods such as screening Poisson reconstruction are utilized to merge the depth data to produce the mesh [33]. It accomplishes this by interpolating an indicator function to provide a continuous and smooth surface. Additional post-processing steps, such as graph cuts-based optimization [34], may be utilized to improve details even further. Realism and color are added when texturing with the use of input images.

2.0.6 NeRF (Neural Radiance Field)

Neural radiance fields (NeRF) are a recent breakthrough technique for novel view synthesis and 3D scene modeling using implicit neural representations [7]. NeRF represents a scene as a continuous 5D radiance field (3D position + 2D viewing direction) using a multilayer perceptron (MLP).



Figure 2.7: Conceptual Illustration of Neural Radiance Fields (NeRF) [7]

Algorithm 1 Training Neural Radiance Field (NeRF)

```

1: Initialize MLP with random weights.
2: for each training iteration do
3:   Select a subset of input images.
4:   for each selected image do
5:     Sample a set of rays passing through image pixels.
6:     for each ray do
7:       Sample points along the ray.
8:       Query MLP for color and density at each point.
9:       Calculate the rendered color of the ray using volume rendering.
10:    end for
11:    Calculate loss between rendered and actual pixel colors.
12:    Update MLP weights to minimize the loss.
13:   end for
14: end for

```

The MLP maps each (x,y,z) location in space to an RGB color value and volume density scalar. The color indicates the emitted radiance, while density encodes occlusion. Querying this MLP at sampled points along camera rays enables volumetric ray marching to render novel views. The integral of density * color approximates the total radiance along each ray. Compared to discrete voxel grids or meshes, continuous coordinate-based modeling better captures smooth variations in structure, appearance, and lighting. The MLP can represent complex scenes in a memory-efficient compact latent code rather than an explicit 3D model. Adjusting MLP weights based on rendered and real image differences allows for optimizing the scene representation.

Key advantages of the NeRF approach include:

- Coordinate MLPs effectively model local relationships.
- Continuous representation enables high-quality view interpolation.
- Volume density handles complex occlusion effects.
- View-direction encoding models view-dependent phenomena like highlights.

NeRF obtains impressive results in reconstructing 3D scenes from only a sparse set of input views (e.g. 20-50 images). However, it relies on accurate camera pose estimation and clean photographic inputs. The utility for noisy domains like biomedical imaging is still being explored. For training purposes, you will need the camera parameters for the images that have been provided. These camera parameters are typically computed by SfM tools like COLMAP [7].

Self-Supervised Learning in NeRF for TEM Data

Self-supervised learning, a robust approach for training neural networks in the absence of labeled data, leverages the intrinsic structure of data to guide representation learning, especially in fields where manual annotations are scarce [35]. In our work, we explore this paradigm in the context of TEM data, focusing on a self-supervised framework specifically designed for neural radiance fields (NeRF) to address unique challenges in TEM imaging.

NeRF models are typically trained with posed 2D image datasets and associated camera parameters, utilizing view reconstruction losses for supervision [7]. However, in TEM, obtaining accurate poses is challenging due to issues like low signal-to-noise ratios and incomplete angular coverage, rendering standard structure-from-motion techniques ineffective. Our approach overcomes these hurdles by leveraging principles of geometry [36] and appearance consistency [36], replacing traditional supervision methods with self-supervision that relies on the geometric and appearance consistencies across input views.

In our proposed method, NeRF synthesizes target images from a TEM tilt series using a subset of the images, thus inferring a consistent 3D representation. The process involves random sampling of training directions, predicting images along held-out directions through volumetric raymarching, and using the MLP scene representation. The reconstruction loss between predicted and actual images serves as self-supervision for the model, enabling it to learn to predict complete volumes from sparse, noisy data. This method allows NeRF to implicitly learn priors suitable for electron tomography. Our experiments show that this self-supervised, coordinate-based approach effectively reconstructs higher fidelity volumes from limited input compared to other methods, highlighting its potential to address key computational challenges in TEM analysis [15].

2.0.7 Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN)

An overview of Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [37] shows that this technology is a major step forward in the super-resolution of images. This deep learning model is an essential tool in industries that need high-resolution imaging because it is made to boost image resolution beyond what the sensor can handle. ESRGAN is a development of the previous Super-Resolution GAN (SRGAN) [38], but it has undergone significant improvements that enable it to produce images with finer features and more realistic textures [39].

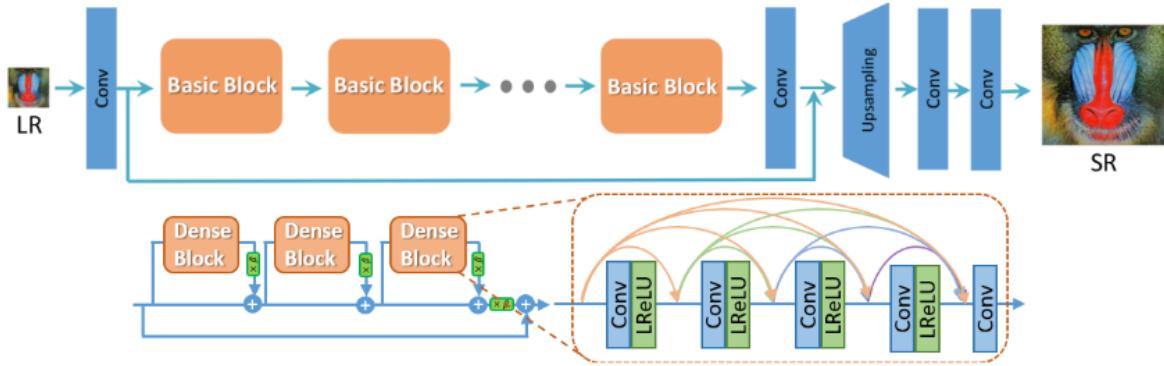


Figure 2.8: ESRGAN Architecture [37]

The implementation of Residual-in-Residual Dense Blocks (RRDB) to enhance the model's depth and parameters and make optimization simpler is one of ESRGAN's primary advantages over SRGAN. Additionally, it introduces the Relativistic GAN (Ra-GAN) loss, which helps to improve texture details and sharpen edges. Furthermore, ESRGAN uses a perceptual loss that makes use of pre-trained features from VGG networks, which improves the super-resolved images' perceived quality [37].

Especially useful in TEM context, ESRGAN may infer high-resolution images four times larger than the input [37]. The more realistic textures of the model are essential for improving medical images and computational microscopy. The ability of ESRGAN to synthesize fine details is utilized as a post-processing step for processed TEM data in our thesis study. This application is essential in order to improve the quality of TEM images, where the clarity of microscopic structures and details is crucial.

Our thesis tackles the problem of enhancing the resolution and perceived quality of TEM pictures through the implementation of ESRGAN. Our goal in using ESRGAN is to improve these images' diagnostic utility and aesthetic appeal so that they may be used for more in-depth examination and interpretation. One example of how cutting-edge deep learning techniques are being used in practice to improve scientific imaging is the incorporation of ESRGAN into the pipeline for processing TEM pictures. This method advances the science of microscopy while also establishing a standard for the use of super-resolution techniques in other fields where improved picture quality and resolution are necessary.

CHAPTER 3

Related Work

3.0.1 Inverse problem

The process of reconstructing 3D structures from TEM tilt series is a challenging and ill-posed inverse problem. In this process, a volume is reconstructed from finite and noisy 2D projections by inverting the advanced visualization model, which projects the 3D structure to the observed dimensions [40]. This allows the volume to be recovered from the 2D projections. Traditional algorithms, such as back projection, are incapable of fully addressing the ill-posedness of the problem because of its structural dimension.

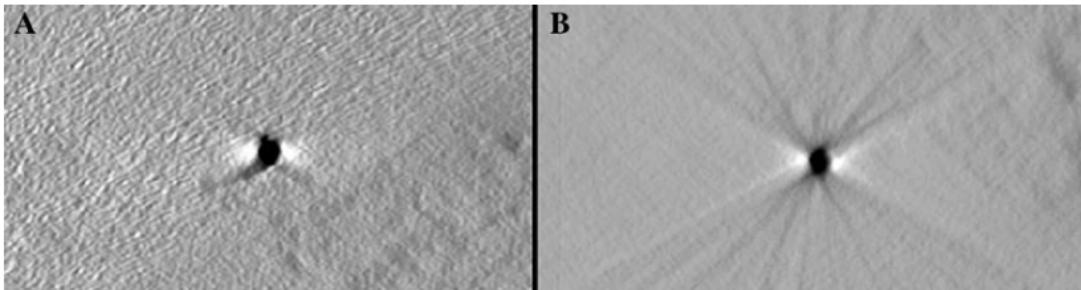


Figure 3.1: Comparative Reconstruction of a 10 nm Diameter Gold Particle: (A) Using Linear Backprojection in IMOD Software, showing deviation from spherical shape. (B) Using Curvilinear Backprojection in TxBR, accurately representing a spherical particle [41].

Regularized iterative reconstruction approaches have been created as a means of overcoming restrictions that are unique to analytical methods [42]. These strategies integrate past knowledge in order to limit the space available for solutions. Total variation (TV), a technique for regularization that creates reconstructions with smoothed intensity variations while keeping edges [42], is a technique that is extensively utilized. Compressive sensing is an additional method that makes use of sparsity in transform domains such as gradients or wavelets [43]. In addition, patch-based sparse coding methods have been utilized in order to discover an exhaustive collection of local basis functions for the purpose of denoising [44]. However, when the amount of data grows larger, these methods experience a rise in the computational burdens they must bear, and choosing the appropriate regularization parameters becomes a process that is not easy to do.

The inclusion of a missing wedge, which results from the limited tilt range, low signal-to-noise ratios, and high volumes of the 3D data [45], further complicates the TEM reconstruction process. In order to overcome these obstacles, more robust regularization strategies that are specifically adapted to the imaging physics of TEM need to be developed. Estimating the point spread function of the microscope using model-based methods has been suggested as a way to invert the blurring effects of the instrument [45].

To enhance the quality of the reconstruction, one method that does so by capitalizing on the self-similarity that exists between blocks is known as non-local means filtering [41]. In recent years, learning-based algorithms, such as dictionary learning and deep convolutional networks [44], have demonstrated promising results as post-processing filters or end-to-end reconstruction methods.

In the realm of TEM reconstruction, the development of more recent deep 3D representations, such as generative adversarial networks and neural radiance/volumetric fields [46], has made room for the introduction of new prospects. These methods involve training networks to map real noisy TEM projections to cleaner target volumes, which enables the networks to implicitly encode appropriate priors for accurate reconstruction. However, in order for these methods to work, a substantial amount of training data that covers the entire spectrum of viewing angles is required. Despite these developments, modeling the process of TEM image creation and adapting regularization limitations are still open difficulties that need to be addressed. There is a possibility that hybrid approaches, which combine model-based reconstruction with learned regularization techniques, could provide a more resilient solution.

3.0.2 Atomic Resolution

Atomic-resolution transmission electron microscopy (TEM) has emerged as a revolutionary technique for materials characterization by directly imaging individual atoms [47]. Modern TEMs use aberration correctors and monochromators that eliminate chromatic blurring and lens flaws to achieve sub-angstrom resolution [48]. Advanced detectors and highly stable devices have created new opportunities for quantitative analysis [49].

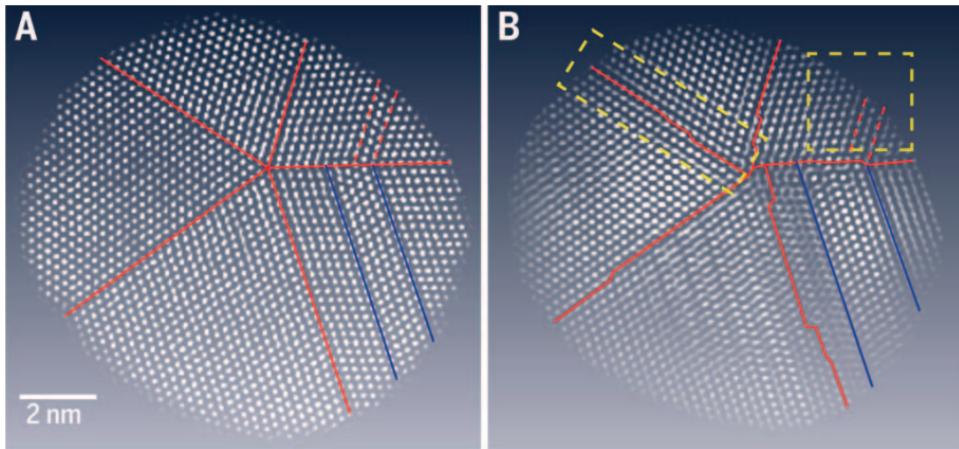


Figure 3.2: Atomic Resolution 3D Imaging of Platinum Nanoparticle: (A) STEM image showing flat twin boundaries. (B) AET reconstruction reveals atomic steps at twin boundaries, with a grain boundary and stacking fault [50].

On atomic structures, many imaging techniques offer complimentary information. Atomic number contrast images are created through high angle annular dark field (HAADF) imaging, where the intensity scales with Z^2 [50]. Elemental distributions are mapped by atomically resolved energy dispersive X-ray (EDX) spectroscopy. High resolution TEM can image light elements and even depict atom columns. For high precision data, scanning TEM (STEM) raster scans a focused probe [51]. High beam currents within a tiny probe are necessary for atomically detailed imaging, though. As a result, the electron dosages exceed what many materials can tolerate in terms of radiation [52]. The fundamentally probabilistic electron-sample interactions thus mask important atomic organization details with quantum noise [53]. Robust denoising techniques designed for TEM are necessary to obtain quantitative data.

Through statistical post-processing, techniques like multi-frame averaging [51] and principal component analysis enhance signal-to-noise. Sparse regularization techniques take advantage of structural redundancy to reduce noise. Compact representations for denoising are discovered through dictionary learning [50]. Convolutional neural networks have most recently demonstrated the ability to learn potent priors from atomistic image simulations [50]. In general, the developments that have been made in aberration corrected TEM have effectively actualized single-atom sensitivity and precision [53]. Researchers now have capabilities never before seen to unearth new insights through quantitative atomic-scale characterization. These skills are made possible with the assistance of specific denoising techniques, which help researchers overcome resolution restrictions imposed by noise.

3.0.3 Noise Modeling

Accurately modeling and characterizing noise is critical for developing effective reconstruction and denoising methods for transmission electron microscopy (TEM) images. Multiple studies have investigated the noise properties and sources in TEM.

- **Shot Noise**

One of the primary sources of noise is shot noise stemming from the quantum nature of electrons and the stochastic process of electron-sample interaction. Reconstruction and denoising techniques for transmission electron microscopy (TEM) images require precise noise characterization and modeling. The sources and characteristics of noise in TEM have been the subject of numerous investigations [54][55]. The number of electrons scattered from each part of the specimen fluctuates, leading to signal-dependent shot noise. Robust statistical distributions capture this behavior.

- **Detector Noise**

TEM detectors also introduce additional noise from readout electronics and amplification [56]. On CCD cameras, dark current shot noise and readout noise are present. Scintillator-photomultiplier detectors show signal-dependent Poisson noise characteristics [57]. Accurate detector models enable simulation of cumulative noise.

- **Beam Current Noise**

Fluctuations in beam current and brightness over time also contribute noise in TEM imaging [58]. Monitoring beam current during acquisition allows normalization to reduce this noise [58] [59]. But residual fluctuations persist and should be incorporated into models.

- **Noise Texture**

The microscope point spread function and optical transfer function modulate the texture of noise in the images. Accurately modeling these effects based on system parameters enables generating realistic synthetic noise for training machine learning models [60].

- **Multiresolution Modeling**

Noise also exhibits signal-dependency and non-stationarity over spatial frequencies [61][62]. Variance stabilization using multiresolution transforms has been proposed to normalize noise over different scales [63].

Overall, rigorous characterization and modeling of the multiple noise sources and their interactions is key to developing optimized TEM reconstruction and restoration techniques. Both model-based and learning-based methods benefit from accurate noise models matched to real TEM imaging.

3.0.4 3D Convolutional Neural Networks

3D Convolutional Neural Networks (CNNs) are a powerful tool that leverage the unique characteristics of 3D context and convolution operations to perform a wide array of tasks such as segmentation, classification, and reconstruction of volumetric data. Among the various architectures available, the 3D U-Net has demonstrated exceptional performance, particularly in the field of medical image analysis, by utilizing encoder-decoder convolutions [64].

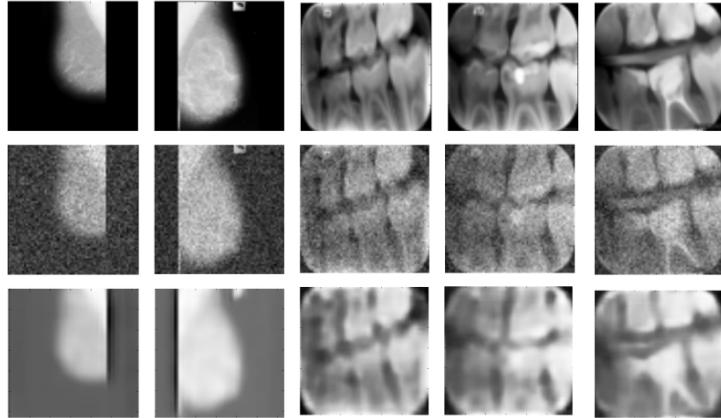


Figure 3.3: First row - Original real images, Second row - Noisier versions, Third row - Images after denoising with CNN DAE [64].

One of the key benefits of 3D CNNs is their ability to act as data-driven filters that can denoise Transmission Electron Microscopy (TEM) volumes, thus enhancing the interpretability of these volumes. For instance, 3D convolutional autoencoders that have been specifically trained to reconstruct TEM data can effectively serve as noise suppression filters [65]. The process of applying 3D CNN denoising before proceeding with coordinate-based Multilayer Perceptron (MLP) modeling may significantly help condition the data, making it more suitable for further analysis.

Combining Volumetric CNNs and Coordinate MLPs for Improved Performance In recent years, hybrid methods that merge the functionalities of volumetric CNNs and coordinate MLPs have been developed, demonstrating potential for improved reconstruction outcomes by leveraging their complementary strengths [46] [44]. In such a hybrid approach, volumetric CNN encoders first aggregate global context from the 3D input data. Subsequently, coordinate-based MLP decoders model local relationships at each individual location.

This unique combination enables the joint learning of multi-scale representations, in which the CNN provides top-down semantic guidance, while the MLP preserves the bottom-up spatial details. In the context of TEM data, this dual approach could effectively capture both anatomical priors and the fine structural variations that are typically present. The global-local modeling provided by this combination may enable accurate reconstruction from sparse, noisy tilt series projections, thereby potentially revolutionizing the way we handle and interpret such data.

3.0.5 Denoising

Reducing noise in transmission electron microscopy (TEM) images is critical for enabling accurate reconstruction and analysis. However, the low electron doses used in TEM result in extremely low signal-to-noise ratios. Conventional linear filters like Gaussian smoothing remove noise at the expense of blurred structural details. More advanced model-based methods are not robust to non-Gaussian noise encountered in TEM.

A variety of denoising methods have been developed for TEM images, including median filtering, Wiener filtering, wavelet transform-based denoising, and deep learning-based denoising [66]. Median filtering is a simple and efficient method, but it can blur image edges. Wiener filtering can preserve image edges, but it can be computationally expensive. Wavelet transform-based denoising is a good compromise between efficiency and image quality [66].

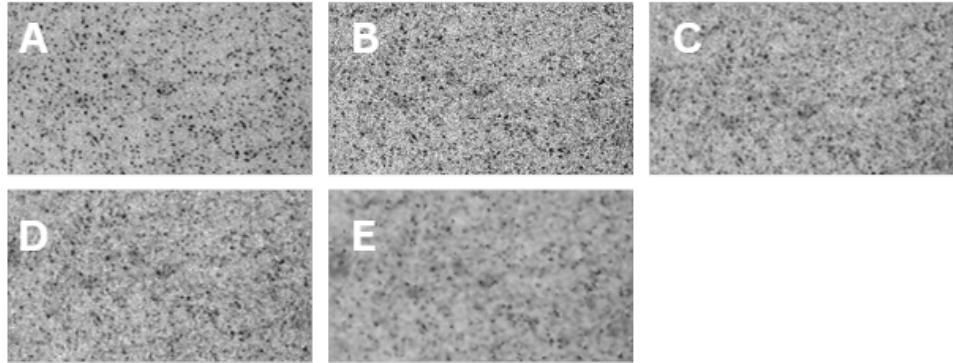


Figure 3.4: TEM Imaging and Denoising of Cadmium Sulfide Nanoparticles: **A.** Original Image, **B.** Image with Gaussian Noise, **C.** Denoised with Average Filter, **D.** Denoised with Median Filter, **E.** Denoised with Weiner Filter [66].

For denoising, sparse coding techniques take advantage of priors in the image such as non-local self-similarity and local sparsity. Sparsifying transforms are used by methods such as K-SVD [67] and BM3D to group comparable patches and filter noise. Computational expenses do not scale well with the magnitude of medical images, despite being effective. There are difficulties in choosing regularization parameters optimally.

Deep learning approaches have recently shown great promise for image denoising by learning data-driven filters [44]. Convolutional networks trained as discriminators between clean and noisy image patches can implicitly model complex image priors. Recurrent inference further boosts quality[44]. Autoencoder architectures directly optimize reconstruction loss [44]. Multi-image network training leverages complementary information across tilt series.

Applying and tailoring deep denoisers to 3D TEM data could significantly enhance reconstruction quality from noisy tilt projections. The high capacity of deep networks may better capture noise characteristics compared to hand-crafted models. Overall, learned denoising provides new opportunities to overcome resolution limits imposed by noise in TEM imaging.

CHAPTER 4

Datasets

This chapter delves into the datasets employed for experimental validation in this thesis, encompassing real transmission electron microscopy (TEM) and scanning TEM (STEM) images, alongside synthetically generated volumes. It provides comprehensive details on the source, composition, imaging parameters. The utilization of both real and synthetic data is crucial for the effective training, evaluation, and analysis of the proposed methods, ensuring a robust and comprehensive understanding of their performance across diverse data types.

4.0.1 TEM Datasets

Source for all TEM datasets: CAU Technical Faculty (Synthesis and Real Structure Group)

- **Dataset 1**

This TEM dataset showcases a single nanoparticle image, featuring a range of images with file sizes varying from 20KB to 280KB. Captured at a 20nm scale, these images present a stark contrast with their completely dark background. This dataset provides a diverse array of image qualities, offering a comprehensive view of the nanoparticle across different imaging conditions and resolutions.

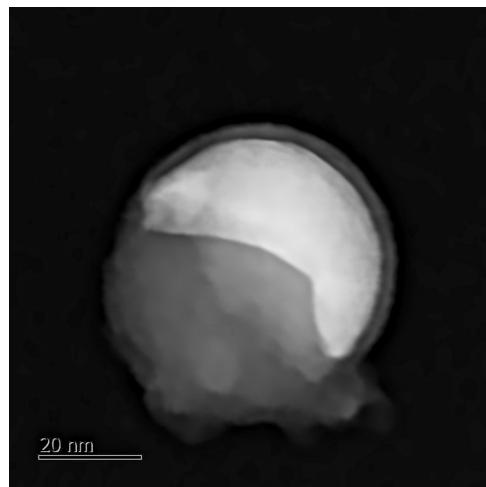


Figure 4.1: Single image from TEM Dataset 1

Table 4.1: Table Summary: Characteristics of TEM Dataset 1

Total Image	10
Dimensions	639 X 639 to 1496 X 1496
Format	BMP
Image Size	20KB to 280KB

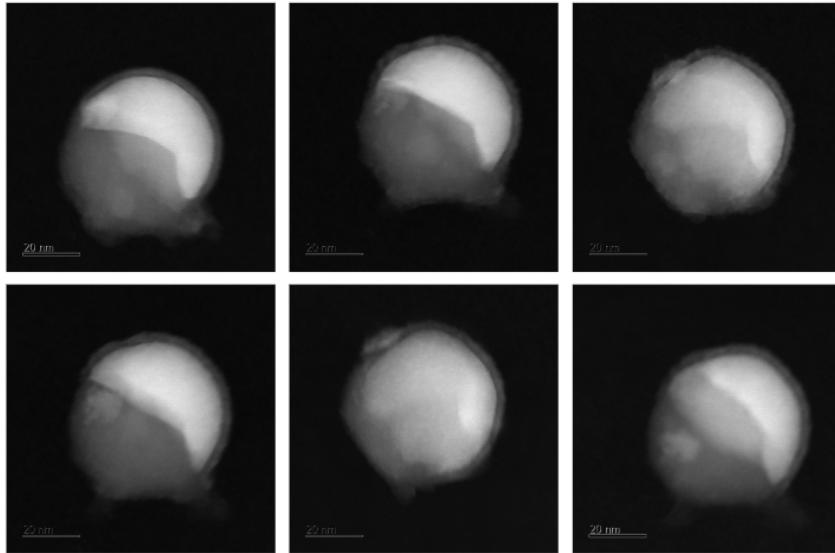


Figure 4.2: TEM Dataset 1

To prepare the dataset for NeRF, which requires JPG format, a script converts all BMP images in the folder to JPG. Given the dataset's ten images of varying dimensions, preprocessing is vital to standardizing their shape and aligning all images to the sizes of the first image. This uniformity is crucial for reducing noise and achieving consistent, reliable experiment results.

Listing 1 Resize all images to match the dimensions of the first image

```

1   import cv2
2   import numpy as np
3
4   for idx, file_name in enumerate(grayscale_files):
5       grayscale_img = cv2.imread(os.path.join(folder_path,
6           file_name), cv2.IMREAD_GRAYSCALE)
7
8       # If it's the first image, store its dimensions
9       if idx == 0:
10           first_image_height, first_image_width, _ =
11               enhanced_img.shape
12
13       enhanced_img = cv2.resize(enhanced_img,
14           (first_image_width, first_image_height))

```

- **Dataset 2**

This TEM dataset features numerous nanoparticles and is more zoomed out at a scale of 200 nm, allowing multiple nanoparticles to be visible in each image. The particles appear white against a completely dark background, making it challenging to discern details with the naked eye. This level of zoom offers a broader view of the sample, but with less detail compared to higher magnification images.

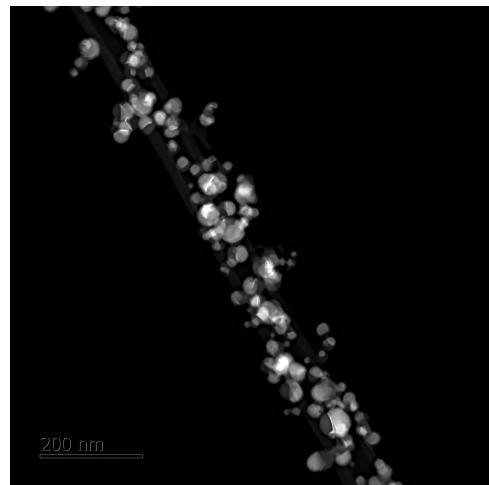


Figure 4.3: Single image from TEM Dataset 2

Table 4.2: Table Summary: Characteristics of TEM Dataset 2

Total Image	23
Dimensions	2048 X 2048
Format	BMP
Image Size	123KB

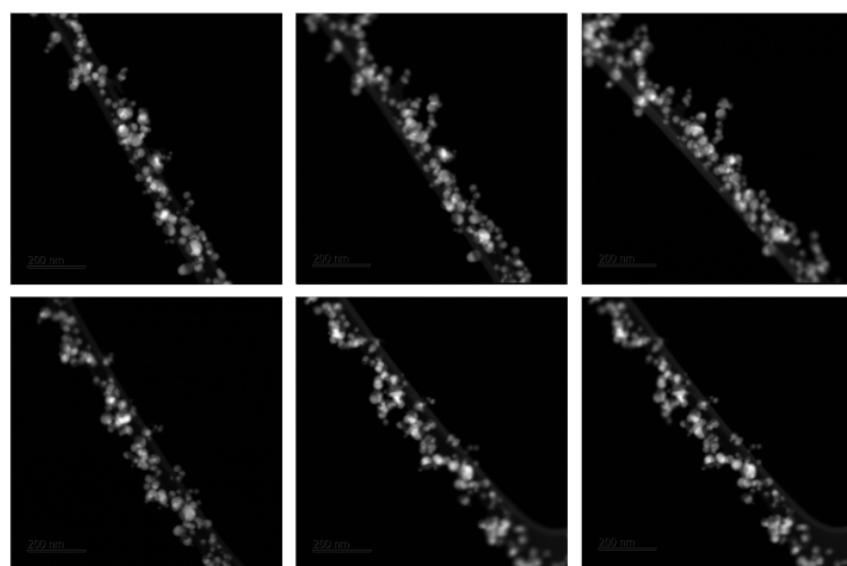


Figure 4.4: TEM Dataset 2

- **Dataset 3**

This dataset features multiple nanoparticles, but lacks specific scale information. The images, clearer than those in other datasets, distinctly show numerous nanoparticles with a V-shaped edge. Despite this clarity, the image quality is moderate with a file size of around 30KB. Notably, the images exhibit considerable noise, affecting the overall visibility and detail of the nanoparticles. This aspect makes it more challenging to analyze and interpret the fine details within the images.

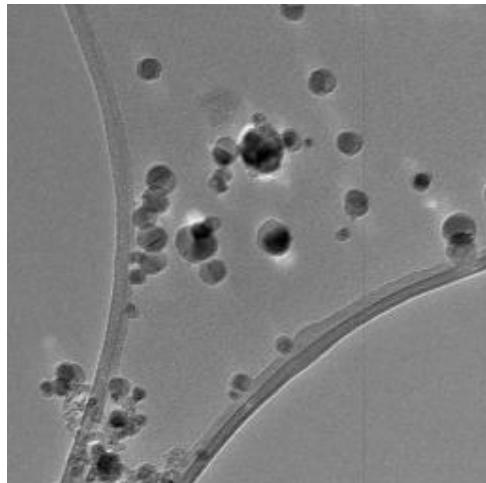


Figure 4.5: Single image from TEM Dataset 3

Table 4.3: Table Summary: Characteristics of TEM Dataset 3

Total Image	48
Dimensions	256 X 256
Format	BMP
Image Size	350KB

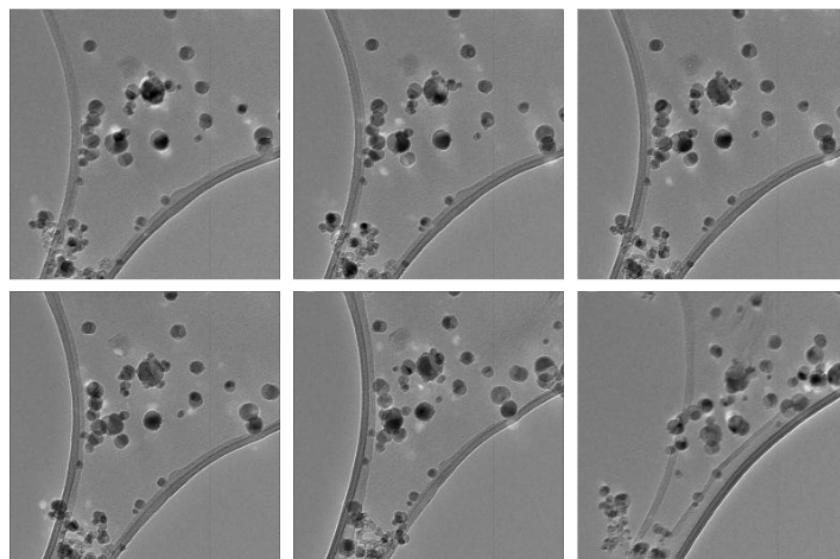


Figure 4.6: TEM Dataset 3

- **Dataset 4**

This dataset, similar in clarity to 4.6, features numerous nanoparticles. These particles are notably clustered at the top, middle, and bottom sections, forming concentrated groups. The background is gray, with lighter areas around the edges of the images, providing a contrast to the clustered nanoparticles. The unique arrangement and background contrast offer a distinctive view, differentiating it from other datasets and emphasizing the clustered nature of the nanoparticles.

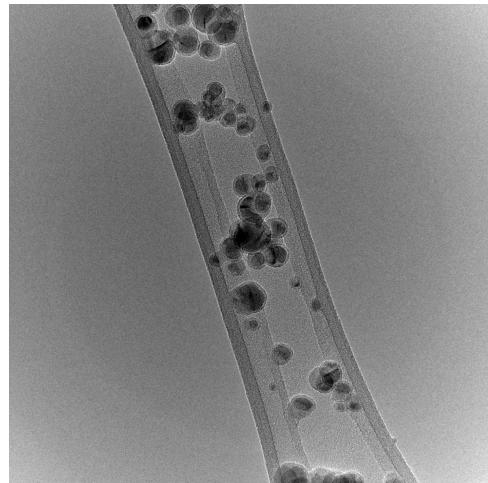


Figure 4.7: Single image from TEM Dataset 4

Table 4.4: Table Summary: Characteristics of TEM Dataset 4

Total Image	20
Dimensions	1421 X 1421
Format	JPG
Image Size	350KB

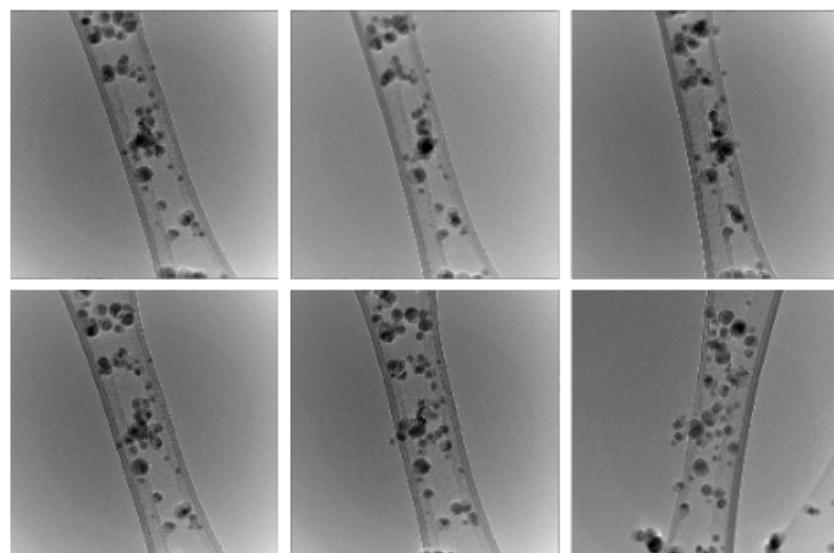


Figure 4.8: TEM Dataset 4

4.0.2 STEM Datasets

The dataset contains a large, clear particle and a smaller, blurry one, linked by a line. Despite its large 1MB size, the image is unclear and lacks a magnification scale.

Source for all STEM datasets: CAU Technical Faculty (Synthesis and Real Structure Group)

- **Dataset 1**

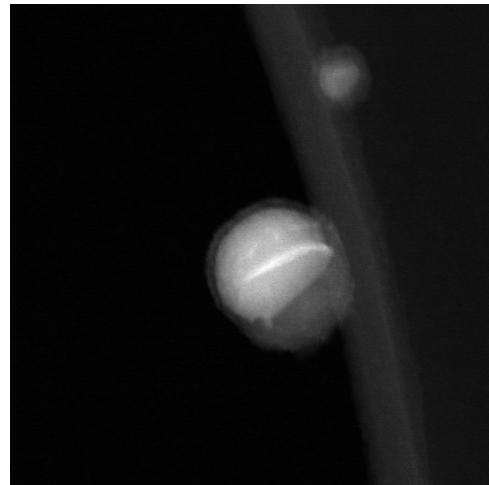


Figure 4.9: Single image from STEM Dataset 1

Table 4.5: Table Summary: Characteristics of STEM Dataset 1

Total Image	17
Dimensions	1024 X 1024
Format	JPG
Image Size	1MB

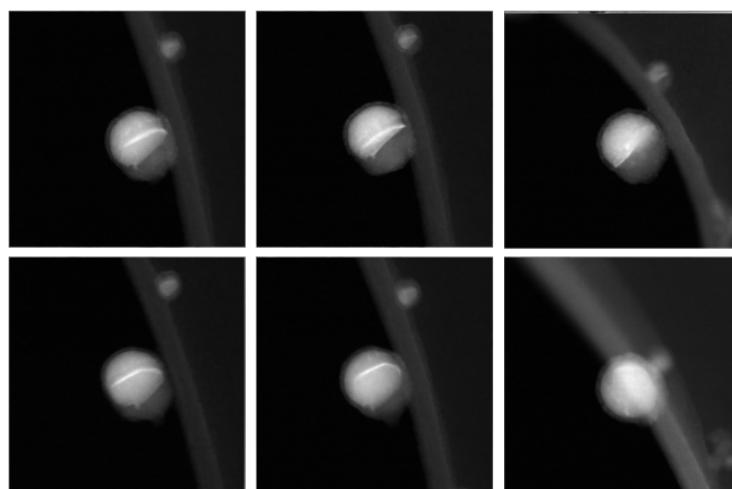


Figure 4.10: STEM Dataset 1

- **Dataset 2**

In this dataset, visual inspection reveals a relatively larger nanoparticle centrally positioned, with two smaller, blurry particles in the background. The image quality is notably low, with a file size of only 80KB, making it challenging to discern detailed information about the nanoparticles. The poor clarity and limited size of the image significantly hinder the ability to gain any meaningful insights into the characteristics of these nanoparticles.

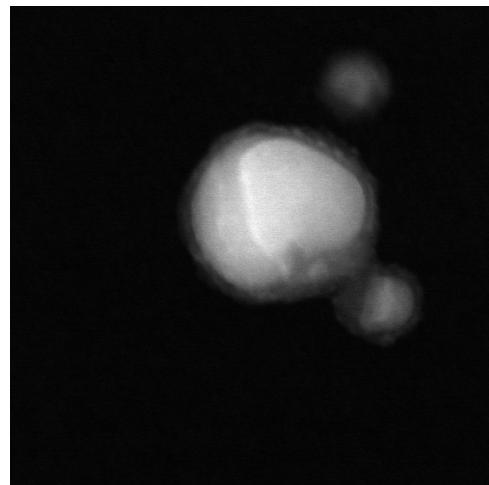


Figure 4.11: Single image from STEM Dataset 2

Table 4.6: Table Summary: Characteristics of STEM Dataset 2

Total Image	16
Dimensions	1024 X 1024
Format	JPG
Image Size	80KB

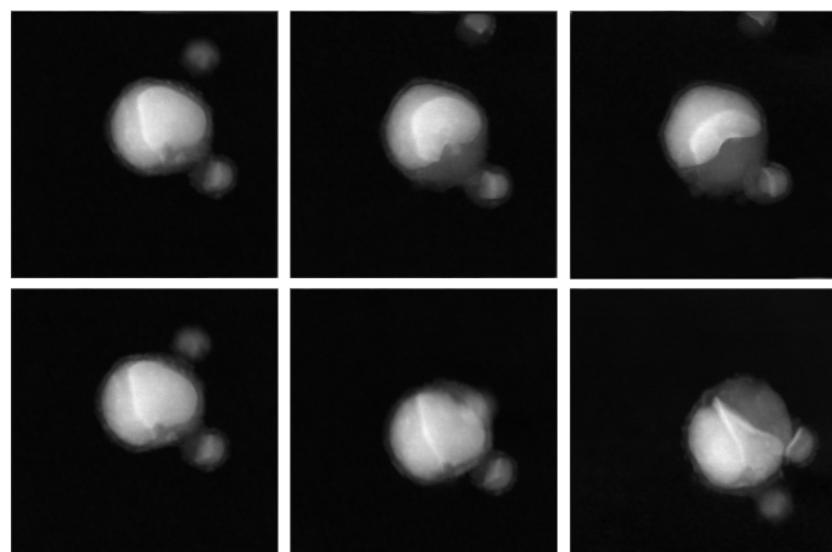


Figure 4.12: STEM Dataset 2

4.0.3 Synthetic Dataset

- **Dataset 1**

This synthetic dataset, not an actual TEM or STEM dataset, was created using Blender to mimic the appearance of TEM/STEM data from [68]. Despite its clarity and lower noise levels, each image is only about 80KB. The collection comprises approximately 120 images, offering a rich perspective from various angles. The dataset features a grey background with light nanoparticle shapes at the center, providing a clear and informative visual representation despite its synthetic origin.

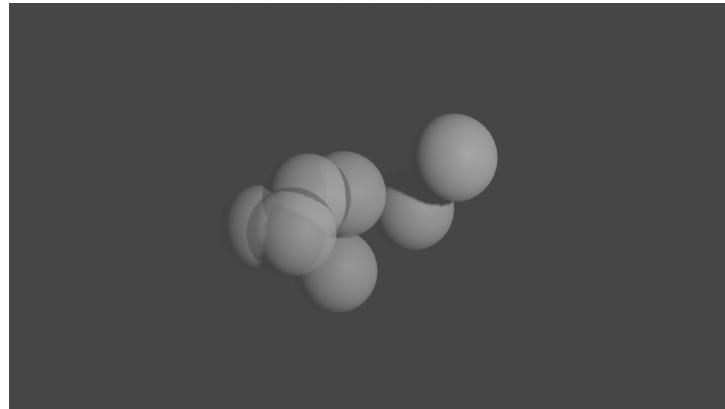


Figure 4.13: Single image from Synthetic Dataset [68]

Table 4.7: Table Summary: Characteristics of Synthetic Dataset

Source	Blender Generated Data
Total Image	120
Dimensions	800 X 800
Format	JPG
Image Size	80KB

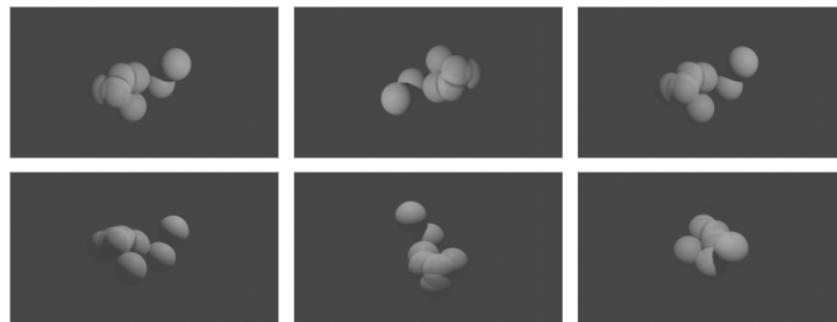


Figure 4.14: Synthetic Dataset [68]

CHAPTER 5

Methodology

5.0.1 Original Neural Radiance Fields (NeRF)

Neural Radiance Fields (NeRF) [7] as a foundational notion established a new paradigm for scene representation. The main concept is to describe a scene as a continuous, high-dimensional radiance field, F_Θ , where Θ represents the multilayer perceptron's (MLP) parameters. NeRF is based on computer graphics theory and uses deep learning methods to produce synthesized views with previously unheard-of levels of photorealism. In Figure 5.1, an overview of NeRF is presented. On the left side of the image, the black dots indicate areas where the network has not yet learned any color information. In contrast, the second image demonstrates the colored output corresponding to a single pixel.

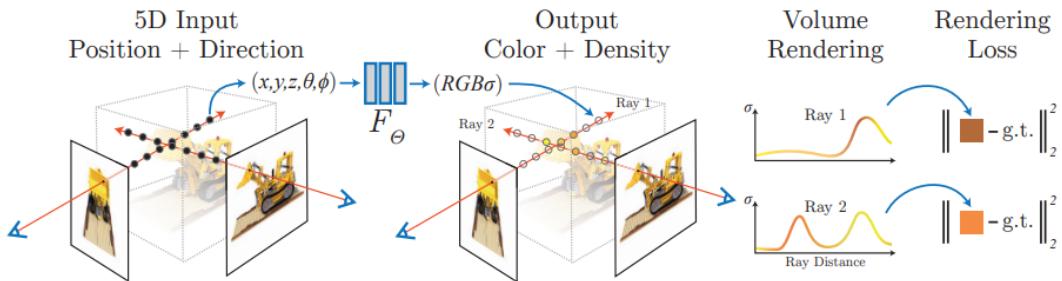


Figure 5.1: NeRF Overview: Image creation involves sampling 5D coordinates along camera rays, using a multilayer perceptron for color and density, and employing volume rendering for image compilation [7].

Given any 3D location in space $\mathbf{x} = (x, y, z)$ and a specific 2D viewing direction $\mathbf{d} = (\theta, \phi)$, the MLP is designed to output both an RGB color vector $\mathbf{c} = (r, g, b)$ and a scalar volume density σ [7]. This relationship is mathematically expressed as:

$$F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$$

Where:

- F_Θ is the neural radiance field parameterized by network parameters θ .
- $\mathbf{x} = (x, y, z)$ represents the 3D coordinates.
- $\mathbf{d} = (\theta, \phi)$ denotes the 2D viewing direction.
- $\mathbf{c} = (r, g, b)$ is the emitted color.

- σ is the volume density.

Volumetric ray marching is used in NeRF[7] to accomplish the rendering process. Using this method, rays are traced into the scene from the camera origin, and sampling points are taken along each ray. The MLP is contacted at each of these sample places in order to get density and color values. An alpha compositing technique is used to accumulate color along a ray. This can be stated as:

$$C(\mathbf{r}) = \sum_{i=1}^N T_i(1 - \exp(-\sigma_i \delta_i)) \mathbf{c}_i,$$

where $T_i = \exp(-\sum_{j=1}^{i-1} \sigma_j \delta_j)$ represents[7] the transmittance, accounting for the absorption and scattering of light as it travels through the medium. Here, δ_i denotes the distance between adjacent sampled points along the ray.

Reconstruction loss vs rendered loss is used to optimize the whole NeRF model by reducing the discrepancy between rendered and captured views. Through this optimization, the MLP is able to pick up on both geometric and photometric subtleties, resulting in a very accurate and detailed representation of the scene.

5.0.2 Extending NeRF to Handle Transmission Electron Microscopy (TEM) Data

The standard NeRF pipeline relies on accurate camera pose estimation, typically obtained through structure-from-motion methods like COLMAP. However, TEM images pose challenges due to their noise levels and unique characteristics, making traditional pose estimation techniques like COLMAP ineffective.

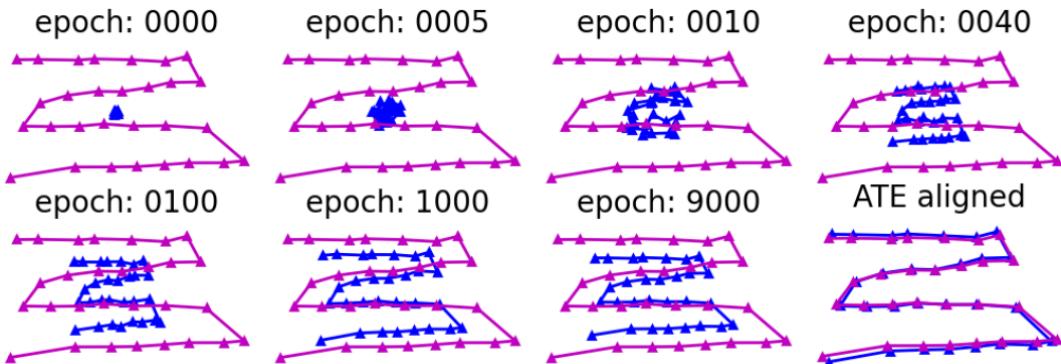


Figure 5.2: Optimization of Camera Pose in NeRF-MM. The purple path represents the camera trajectory derived from COLMAP, while the blue path illustrates the camera trajectory as determined by NeRF-MM. [69]

Introduced by Wang et al. in 2021 [69], NeRF-MM (Neural Radiance Fields Without Known Camera Parameters) represents a substantial advancement of the original NeRF model. This enhanced version is tailored to address the intricacies involved in generating camera poses during training, as depicted in 5.2. Such a feature is especially relevant in the realm of TEM (Transmission Electron Microscopy) imaging, where accurately determining camera positions is crucial.

The primary innovation in NeRF-MM is its capability to process and interpret multimodal data. This is achieved by incorporating additional scene descriptors or feature vectors for each 3D point in the space. This approach is crucial in TEM imaging, where different materials can exhibit varied electron interactions, significantly impacting the final image quality.

NeRF-MM introduces a conditioning model, C_Φ , which is central to its ability to handle multimodal data. The conditioning model is defined as:

$$C_\Phi : \mathbf{x} \rightarrow \mathbf{f} \quad (5.1)$$

Where \mathbf{x} is the 3D coordinate in the scene, and \mathbf{f} represents the generated feature vector. This vector encodes additional information about the scene, such as material properties or lighting conditions, which are not captured by the spatial coordinates alone.

The radiance field function in NeRF-MM is modified to incorporate these feature vectors:

$$F_\Theta : (\mathbf{x}, \mathbf{d}, \mathbf{f}) \rightarrow (\mathbf{c}, \sigma) \quad (5.2)$$

Here, \mathbf{d} represents the viewing direction. The function outputs the color \mathbf{c} and density σ at the point \mathbf{x} , considering the feature vector \mathbf{f} and viewing direction \mathbf{d} .

A critical aspect of NeRF-MM is its advanced approach to volumetric integration. The model employs a sophisticated sampling strategy along the rays passing through the scene. This is mathematically represented as:

$$\text{Color}(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt \quad (5.3)$$

In this equation, $T(t)$ denotes the accumulated transmittance along a ray \mathbf{r} from the near boundary t_n to the far boundary t_f [69]. This intricate sampling method ensures a detailed and accurate reconstruction of the scene, which is especially beneficial in TEM imaging where the subjects often exhibit complex geometries and material properties.

To optimize NeRF-MM for the specific challenges of TEM imaging, such as high noise levels and fine-scale detail capture, we have introduced targeted modifications to the model. These include:

- Adjusting the depth and width of the MLP to increase the model's capacity for handling complex structures typical in TEM data.
- Fine-tuning the conditioning model C_Φ to better represent the unique material properties encountered in TEM imaging.
- Implementing a noise-aware training strategy to improve the model's robustness to the high noise levels commonly found in TEM datasets.

A notable feature of NeRF-MM is its independence from external camera parameter estimation tools, such as COLMAP. This is particularly advantageous in TEM imaging, where precise camera positioning and calibration can be challenging. NeRF-MM's advanced feature vector representation and sampling methods allow it to inherently adapt

to scenes with minimal or no camera parameter information, facilitating its application in TEM imaging.

NeRF-MM, with its advanced architecture and multimodal data handling capabilities, represents a significant evolution in the field of 3D reconstruction, particularly for TEM imaging. Its unique approach to incorporating additional scene descriptors, sophisticated volumetric integration, and optimizations tailored for TEM data, make it an ideal choice for addressing the challenges inherent in TEM imaging scenarios. By effectively capturing and processing the complex interplay of material properties and electron beam interactions, NeRF-MM enables the creation of highly detailed and accurate 3D models from TEM datasets.

Adopting NeRF-Dark Techniques: NeRF-Dark, as proposed by Mildenhall et al. in 2021 [70], marked a significant advancement in handling low-light conditions, a feature beneficial for TEM imaging where illumination is often limited. Building on this, RawNeRF, a variation trained on linear raw images, demonstrates remarkable capabilities in reconstructing scenes from extremely noisy images captured in near-dark conditions. This makes it particularly suitable for the challenges faced in TEM imaging.

NeRF in the Dark referred as RawNeRF outperforms conventional single and multi-image deep raw denoisers by effectively combining information from all input images for reconstruction [70], even in the absence of explicitly learned image priors or clean training data. This approach allows RawNeRF to serve as a powerful multi-image denoiser for wide-baseline static scenes, efficiently aggregating observations from widely spaced input images. Moreover, the linear High Dynamic Range (HDR) scene representation of RawNeRF supports novel view synthesis tasks, including adjustments in focus, exposure, and tonemapping post-rendering. This flexibility is a significant leap forward in scene reconstruction and view synthesis [70].

RawNeRF demonstrates exceptional noise robustness, handling high levels of image noise and surpassing traditional NeRF and other denoising methods in producing accurate scene representations. This capability makes it particularly effective for TEM imaging, which often contends with significant noise challenges. Merging principles from NeRF, low-level image processing, and HDR imaging, RawNeRF contributes substantially to advancements in scene reconstruction and view synthesis. Utilizing this interdisciplinary approach, we have integrated RawNeRF’s denoising methodology into our updated NerfMM variant. This integration not only enables the 3D reconstruction of TEM images but also enhances denoising capabilities, surpassing the performance of NerfMM alone. This synergy ensures high-quality image reconstruction from TEM data, typically challenged by darkness and noise.

Utilizing NAN-NeRF Innovations: NAN-NeRF, a breakthrough by Pearl et al. in 2022 [71], marks a significant advancement in NeRF’s application to nanoscale imaging, a domain closely aligned with TEM that often grapples with capturing fine details at a small scale. This innovation is particularly relevant for our methodology, addressing TEM’s challenges in resolving intricate details at the nanoscale. NAN-NeRF excels in reconstructing accurate 3D models from highly limited and noisy data, a capability crucial for TEM, where data quality and quantity are often limited due to sample sensitivity and electron dose constraints.

In our approach, we've adopted key techniques from NAN-NeRF, such as replacing NerfMM's ReLU function with NAN-NeRF's Leaky ReLU and Sigmoid functions. Additionally, we've implemented separate Adam optimizers for NerfMM and other parameters from NAN-NeRF. These modifications are tailored to enhance our methodology, aiming to achieve more precise 3D reconstructions and better handle the intricate challenges posed by TEM imaging at the nanoscale. This integration of NAN-NeRF techniques into our framework not only improves the denoising capabilities but also ensures a more nuanced and detailed approach to reconstructing high-quality images from TEM data.

Hyperparameter Adjustments for TEM Data: In response to the lower signal-to-noise ratio typical of TEM images, our methodology incorporates essential hyperparameter adjustments in the NeRF model to enhance sensitivity to subtle data variations, crucial for distinguishing actual features from noise. Key modifications include a refined learning rate for more precise model updates, and adjustments to the depth and width of the NeRF model's MLP to balance modeling complex TEM structures and avoiding noise overfitting. Advanced denoising techniques and exposure correction algorithms are also employed to improve feature visibility in low-light TEM images.

Our experimentation extended to a comprehensive tuning of all possible hyperparameters to optimize the model for TEM imaging. We incorporated activation functions from NAN-NeRF and fine-tuned network layers to achieve the best results. To validate the model's accuracy, we reserved one sample image with a known camera position, separate from the training set. We also adopted different optimization algorithms from NAN-NeRF. While the best results were obtained at 10,000 epochs, this was time-intensive, taking approximately 10 to 14 hours. Therefore, for baseline experiments, we opted for 200 to 1000 epochs, which significantly reduced processing time to a few minutes to an hour. After establishing successful baseline results, we conducted all experiments at 10,000 epochs to maximize accuracy and detail in the reconstructed images.

In summary, by integrating these advanced NeRF variations and meticulously adjusting hyperparameters, our methodology establishes a robust framework for handling the noise and imaging challenges characteristic of TEM data. This approach not only enhances the fidelity and quality of TEM imaging but also broadens the scope of its applications in material science and nanotechnology research.

5.0.3 Post-Reconstruction Processing with ESRGAN

In the post-reconstruction phase of our TEM imaging analysis, the 3D model's frames undergo further processing for denoising. Among various methods explored, Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN)[37] proved to be exceptionally effective. The ESRGAN model operates by transforming a low-resolution, noisy input image into a high-resolution, denoised output.

The ESRGAN model utilizes a generative adversarial network (GAN) framework comprising a generator and a discriminator, trained simultaneously. The generator aims to produce images that closely resemble real, high-resolution images, while the discriminator evaluates their authenticity. Through this adversarial training, ESRGAN learns to enhance details and reduce noise effectively. It incorporates advanced loss functions, including perceptual and adversarial loss, to improve image quality and detail fidelity.

This capability makes ESRGAN particularly adept at handling the complex textures and subtle details characteristic of TEM imagery, thereby significantly enhancing the quality and utility of the reconstructed models.

Implementation of ESRGAN in Our Workflow: In our methodology, the application of ESRGAN is a principal post-processing step after extracting frames from the 3D model. However, we designed this step to be flexible, allowing for the possibility of incorporating other advanced methods in the future. Our decision to use ESRGAN was based on comprehensive trials involving various GANs and denoising techniques. Among these, ESRGAN stood out for its exceptional performance in enhancing a specific dataset we evaluated. Its proficiency in preserving and recovering intricate details, coupled with its effective noise reduction, was instrumental in our choice to integrate it into our workflow.

Performance Metrics and Results: The proficiency of ESRGAN is quantifiably demonstrated through its remarkable performance in peak signal-to-noise ratio (PSNR) and Structural Similarity Index Measure (SSIM) metrics. In our evaluations, ESRGAN achieved a PSNR of 38.70, outperforming other methods, and recorded an SSIM value of 0.95. These metrics are especially significant in TEM imaging, as they are directly indicative of the clarity and accuracy of the reconstructed images.

5.0.4 Integrating Traditional Denoising Techniques

In addition to advanced techniques like ESRGAN, traditional denoising methods also play a significant role in post-processing the frames of the 3D model. These methods are essential for handling various types of noise and artifacts present in the images.

Traditional denoising techniques employed include wavelet denoising, Gaussian blur, median filtering, bilateral filtering, and non-local means denoising. Each method offers unique advantages in reducing noise while preserving the essential features of the image.

- **Wavelet Denoising:** This method employs wavelet transforms to decompose an image into various frequency components. By selectively thresholding these components, it effectively separates noise from significant image features. Wavelet denoising is particularly effective for TEM images with non-uniform noise distribution or images where noise characteristics vary across different regions.
- **Gaussian Blur:** Gaussian blur involves convolving the image with a Gaussian kernel. This process smooths the overall image, reducing high-frequency noise elements like grainy textures. It's especially useful in TEM images where the primary concern is overall image graininess rather than preserving high-frequency details.
- **Median Filtering:** In median filtering, each pixel is replaced with the median value of its surrounding pixels. This technique is particularly effective in removing 'salt-and-pepper' noise, common in digital image sensors and often encountered in TEM imaging.
- **Bilateral Filtering:** Bilateral filtering is a more sophisticated approach that combines smoothing with edge preservation. By considering both spatial and intensity differences, it smooths the image while maintaining edge details, making it suitable for TEM images where edge preservation is crucial.

- **Non-Local Means Denoising:** This technique works by averaging similar patches across the image, thus preserving textures and details. It's particularly effective in TEM imaging for maintaining structural integrity in the presence of noise, as it does not rely on local neighborhood intensity values alone.

These methods were systematically applied to the frames extracted from the 3D model. The process involved reading each frame, applying the denoising techniques, and saving the enhanced images into respective subfolders for each method. This approach ensured that the most suitable denoising technique could be selected based on the specific characteristics of the noise present in each frame.

5.0.5 Adapted Framework for TEM Data Analysis

In summary, the methodology presented in this chapter represents a pioneering convergence of advanced computational techniques and traditional image processing methods, specifically tailored for the unique challenges of Transmission Electron Microscopy (TEM) data analysis. By integrating state-of-the-art Neural Radiance Fields (NeRF) models with a combination of modern and traditional denoising strategies, we have established a comprehensive framework that significantly enhances the quality and usability of TEM data.

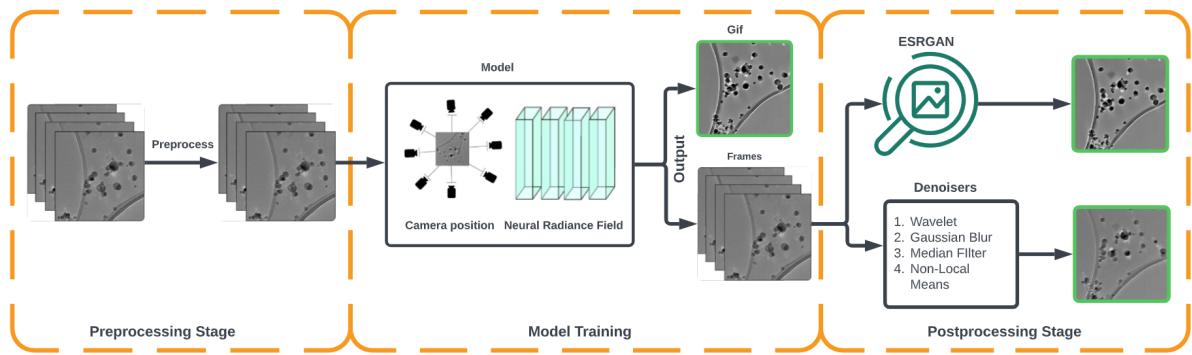


Figure 5.3: Comprehensive overview of our framework: Integrating NeRF and advanced Denoising Techniques, ESRGAN for Enhanced TEM Image Analysis

As depicted in Figure 5.3, our framework illustrates the power of integrating diverse computational methodologies into a cohesive process. The **green** borders in the architecture diagram highlight three key outcomes:

1. **3D Reconstruction from 2D Input Images:** The first green-bordered image represents the transformation of traditional 2D TEM data into comprehensive 3D models using NeRF, providing deeper insights into material structures.
2. **Output from ESRGAN:** The second green-bordered image showcases the enhanced output from the application of ESRGAN, focusing on resolution enhancement and noise reduction in TEM images.
3. **Optimal PSNR from Traditional Denoising Algorithms:** The final green-bordered image represents the best result achieved using traditional denoising algorithms, emphasizing the importance of these techniques in preserving image fidelity.

CHAPTER 6

Implementation

This section provides a comprehensive overview of the key steps taken to implement our integrated framework for denoising and analysis of TEM images, as depicted in Figure 5.3. We build upon recent advances in Neural Radiance Fields, specifically the NeRF-MM architecture along with techniques from NAN-NeRF and NeRF-Dark, to address the noise and imaging challenges associated with TEM data. It outlines the development of these methods, specifically aimed at addressing denoise TEM Images. The implementation process is structured to illustrate each step, from initial environment setup to final execution. Emphasis is placed on key parts of the code and their contributions to the overall methodology, ensuring a comprehensive understanding of the applied techniques and how they integrate to achieve the objectives of this thesis.

6.0.1 Setup Overview

The NeRF-MM implementation was conducted within a Python environment, leveraging crucial libraries such as Torch and NumPy. Initial steps included setting the `CUDA_LAUNCH_BLOCKING=1` environment variable for synchronized CUDA operations and cloning the NeRF-MM repository from GitHub. To ensure the reproducibility of results, seed values for random, NumPy, and Torch were carefully set, and the deterministic behavior of PyTorch’s backend (CUDNN) was enabled. The implementation extensively used core modules and utility functions from the NeRF-MM library, notably `encode_position` and `volume_rendering`, which play a vital role in positional encoding and volume operations within the neural radiance field framework.

6.0.2 Key Implementation Steps

The primary steps involved in the NeRF-MM implementation were as follows:

1. **Data Preprocessing:** The data preparation phase involved importing and processing the TEM images, initially in grayscale BMP format, to fit our model’s requirements. To align with our needs, these images were converted to JPG or PNG format and transformed into RGB color space. The dataset also had images of varying dimensions, posing a challenge for model training. To resolve this, we used a Python script for preprocessing, standardizing the shape and size of all images. This step was essential to ensure data consistency and efficient model training.
2. **Model Architecture:** We enhanced the TinyNerf class from NerfMM[69] by integrating features from NAN-Nerf [71] and Nerf-Dark [70], optimizing it for learning

from noisy TEM data. Our updates included changes in layers, activation functions, and hyperparameters to better handle sparse TEM image sampling and noise.

Algorithm 2 Enhanced TinyNerf Model Architecture

```

1: Class TinyNerf:
2: Define Sequential Layers like Nerf-Dark:
3: for each layer in Sequential Layers do
4:   Add Linear Layer with Leaky ReLU Activation
5: end for
6: Define Density and RGB Layers
7: Define RGB Output Layer
8: Function Forward Pass:
9: for each input do
10:  Compute RGB Value
11:  Concatenate RGB and Density
12:  Return Concatenated Output
13: end for
```

The revised TinyNerf effectively processes positional and directional encodings, improving scene density and RGB output computation. These enhancements enable our model to capture fine details and lighting variations in limited view radiance fields.

This synergizing of neural network advancements in our TinyNerf adaptation improves representation learning and novel view synthesis from challenging TEM datasets.

3. **Training and Optimization:** The training and optimization of the NeRF-MM model were critical stages of the implementation. This process involved setting up and training a TinyNerf neural network model, fine-tuning camera pose parameters, and optimizing the focal length. A critical optimization technique implemented was an early stopping mechanism. The training process monitored the updates in camera parameters and halted the training prematurely if these parameters ceased to significantly change after a certain number of epochs. This approach ensured efficiency by avoiding unnecessary computations when the model had converged.

Algorithm 3 Training Process and Optimization

```

1: Initialize training parameters
2: for each epoch in N_EPOCH do
3:   Compute L2 loss using train_one_epoch function
4:   Calculate train_psnr from L2 loss
5:   Log PSNR value for current epoch
6:   Check for convergence based on camera parameter updates
7:   if convergence criteria met then
8:     Break the loop
9:   end if
10:  Update learning rate and training steps
11: end for
```

In addition to image enhancement using a pre-trained model, our training loop incorporated crucial elements for optimal performance evaluation. This included calculating the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) as metrics to quantify image quality improvements. To enhance the training process, we utilized a staircase exponential decay scheduler to dynamically adjust the learning rate. This strategic approach contributed to the model's effective convergence, ensuring stable and optimal performance.

4. **Rendering and Visualization:** This phase focused on rendering 3D views using a spiral camera trajectory and visualizing camera poses 7.8. The NeRF model rendered novel viewpoints, dynamically representing the scene. To optimize computational efficiency, images were resized without compromising visual quality. Rendered images and depth maps were saved as GIFs for detailed scene analysis. Additionally, camera pose visualization through animated plots provided insights into the model's spatial understanding, dynamically illustrating camera movement over training epochs.
5. **Post-processing:** After generating the GIFs, key post-processing steps were applied to all frames for quality enhancement. This process includes:
 - (a) **Enhancing with ESRGAN:** Image enhancement was achieved using the ESRGAN model. This method focuses on upscaling and improving image resolution. The pre-trained ESRGAN model from TensorFlow Hub was utilized for this enhancement.

Algorithm 4 Enhance Image using Pre-trained Model: ESRGAN

- 1: Load a pre-trained model from TensorFlow Hub
 - 2: Convert the image at image_path to a TensorFlow tensor and preprocess it
 - 3: Normalize the RGB image by dividing by 255.0
 - 4: Expand the dimensions of the RGB image to match the model input shape
 - 5: Apply the pre-trained model to the preprocessed image
 - 6: Squeeze the enhanced image to remove the batch dimension
 - 7: Create directories in the save path if they don't exist
 - 8: Save the grayscale output to the specified save path
-

- (b) **Denoising:** This step involves applying various denoising techniques to reduce noise artifacts. Methods such as wavelet, Gaussian blur, median filter, bilateral filter, and non-local means were used. Each technique was applied to the frames and the denoised images were saved in respective subfolders.

These techniques significantly improve the visual quality of the images, rendering them more suitable for in-depth analysis.

6.0.3 Code Availability

The complete source code for this project, including all scripts and detailed documentation, is available on GitHub. It can be accessed at:

[https://github.com/Mithunjack/Thesis-NeRFs.](https://github.com/Mithunjack/Thesis-NeRFs)

CHAPTER 7

Experiments & Results

In this pivotal chapter, we embark on a journey through the experiments and results that form the core of this thesis. Our exploration begins with the initial steps of setting up the dataset using the NeRF original codebase, and it culminates in the realization of our ultimate thesis goal. Within these pages, we delve into the intricacies of each experiment, documenting both the successes and setbacks encountered along the way.

7.1 Preliminary Evaluation of NeRFStudio

Our initial foray into Neural Radiance Fields (NeRF) began with the exploration of NeRF-Studio, a crucial platform for understanding and utilizing NeRF techniques. This exploration included setting up NeRFStudio and employing their publicly available datasets. The outcomes of these initial trials closely matched the results documented in the original NeRF paper, confirming the tool's effectiveness and reliability. This preliminary phase was instrumental in establishing a solid foundation for our subsequent, more in-depth experiments involving custom datasets. These early trials with NeRFStudio were pivotal in shaping our approach and understanding of NeRF's capabilities and limitations.

7.1.1 Utilization of Custom Datasets in NeRFStudio

In our quest to deepen our understanding of Neural Radiance Fields, we initiated an experiment using a custom dataset. This dataset comprised a small collection of images, numbering around 5 to 6, all depicting the same object - a clock. These images were captured from various angles using an iPhone X, resulting in high-quality images with clear details. Each image was sizeable, ranging from 9MB to 11MB, and shared the same resolution of 3024 X 4032 pixels. This collection of images, distinct in their clarity and focus on the singular object, provided a valuable resource for our study.

The experiment with the dataset in 7.1 revealed limitations in the NeRF and NeRF-Studio models. It became evident that using a small dataset of only 5 to 6 images led to suboptimal reconstruction quality. The images, despite their clarity, resulted in a blurry reconstruction when processed through these models. This outcome highlighted the significant impact of dataset size on the effectiveness of NeRF-based reconstructions, emphasizing the need for larger datasets to achieve higher quality in 3D model reconstruction.



Figure 7.1: Sample Image from the Clock Dataset.

7.1.2 Scaling Dataset Size

In our subsequent experiment, we employed a larger custom dataset as detailed in 7.2. This dataset was comprised of over 50 high-resolution images of a single object, captured using an iPhone X, with each image size ranging between 10 to 15 MB. The high-quality nature of these images was evident, as even when magnified three times, the images remained clear, with no blurriness. This experiment emphasized the importance of dataset quality and size in achieving optimal reconstructions with NeRF models.



Figure 7.2: Dataset of 2D Images from Various Camera Positions.

The results of this experiment demonstrated a significant improvement in reconstruction quality compared to the smaller dataset.

The still frames from the generated video provided insight into the potential of NeRF when working with datasets that meet the minimum image count requirements. Through



Figure 7.3: Still Frames from Video Generated by NeRFStudio Using 2D Images.

these experiments, we gained valuable insights into the capabilities and limitations of Neural Radiance Fields, emphasizing the significance of dataset size. These initial steps set the stage for further exploration and optimization in subsequent experiments.

7.2 Exploration of TEM Image Experiments

In this experiment, we ventured into the realm of Transmission Electron Microscopy (TEM) images, a domain known for its unique challenges in image reconstruction. Armed with a better understanding of NeRFStudio, we set out to apply Neural Radiance Fields (NeRF) to our original TEM dataset.

Our initial attempt involved training NeRFStudio with Dataset 1, comprising TEM images of our target subject. However, our efforts encountered an unexpected obstacle. Within a matter of minutes, an error emerged, indicating that COLMAP was unable to create the "transforms.json" file, rendering it impossible to proceed with training.

```
=====
Loading database
=====

Loading cameras... 0 in 0.000s
Loading matches... 0 in 0.000s
Loading images... 0 in 0.000s (connected 0)
Building correspondence graph... in 0.000s (ignored 0)

Elapsed time: 0.000 [minutes]

WARNING: No images with matches found in the database.

ERROR: failed to create sparse model
```

Figure 7.4: COLMAP Error During Training.

The underlying challenge became apparent—TEM images, characterized by their inherent noise and low quality, posed a formidable hurdle. In the context of NeRF, precise camera position information is essential for successful 3D image generation from 2D inputs. Unfortunately, the noise and quality limitations of TEM images made it extremely challenging for NeRFStudio, utilizing the Nerfacto model, to reconstruct accurate camera positions, resulting in the encountered issues.

7.3 Exploring Alternative Nerf Models for TEM Images

Recognizing the limitations of NeRFStudio with TEM images, we embarked on a quest to discover alternative Nerf models that might better address the challenges posed by TEM data. Our journey into the realm of Neural Radiance Fields led us to experiment with various models, each with its unique characteristics and capabilities. In the following table, we summarize the Nerf models we explored and the specific limitations we encountered when working with TEM images:

Nerf Models	COLMAP Limitation with TEM Images
NeRFStudio(Nerfacto)	Dependency on COLMAP ✗
LeRF	Dependency on COLMAP ✗
Mip-NeRF	Dependency on COLMAP ✗
Instant-NGP	Dependency on COLMAP ✗
MultiNeRF	Dependency on COLMAP ✗
NanNeRF	Dependency on COLMAP ✗
NeRF in the Dark	Dependency on COLMAP ✗
IBRNeRF	Dependency on COLMAP ✗
RefNeRF	Dependency on COLMAP ✗
SCNeRF	Dependency on COLMAP ✗
gNeRF	Dependency on COLMAP ✗
NeRFMM	Not Dependent on COLMAP ✓

Table 7.1: Summary of Nerf Models Experimented with TEM Images.

This research, we delved into various Nerf models to address the specific challenges presented by TEM data, particularly focusing on achieving accurate 3D reconstructions. Most existing models were constrained by their reliance on COLMAP, a limitation that often hindered their effectiveness with TEM data. Our exploration led us to NeRFMM, a notable model that operates independently of COLMAP, offering a significant breakthrough for TEM image reconstruction. Encouraged by the success of NeRFMM, we extended our search to include the latest advancements in Nerf technology, seeking models with similar capabilities and compatibility with TEM data. Among these, LUNeRF emerged as a particularly promising model, demonstrating potential in preliminary assessments. However, the lack of public access to LUNeRF’s codebase has been a barrier, limiting our ability to fully explore and validate its applicability in the context of TEM data analysis. This ongoing exploration underlines the dynamic nature of our research and the continual search for innovative solutions in the field of 3D image reconstruction.

7.4 Exploration of COLMAP Alternatives for TEM Images

Recognizing that nearly 95% of our attempts to apply NERF models to TEM images resulted in failure, we conducted an in-depth investigation into the underlying issue. It became apparent that COLMAP played a pivotal role as a critical dependency in these failures. Our primary goal was to identify alternative software tools capable of processing TEM images and generating the camera’s exact positions or an equivalent format, which could subsequently be converted into a `transform.json` file. Such a file format held the potential to enable the execution of NERF models that were originally designed for COLMAP compatibility.

Our experiment involved rigorously testing four different software alternatives—IMOD, SerialEM, TomoPy, and Scipion—each with its unique capabilities in image processing. The critical criteria for these tests were the software’s ability to handle the specific data format of TEM images and their compatibility with the `transform.json` file format, a necessity for NERF models.

COLMAP Alternatives	Result with our TEM Images
IMOD	Didn’t work ✗
SerialEM	Didn’t work ✗
TomoPy	Didn’t work ✗
Scipion	Didn’t work ✗

Table 7.2: Summary of COLMAP Alternatives Experimented with TEM Images.

Our exploration into COLMAP alternatives for TEM data aimed to find tools for successful 3D reconstructions. We tested IMOD, SerialEM, TomoPy, and Scipion, each known for specific imaging strengths, to see if they could replace COLMAP. However, these alternatives fell short, struggling with TEM image specifics required for NERF modeling, such as precise camera positioning and compatibility with `transform.json` file format.

This led to the realization that the uniqueness of TEM images demands a novel approach, not just a replacement for COLMAP. We then shifted our focus to NeRFMM, a promising framework better suited for TEM images. Although not completely independent of COLMAP, NeRFMM represented a significant step forward, offering a new pathway for our research in applying NERF models to TEM data.

7.5 Experimental Analysis with TEM Dataset 3

In our study, we worked with eight distinct datasets, each offering different complexities. This section outlines the development of our entire architecture, as illustrated in Figure 5.3, from inception to completion. To streamline the experimental process, we chose **Dataset 3** (see Figure 4.6) as the primary focus for in-depth analysis. Our methodology involved initially tailoring the model and its hyperparameters to **Dataset 3**, ensuring optimal performance. Once satisfactory results were obtained with this dataset, we then extended the same model configuration to the other datasets. This strategy allowed for a thorough and consistent evaluation of the model's effectiveness across various datasets.

In this phase of our research, we present the cumulative results of our experiments, from model development to post-processing with denoisers and ESRGAN. We successfully improved the image quality across all datasets, achieving impressive metrics such as a PSNR value exceeding 38.70 and an SSIM value of approximately 0.95. These improvements marked a significant enhancement compared to the initial metrics, which started at 28 PSNR and 0.46 SSIM for Dataset 2.

7.5.1 Dataset 3: Training with 200 Epochs

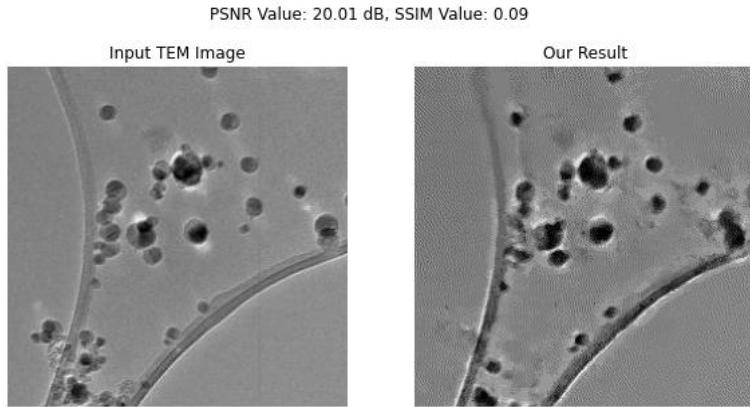


Figure 7.5: Dataset 3 with 200 epochs.

These results represent the preliminary stage of our model's development, indicating a PSNR value of 20.01 and an SSIM value of 0.09. During this initial phase, our model demonstrated the capability to enhance image edges and reduce noise compared to the ground truth image.

7.5.2 Dataset 3: Training with 10,000 Epochs

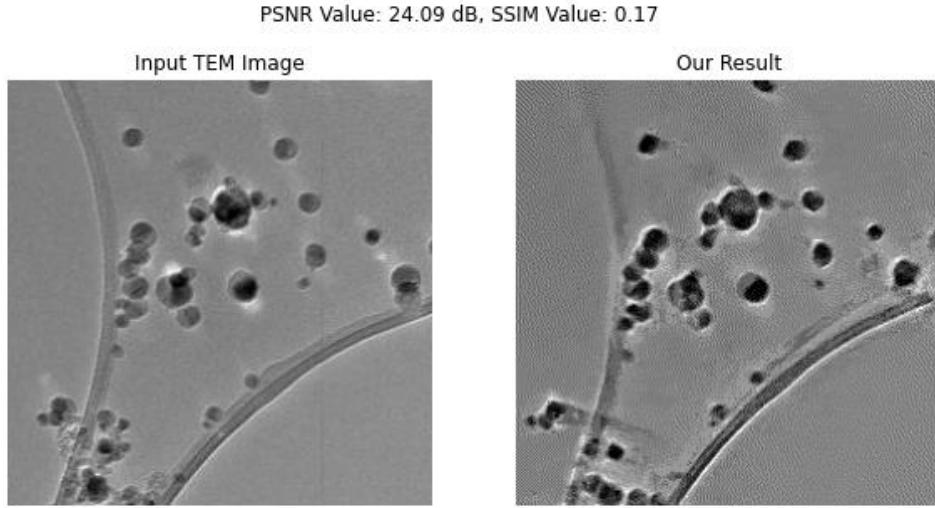


Figure 7.6: Dataset 3 with 10,000 epochs.

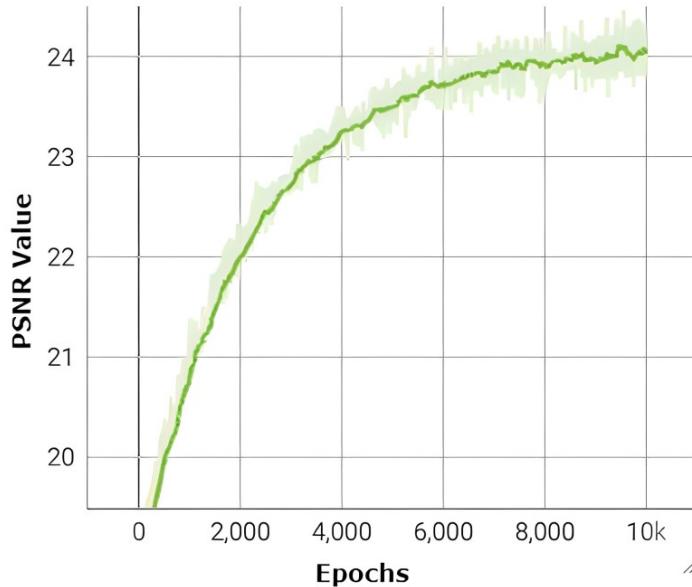


Figure 7.7: Progression of PSNR values over 10,000 epochs for Dataset 3

After undergoing an extensive training process spanning 10,000 epochs over a duration of roughly 8 to 9 hours, the model demonstrated a substantial enhancement in its performance. This intensive training phase led to a notable increase in the PSNR value, escalating from 20.01dB to 24.09dB, as depicted in 7.7, which denotes a significant improvement in image quality. While the PSNR value showed considerable improvement, the SSIM value, indicative of image structural similarity, remained relatively stable, exhibiting significant variations. This pattern suggests that despite the improved PSNR,

the model consistently maintained the structural integrity of the images throughout the training period.

7.5.3 Dataset 3: Camera Pose Optimization after 10,000 Epochs

Figure 7.8 showcases the model’s gradual improvement in optimizing camera positions across 10,000 training epochs. Starting from an initial position near (0,0) at epoch 0, the model progressively refines its camera parameter adjustments. This steady enhancement reflects the model’s growing ability to accurately interpret the spatial layout of the dataset. The graph highlights not only the model’s developing proficiency but also the importance of extensive training for precise camera positioning, a key challenge in NeRF models, particularly in overcoming limitations associated with tools like COLMAP.

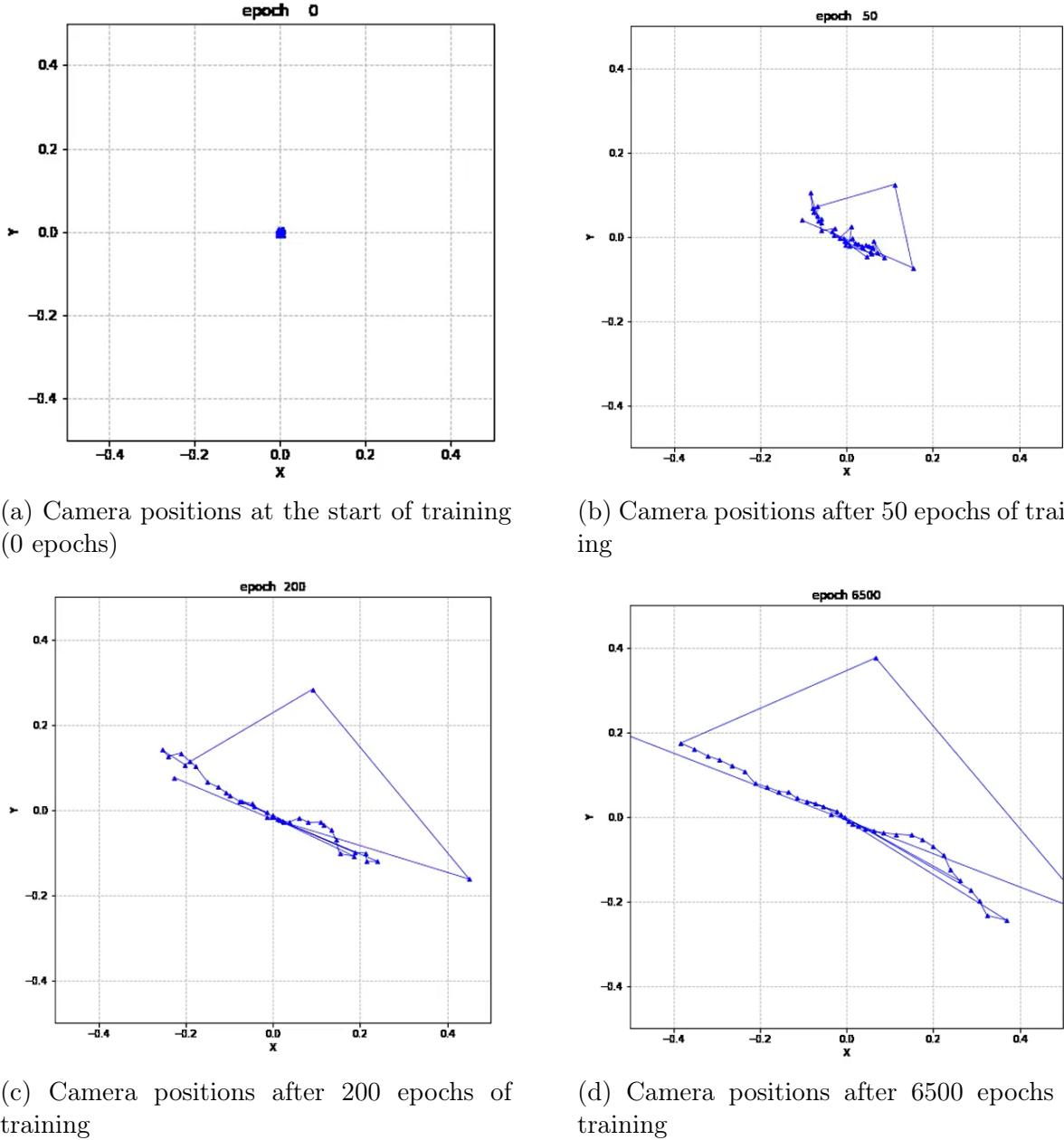


Figure 7.8: Progression of camera position optimization from 0 to 10,000 epochs in Dataset 3

7.5.4 Dataset 3: Exploring Different Activation Functions

Within our model, the primary activation function employed is the widely used **ReLU** (Rectified Linear Unit). However, we embarked on an exploration to investigate if there existed room for improvement in this crucial aspect. Our exploration involved the thorough examination of various alternative activation functions that have shown promise in image denoising tasks from NAN-NeRF [71].

To efficiently conduct our experiments, we initially executed them for 200 epochs, considering that training for 10,000 epochs demanded an extensive time investment of over 10 hours. This initial phase served as a screening process, allowing us to identify any promising results within a shorter timeframe. Subsequently, if any of these early experiments yielded favorable outcomes in terms of PSNR or SSIM values, we proceeded with the full 10,000-epoch training to assess if further improvements could be achieved.

1. Tahn activation function:

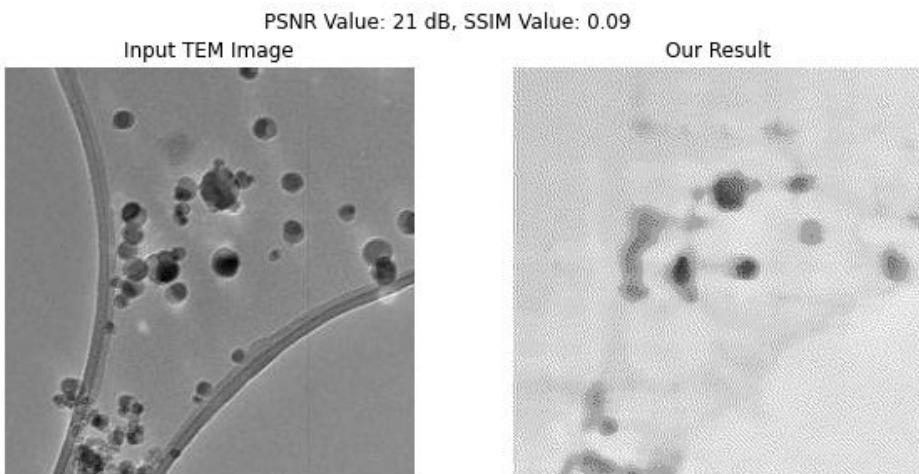


Figure 7.9: Dataset 3 with Tanh Activation function (200 epochs).

We initiated our exploration of alternative activation functions by training the model using the **Tanh** activation function for a limited 200 epochs. In this initial phase, the results indicated a PSNR value of 21 dB and an SSIM value of 0.09. Recognizing the potential for improvement, we extended the training duration to 10,000 epochs, resulting in a significant enhancement in performance. The model achieved a remarkable PSNR value of 28.30 dB and an SSIM value of 0.16.

While the performance boost with **Tanh** activation was notable, it's worth noting that our original choice of **ReLU** still outperformed it, consistently delivering the highest PSNR value of 28.37 dB. This comparison underscored the efficacy of **ReLU** in our model's context.

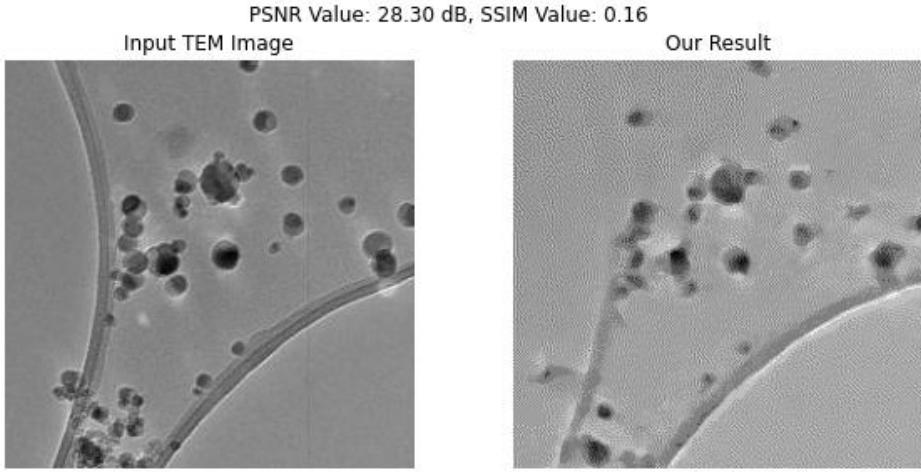


Figure 7.10: Dataset 3 with Tanh Activation function (10,000 epochs).

2. Sigmoid activation function:

In our experimental trials, we investigated the use of the **Sigmoid** activation function, initially training it for 200 epochs. However, this initial attempt yielded underwhelming results, as we did not calculate PSNR and SSIM values. Due to this lackluster performance in the early stages, we chose not to continue further experiments with the **Sigmoid** activation function and refrained from extending the training duration to 10,000 epochs.

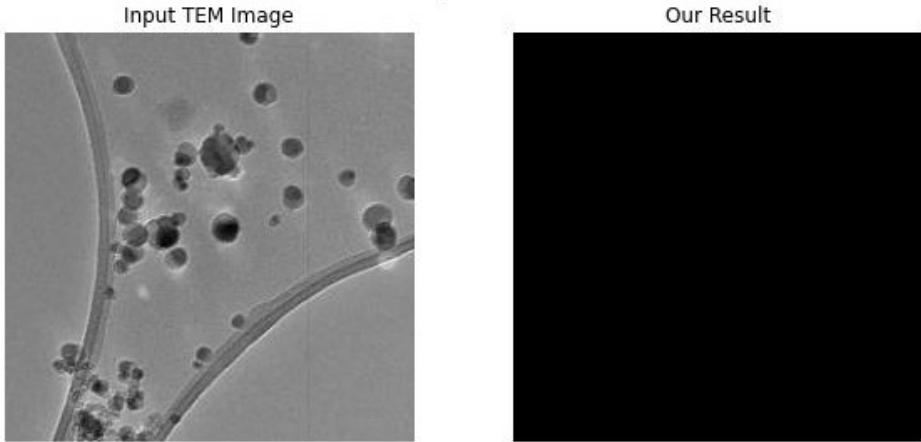


Figure 7.11: Dataset 3 with Sigmoid Activation function (200 epochs).

3. Leaky ReLU activation function:

Our exploration into activation functions culminated with the **Leaky ReLU** activation function. We initiated this experiment with 200 epochs and were pleasantly surprised by the results, which yielded a remarkable 28.30 dB PSNR value and a 0.16 SSIM value. Encouraged by this promising outcome, we decided to further investigate the potential of **Leaky ReLU** by extending the training to 10,000 epochs, despite the substantial time investment required.

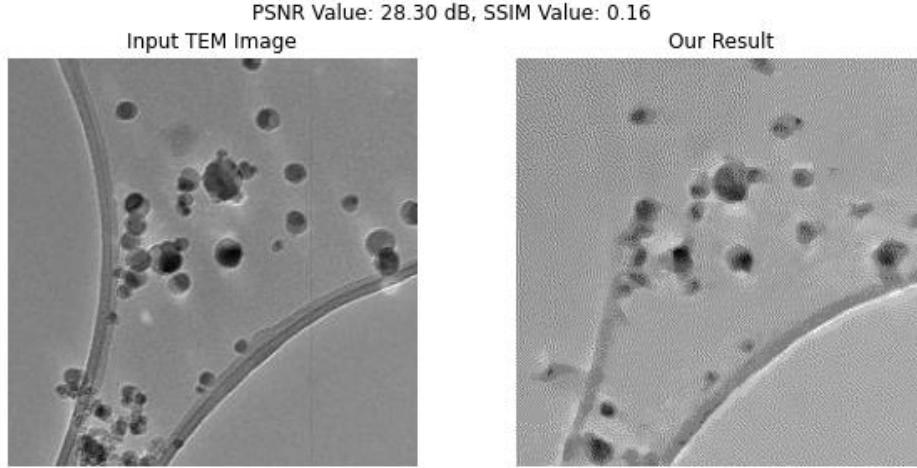


Figure 7.12: Dataset 3 with Leaky ReLU Activation function (200 epochs).

The extended training period, spanning over 10 hours, proved to be worthwhile. The PSNR value exhibited significant improvement, increasing from 28.30 dB to 28.38 dB. Even more noteworthy was the enhancement in the SSIM value, which escalated to 0.18. This achievement surpassed the performance of the original **ReLU** activation function, prompting us to adopt **Leaky ReLU** as the final activation function for our model.

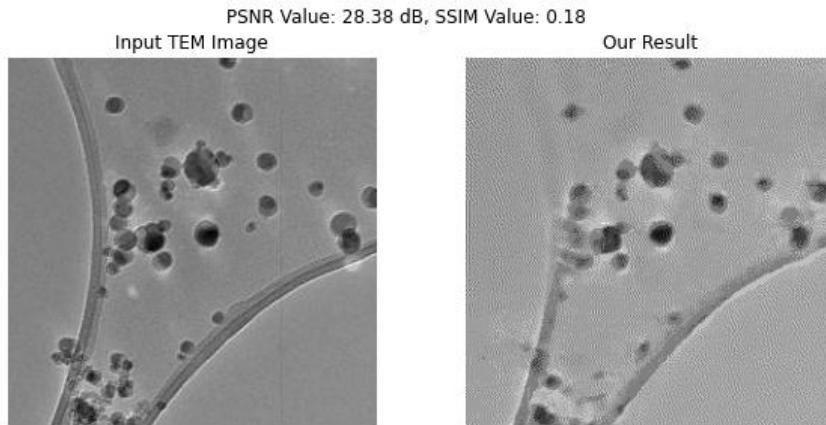


Figure 7.13: Dataset 3 with Leaky ReLU Activation function (10,000 epochs).

7.5.5 Dataset 3 with different Network layer

Our model initially consisted of a baseline configuration with a 4-layer neural network. In pursuit of improved performance, we conducted experiments to explore the impact of increasing the depth of our model. By incrementing the number of layers to 8, we witnessed a substantial enhancement in results 7.14.

The model with 8 layers achieved a noteworthy PSNR value of 29.84 dB and an SSIM value of 0.21. This outcome signified a significant improvement in the model's capacity to denoise and reconstruct images. The increased depth of the neural network contributed to

the model's ability to capture more intricate patterns and details in the images, resulting in superior image quality.

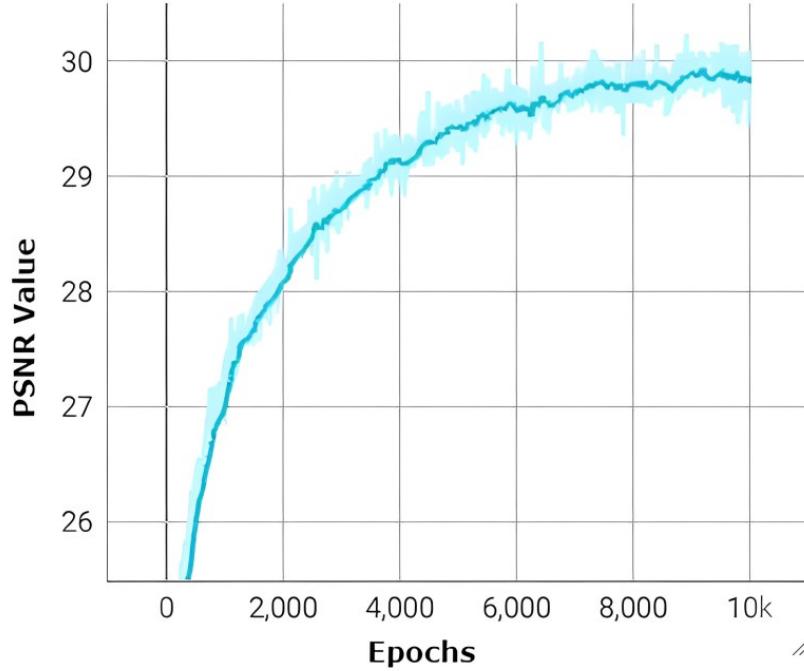


Figure 7.14: Progression of PSNR values over 10,000 epochs after adjusting hyperparameters for Dataset 3

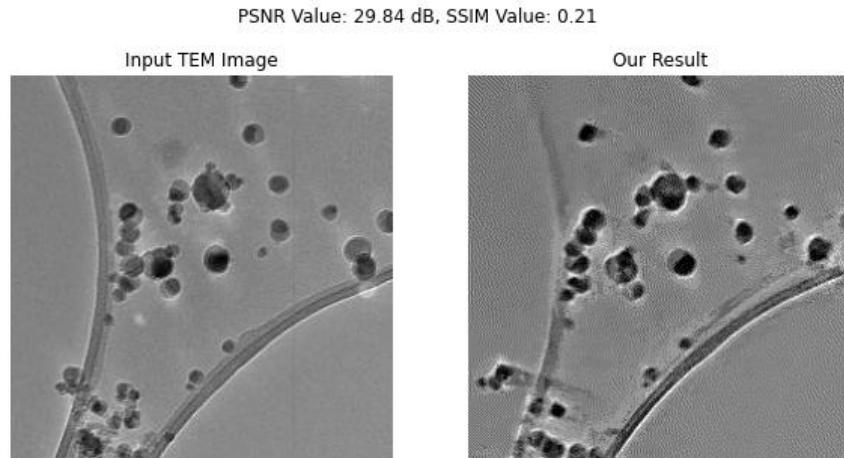


Figure 7.15: Dataset 3 with 8 Layer Architecture.

7.5.6 Enhancing the Model with NeRF-Dark and NAN-NeRF Parameters

Incorporating advanced techniques from NeRF-Dark and NAN-NeRF, we enhanced the quality of our image reconstructions, as detailed in 5 and 6. NeRF-Dark [70], addressed

the challenges of low-light conditions, common in TEM imaging. It enabled us to reconstruct high-quality images from previously unusable data. Additionally, NAN-NeRF [71], excelled at nanoscale imaging, facilitating precise 3D reconstructions from limited and noisy TEM data. These innovations culminated in impressive results, with our images achieving a peak PSNR value of 31.29 dB and an SSIM value of 0.56, showcasing enhanced clarity and reduced noise.

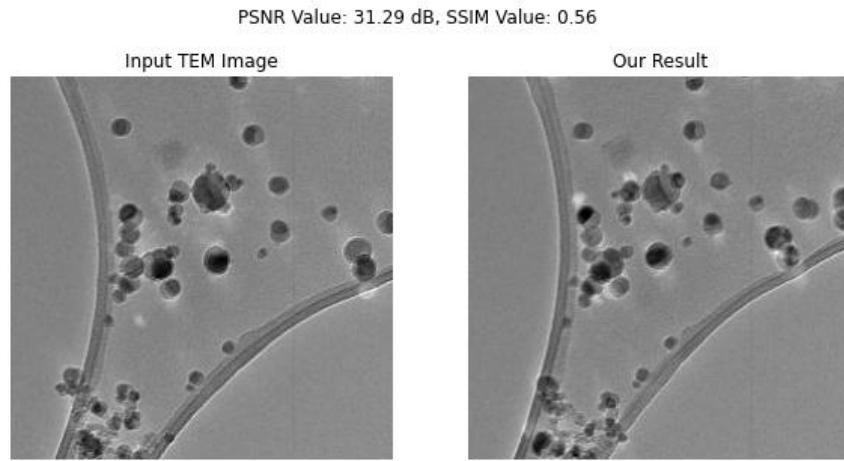


Figure 7.16: Remarkable Enhancements in Image Quality Through NeRF-Dark and NAN-NeRF Integration.

7.5.7 Postprocessing with Traditional Denoisers

In this section, we present the results of our experiments with traditional denoising techniques, as explained in Chapter 5. We evaluated methods including Bilateral Filtering, Gaussian Blur, Median Filtering, Non-Local Means Denoising, and Wavelet Denoising. Chosen for their established utility, these techniques range from Bilateral Filtering, known for edge preservation, to Wavelet Denoising, recognized for its selective denoising capability. The subsequent subsections detail the application of these methods to Dataset 3, showcasing the resulting postprocessed images and discussing their effectiveness.

1. Bilateral Filtering

Bilateral Filtering is a sophisticated denoising technique that excels in enhancing image quality by preserving edges and fine details while effectively reducing noise. In our experiment with Dataset 3, **Bilateral Filtering** achieved a PSNR value of **29.33 dB** and an SSIM value of **0.56**. These results highlight its capability to significantly improve image clarity and the discernibility of individual elements within the image.

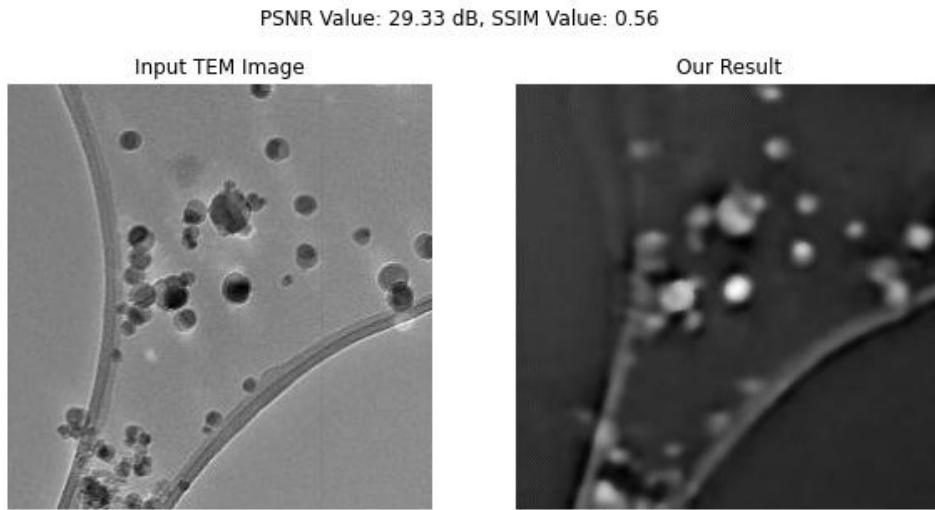


Figure 7.17: Image Postprocessed with Bilateral Filtering.

2. Gaussian Blur

Gaussian Blur is a widely used denoising approach that aims to reduce noise by averaging pixel values within a specified radius. In our experiment, **Gaussian Blur** produced results similar to Bilateral Filtering, with a PSNR value of **29.33 dB** and an SSIM value of **0.56**. While the performance closely aligns with that of Bilateral Filtering, Gaussian Blur remains a viable option for noise reduction.

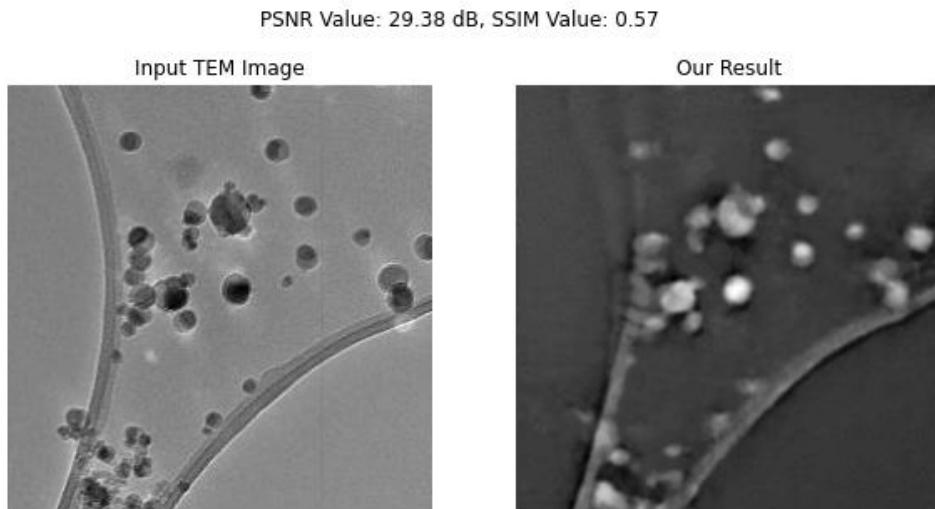


Figure 7.18: Image Postprocessed with Gaussian Blur.

3. Median Filtering

Median Filtering is a technique that effectively reduces noise while preserving image edges and structural details. It calculates the median value of neighboring pixels, making it

robust against outliers. In our experiment, **Median Filtering** achieved a PSNR value of **29.03 dB** and an SSIM value of **0.52**. These results demonstrate its effectiveness in enhancing image quality by improving particle visibility and edge definition.

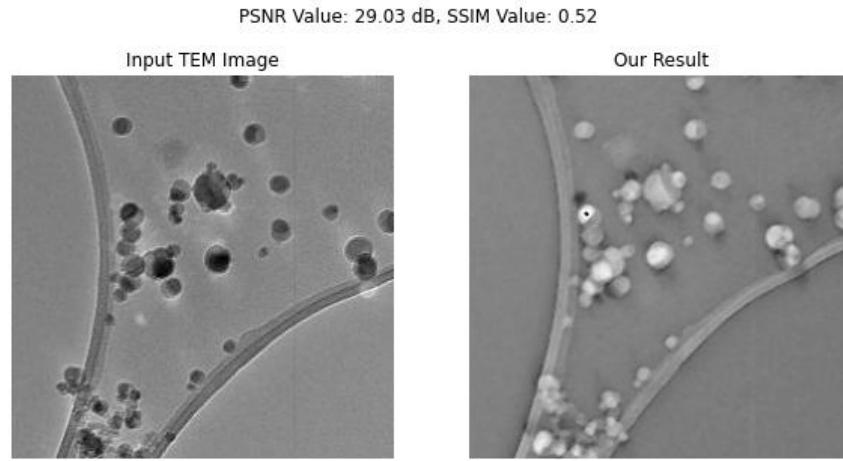


Figure 7.19: Image Postprocessed with Median Filtering.

4. Non-Local Means Denoising

Non-Local Means Denoising is a powerful technique known for its ability to effectively reduce noise while preserving image details. In our experiment, it achieved a remarkable PSNR value of **31.78 dB** and an SSIM value of **0.64**. However, the resulting image appeared slightly blurry, indicating a trade-off between noise reduction and image sharpness. Despite this trade-off, Non-Local Means Denoising remains a valuable tool for noise reduction tasks.

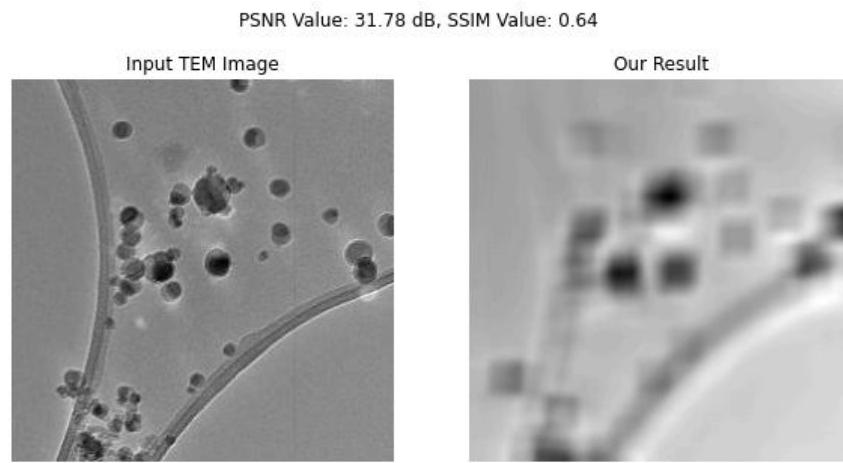


Figure 7.20: Image Postprocessed with Non-Local Means Denoising.

5. Wavelet Denoising

Wavelet Denoising is a sophisticated denoising technique that excels in noise reduction while preserving image details. It decomposes the image into different frequency components and applies denoising selectively. In our experiment, **Wavelet Denoising** outperformed the other techniques, achieving a PSNR value of **32.29 dB** and an impressive SSIM value of **0.68**. These results highlight its effectiveness in reducing noise and enhancing image quality, making it the preferred choice for Dataset 3.

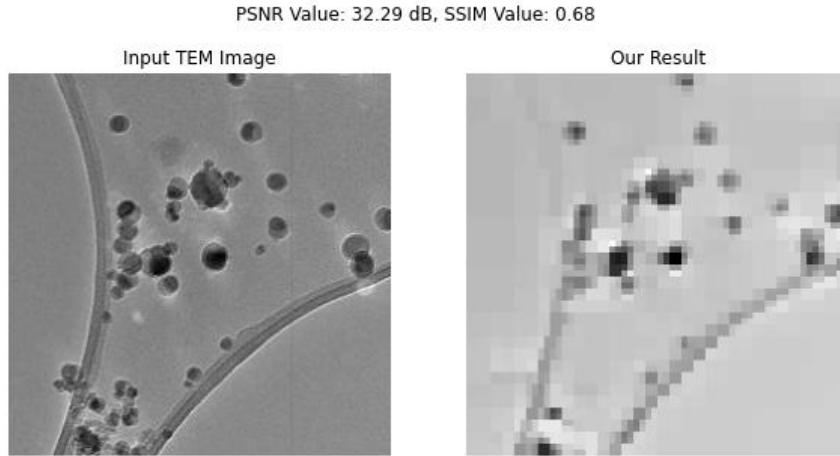


Figure 7.21: Image Postprocessed with Wavelet Denoising.

Among the five denoising techniques evaluated for Dataset 3, Wavelet Denoising emerged as the top performer, achieving a PSNR value of 32.29 dB and an SSIM value of 0.68. This makes Wavelet Denoising the preferred choice for enhancing the image quality of Dataset 3.

7.5.8 Postprocessing with ESRGAN

The final stage of our image enhancement architecture involves the application of Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN). ESRGAN is a powerful method that can infer high-resolution images up to four times larger than the input, making it superior to traditional denoising techniques. In this stage, we take the noisy and relatively small images generated by our model's output and significantly enhance their quality.

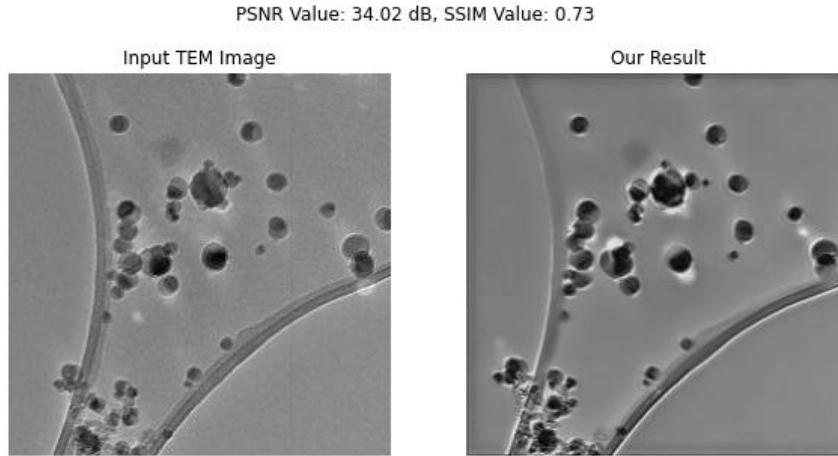


Figure 7.22: Image Postprocessed with ESRGAN.

ESRGAN has proven to be highly effective, achieving exceptional results in terms of both objective image quality metrics and perceptual quality. Specifically, **ESRGAN** yielded a remarkable **PSNR** value of **34.02 dB** and an impressive **SSIM** value of **0.73**. These metrics represent the highest values obtained throughout our entire experimentation process.

From a visual perspective, the images postprocessed with ESRGAN exhibit several noteworthy characteristics. First and foremost, the images are notably sharp, with enhanced clarity and well-defined edges. This sharpness extends to even the smallest particles within the image, further emphasizing the effectiveness of noise reduction. Moreover, ESRGAN excels in eliminating noisy artifacts from the background, resulting in cleaner and more visually pleasing images.

This strategic utilization of ESRGAN as the final step in our image processing pipeline culminates in exceptional outcomes, ensuring that our reconstructed TEM images exhibit the highest levels of clarity, fidelity, and noise reduction.

7.6 Experimental Analysis with TEM Dataset 1

In our experiments with Dataset 1 4.7, we initially expected ESRGAN to outperform other methods. However, the Non-Local Means algorithm delivered surprisingly superior results in this dataset.

Non-Local Means 7.23d achieved a PSNR value of 33.02 dB and an SSIM value of 0.80, surpassing ESRGAN, which had a PSNR of 31.07 dB and SSIM of 0.38. This outcome can be attributed to the dataset's predominantly black background, which favored Non-Local Means and Wavelet denoising, with a PSNR of 33.00 dB and an SSIM of 0.79.

In contrast, the other three denoising algorithms did not yield notably remarkable results, making them less relevant in this context. Therefore, we focused our attention on the outstanding performance of Non-Local Means and Wavelet denoising for Dataset 1.

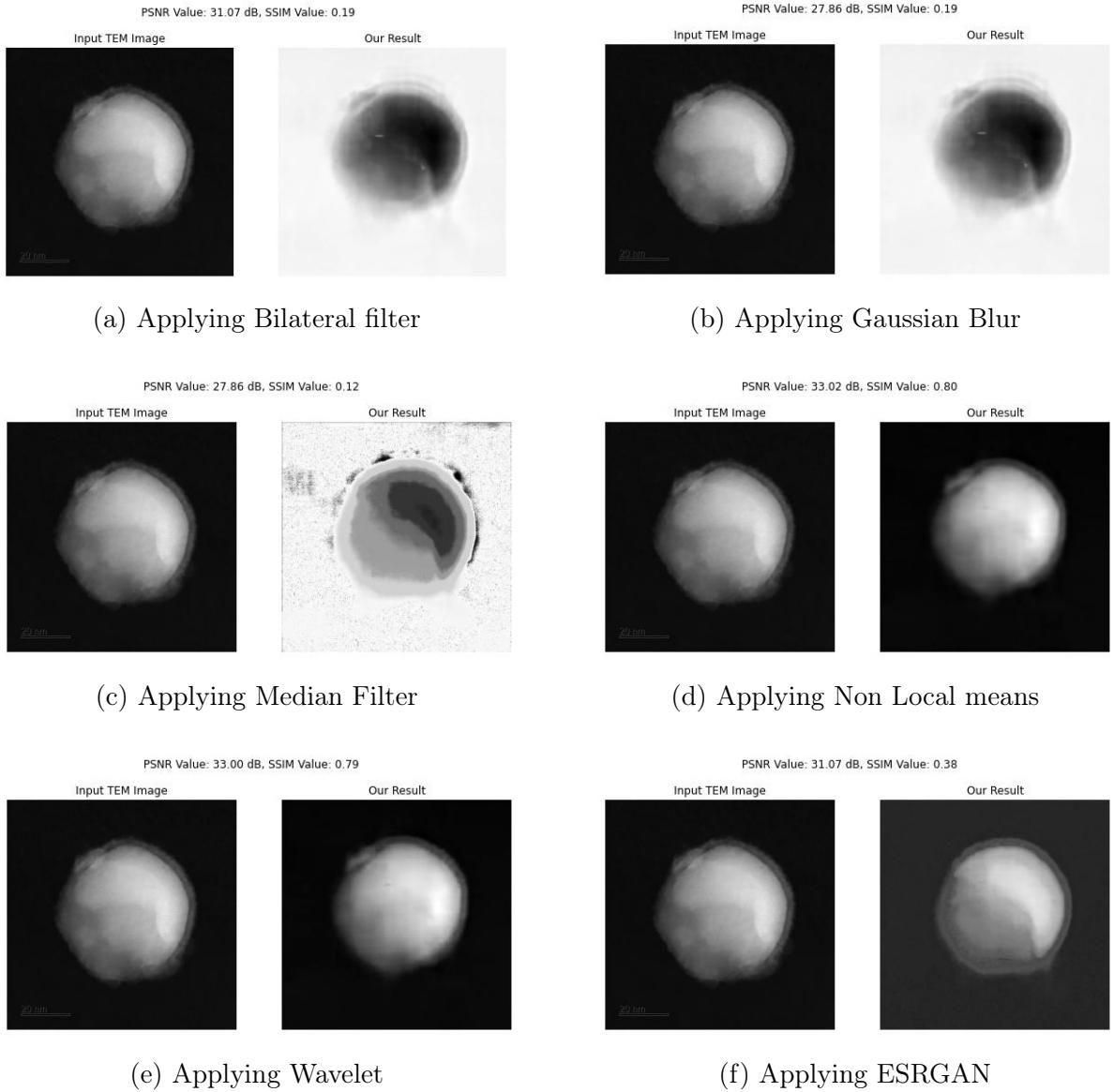


Figure 7.23: Experimental Results for TEM Dataset 1.

7.7 Experimental Analysis with TEM Dataset 2

TEM Dataset 2 (refer to Figure 4.4) exhibited exceptional denoising capabilities, particularly in dark backgrounds. ESRGAN delivered astounding results with a PSNR value of 38.70 dB and an SSIM value of 0.95, nearing the highest possible SSIM value of 1. Furthermore, Non-Local Means and Wavelet denoising also showcased commendable performance, achieving PSNR values of 34.54 dB and 34.57 dB, with SSIM values of 0.75 and 0.74, respectively.

In contrast, Bilateral Filter, Gaussian Blur, and Median Filter yielded average results, with PSNR values around 29.00 dB and SSIM values of approximately 0.70.

The superiority of ESRGAN(refer to figure 7.24f) in this dataset is evident, reaffirming its effectiveness in challenging denoising scenarios.

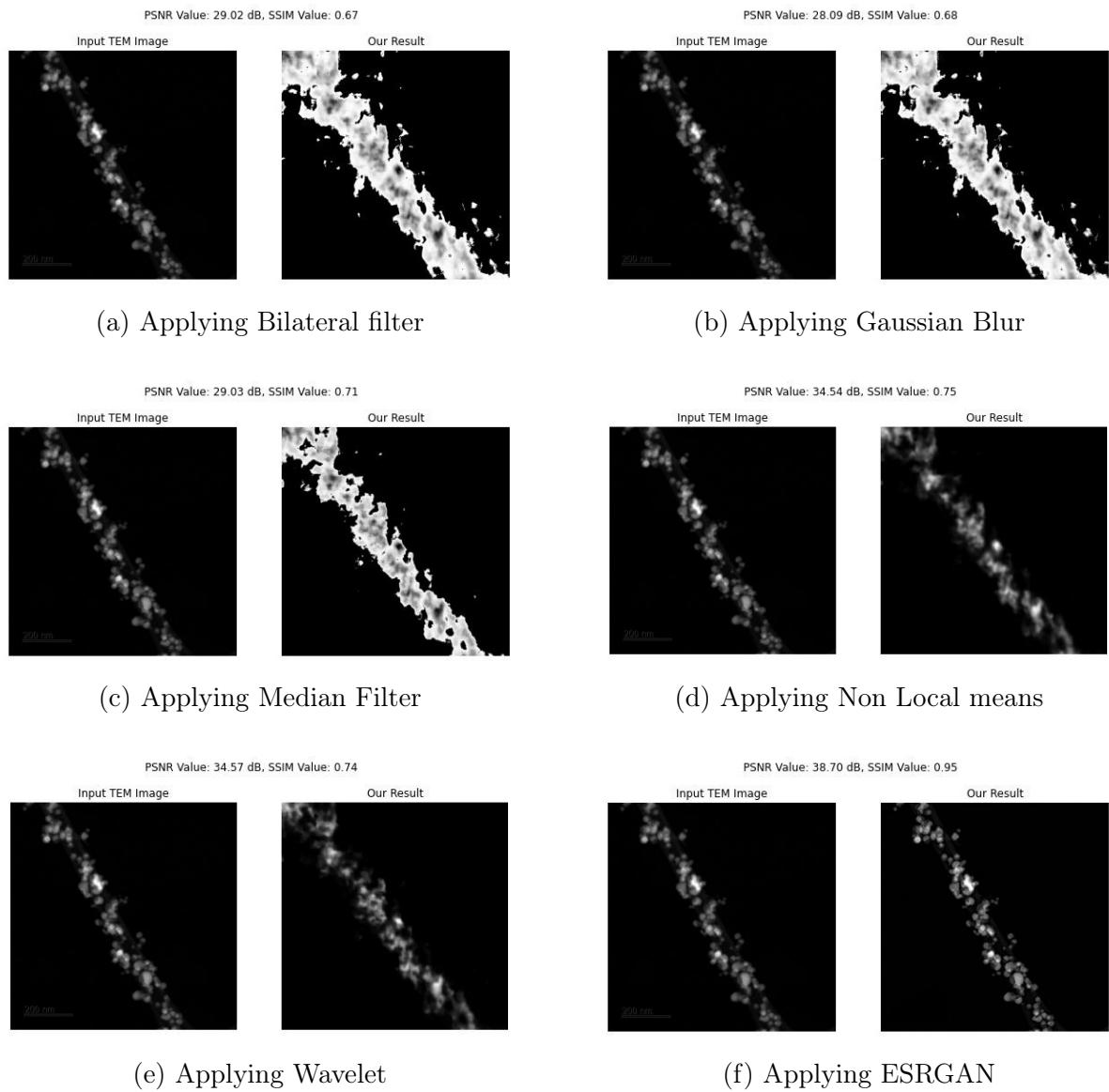


Figure 7.24: Experimental Results for TEM Dataset 2

7.8 Experimental Analysis with TEM Dataset 4

TEM Dataset 4, shown in Figure 4.8, consists of sharply structured particles against a gray background, aiding our model in 3D structure reconstruction. Its larger size of 1421 x 1421 pixels poses unique image processing challenges.

Regarding denoising performance, ESRGAN 7.25f excelled with the highest PSNR value of 31.56dB and an SSIM of 0.72, indicating superior image quality enhancement. The Wavelet method also performed comparably, achieving a PSNR of 31.50dB and an SSIM of 0.70, demonstrating its effectiveness in noise reduction.

The Non-Local method yielded a PSNR of 31.45dB and an SSIM of 0.70, closely matching the Wavelet's results and proving its efficiency. Meanwhile, the other three algorithms averaged a PSNR of around 29.50dB and an SSIM of 0.60, reflecting good image quality improvement, albeit slightly lower than the leading methods.

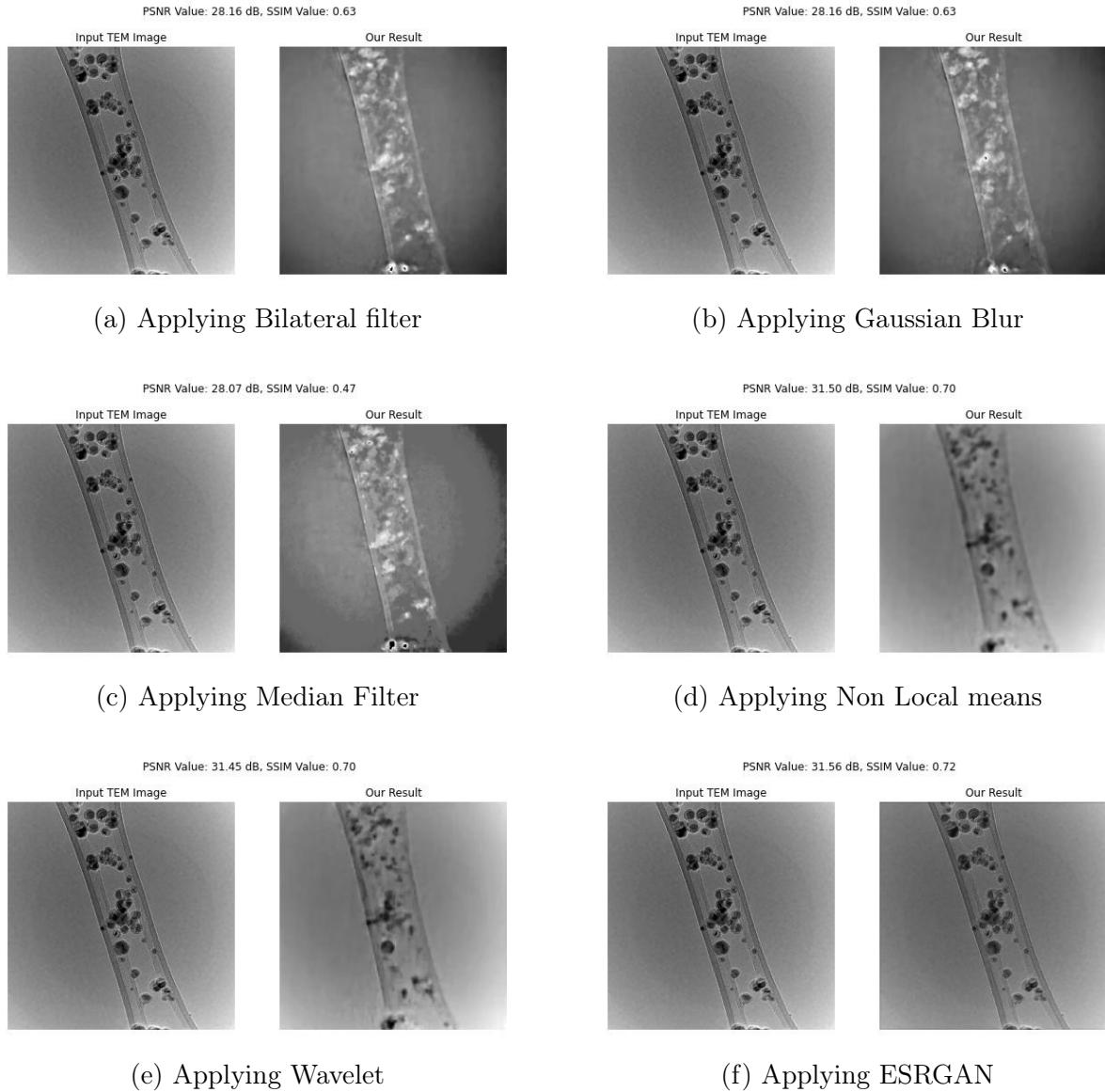


Figure 7.25: Experimental Results for TEM Dataset 4

7.9 Experimental Analysis with STEM Dataset 1

Figure 4.10 shows STEM Dataset 1, featuring a single particle with solid material attachment, posing a unique challenge in image processing. This dataset is the largest we tested, with images of 1024 x 1024 pixels, each about 1 MB.

Unexpectedly, ESRGAN registered the lowest PSNR at 25.06 dB in our experiments, while Wavelet 7.26e and Non-Local means 7.26d excelled with PSNRs of 34.45 and 34.42, respectively. In terms of SSIM, Non-Local means led with 0.84, followed by Wavelet at 0.83, indicating strong structural preservation. Despite its low PSNR, ESRGAN achieved a better SSIM of 0.62, suggesting decent structure retention. The Gaussian blur had the lowest SSIM at 0.10, and Bilateral Filter and Median methods showed moderate PSNR but lower SSIM values, close to the Gaussian blur's performance.

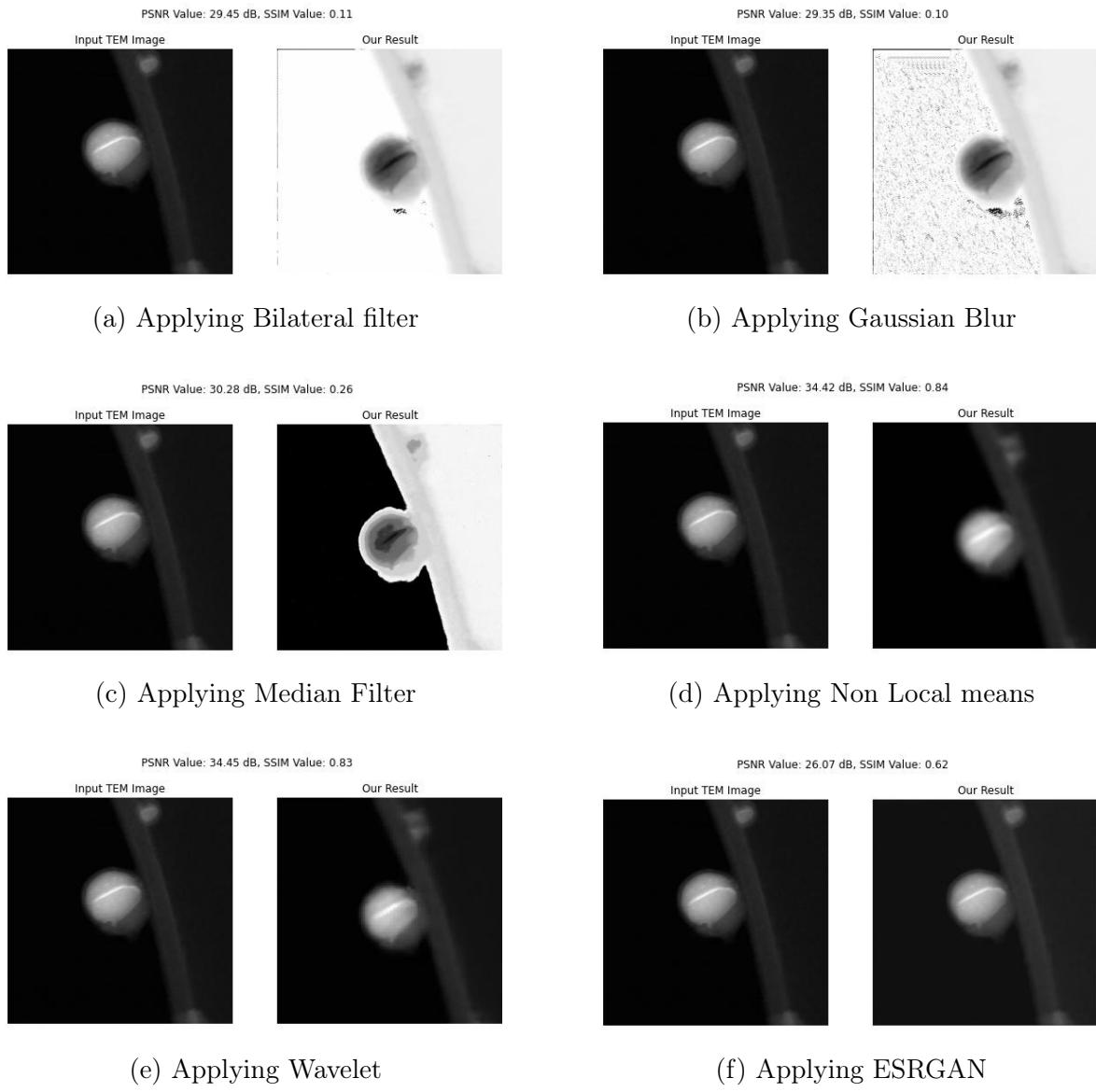
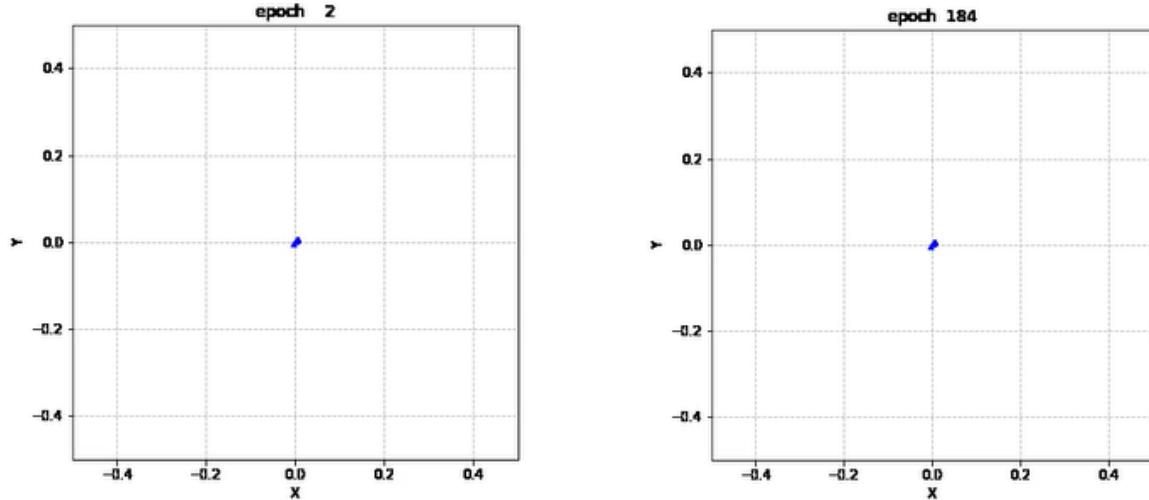


Figure 7.26: Experimental Results for STEM Dataset 1

7.10 Experimental Analysis with STEM Dataset 2

STEM Dataset 2 4.12 represents a challenging experiment that yielded unsatisfactory results. Our Method architecture, which is designed to generate 3D reconstructions from input images, produced completely dark outputs for this dataset. Upon closer examination of STEM Dataset 2 4.12, it becomes apparent that the dataset features a predominantly black and dark background. The images exhibit high levels of noise and generally poor image quality. Notably, all images in this dataset have a file size of 60KB and a resolution of 1024 x 1024 pixels, where our model effectively renders every pixel as a dark point.

A closer inspection of the initial camera position optimization in 7.27a reveals that the camera remains fixed at the (0,0) position. Even after 184 epochs, as shown in 7.27b, the camera position remains unchanged. This persistent stagnation indicates that our model struggles to detect any meaningful features within the images, resulting in a consistent camera position of (0,0) for all images. Consequently, this dataset experiment can be deemed a failure due to the inability of our model to discern relevant information within the images.



(a) Visualization of Camera Position Optimization After 2 Epochs

(b) Visualization of Camera Position Optimization After 184 Epochs

Figure 7.27: Experimental Results for STEM Dataset 2

7.11 Experimental Analysis with Synthetic Dataset

Our Synthetic Dataset experiment aimed to test our model's ability to reconstruct 3D structures. Despite the data being cleaner and clearer, the image size was notably small, around 80KB. This contrasts with a similar small-sized dataset (see Figure 4.12), where our model struggled to generate 3D structures. However, with 120 images in the Synthetic Dataset, we observed a positive outcome, as demonstrated in Figure 7.3, showing that our model can leverage a larger number of smaller images.

We successfully generated a 3D structure from the Synthetic Dataset, and our methods yielded impressive results. ESRGAN achieved the highest PSNR of 34.56 (see Figure 7.28f) with an SSIM of 0.75. Wavelet (Figure 7.28f) and Non-Local (Figure 7.28d) methods surpassed ESRGAN with SSIM values of 0.81 and 0.78, respectively. In contrast, Gaussian blur had the lowest PSNR at 26.19 dB, while Median means recorded the lowest SSIM at 0.41.

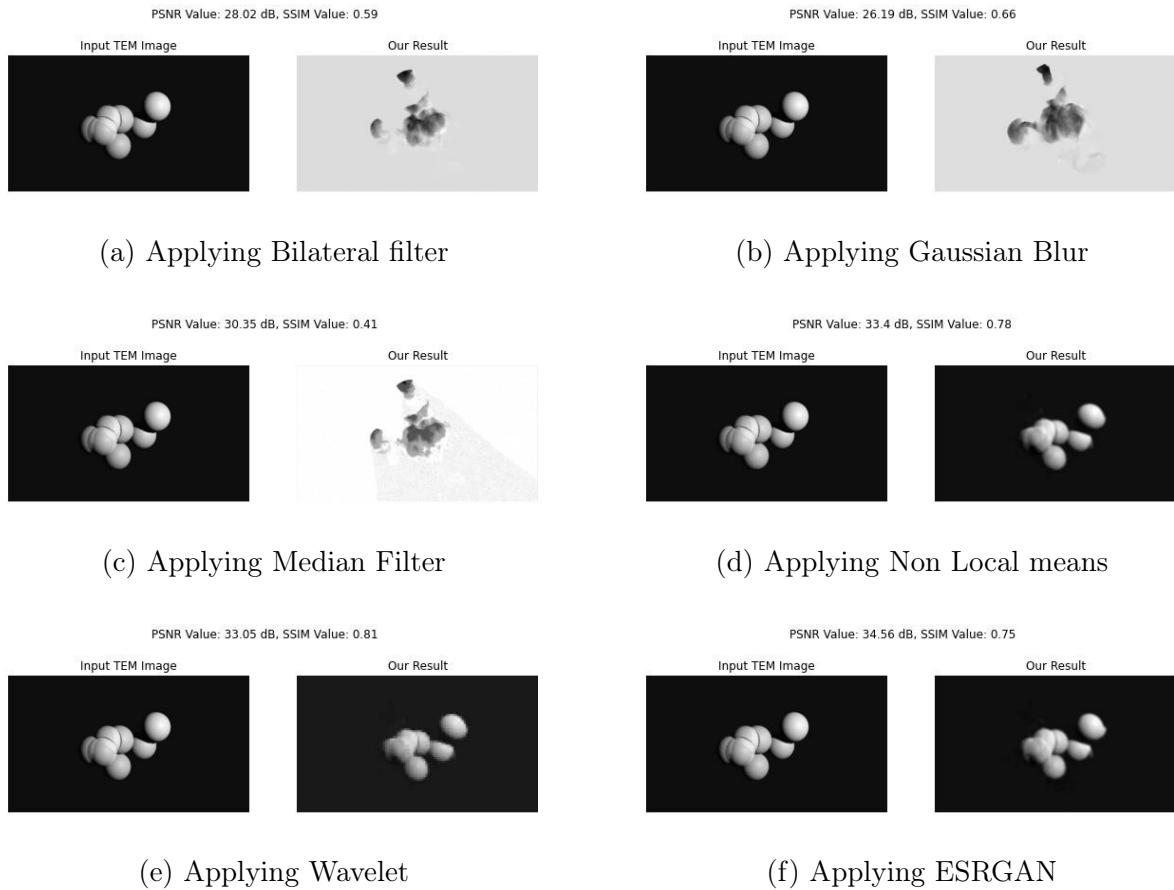


Figure 7.28: Experimental Results for Synthetic Dataset

7.12 Comparative PSNR & SSIM Analysis of Image Denoising

This overview presents the experimental results across all 8 datasets using our model. In the table, cells highlighted in **Green** indicate the **highest PSNR** and **highest SSIM** values for each dataset, while those in **Red** denote the **lowest** values.

Datasets	Bilateral	Gaussian	Median	Non-Local	Wavelet	ESRGAN
TEM Dataset 1	31.07	27.86	27.84	33.02	33.00	31.07
TEM Dataset 2	29.02	28.09	29.03	34.54	34.57	38.70
TEM Dataset 3	29.33	29.38	29.03	31.78	32.29	34.02
TEM Dataset 4	28.16	28.16	28.07	31.50	31.45	31.56
STEM Dataset 1	29.45	29.35	30.28	34.42	34.45	26.07
STEM Dataset 2	X	X	X	X	X	X
Synthetic Dataset	28.02	26.19	30.35	33.4	33.05	34.56

Table 7.3: Comparative Analysis of Image Denoising Techniques Across TEM and STEM Datasets Measured by **Peak Signal-to-Noise Ratio (PSNR)**

Datasets	Bilateral	Gaussian	Median	Non-Local	Wavelet	ESRGAN
TEM Dataset 1	0.19	0.19	0.12	0.80	0.79	0.38
TEM Dataset 2	0.67	0.68	0.71	0.75	0.74	0.95
TEM Dataset 3	0.56	0.57	0.52	0.64	0.68	0.73
TEM Dataset 4	0.63	0.63	0.47	0.70	0.70	0.72
STEM Dataset 1	0.11	0.10	0.26	0.84	0.83	0.62
STEM Dataset 2	X	X	X	X	X	X
Synthetic Dataset	0.59	0.66	0.41	0.78	0.81	0.75

Table 7.4: Evaluation of Image Denoising Efficacy on Various Datasets Using the **Structural Similarity Index Measure (SSIM)**

CHAPTER 8

Conclusions

This thesis has successfully demonstrated significant advancements in denoising 3D Transmission Electron Microscopy (TEM) tomography through the use of Advanced Neural Radiance Fields (NeRF). Our rigorous experimentation and thorough analysis across various datasets have highlighted NeRF's capability to markedly enhance the clarity and interpretability of TEM images, effectively addressing the high noise levels commonly found in such images.

In our study, we tackled the challenge of processing TEM images with NeRF, which was previously a significant concern. Our approach involved adapting NeRF models such as NeRFMM, NAN-NeRF, and NeRF in the Dark, and combining them with various denoising techniques, including ESRGAN. Remarkably, our method achieved impressive results, reaching a maximum PSNR of 38.70 and an SSIM value of 0.95 with our refined solution. Interestingly, traditional methods like Non-Local Means and Wavelet denoising at times outperformed ESRGAN, while other denoisers such as Bilateral, Gaussian, and Median consistently yielded suboptimal results. These findings provide valuable insights for future researchers, suggesting alternative directions to explore beyond these algorithms. Our work significantly improved image quality, as proven by extensive PSNR and SSIM analysis, demonstrating the efficacy of these techniques in enhancing the clarity of TEM data.

Although our solution did not succeed for the STEM Dataset 2, it was effective for all other datasets we experimented with. This research contributes to the fields of computational imaging and electron microscopy, showcasing how NeRF can be effectively utilized to create and denoise 3D TEM images.

CHAPTER 9

Future Scope

As we conclude this thesis, we recognize that the field of 3D TEM tomography and Neural Radiance Fields (NeRF) is rapidly evolving, offering numerous avenues for future research. Our work has laid a foundation, but there is substantial scope for enhancement and exploration. In this chapter, we outline potential directions for future work.

9.1 Advancements in NeRF Models

The ongoing development and refinement of Neural Radiance Fields (NeRF) models offer significant opportunities for future research. One key area is enhancing model efficiency, particularly in terms of computational speed and resource consumption. Another crucial aspect is the development of models capable of handling more challenging scenarios, such as higher noise levels and complex image backgrounds.

As highlighted in our 7 chapter, a promising direction is the exploration of LU-NeRF, a potential NeRF model whose codebase has not yet been published. Implementing and testing LU-NeRF to ascertain if it offers improvements over our current architecture represents a vital area of future work. Additionally, recent advancements have seen the emergence of advanced NeRF models built upon the foundation similar to our NeRFMM. Investigating these models could provide valuable insights and enhancements for processing TEM images.

9.2 Extending Dataset Variety

To enhance the robustness and validate the effectiveness of Neural Radiance Fields (NeRF) models, diversifying the range of datasets used for training and testing is essential. Broadening the scope to include a more extensive array of TEM data, characterized by various noise levels and encompassing different biological and material samples, is crucial for this advancement.

In our research, we encountered challenges in training certain datasets, notably as seen with 4.12. Despite this, we hypothesize that our architecture has the potential to handle a wide range of TEM and STEM data. By exploring a more diverse set of datasets, we aim to uncover additional insights into the capabilities and limitations of our method. Furthermore, minor modifications to our architecture could potentially address the issues encountered with specific datasets like STEM Dataset 2. Such refinements would not

only aid in resolving current challenges but also contribute to the broader applicability and effectiveness of NeRF models in diverse imaging contexts.

9.3 Development of User-Friendly Software Tools

Enhancing the accessibility of Neural Radiance Fields (NeRF) through the development of user-friendly software tools is crucial in democratizing this technology for TEM tomography. Such tools should be designed to cater to a wide spectrum of users, ranging from researchers to professionals in various fields. A prime example to emulate is NerfStudio. This platform allows users to effortlessly upload their datasets and automatically generate outputs based on the methodologies discussed in this thesis. Incorporating features for hyperparameter tuning, similar to those found in NerfStudio, would significantly benefit material scientists and other researchers. This approach not only simplifies the process of using NeRF but also encourages broader experimentation and application in diverse scientific domains.

Bibliography

- [1] Achilleas S. Frangakis. “It’s noisy out there! A review of denoising techniques in cryo-electron tomography”. In: *Journal of Structural Biology* 213 (4 Dec. 2021), p. 107804. ISSN: 1047-8477. DOI: 10.1016/J.JSB.2021.107804.
- [2] David C. Joy. “Noise and Its Effects on the Low-Voltage SEM”. In: *Biological Low-Voltage Scanning Electron Microscopy* (Nov. 2008), pp. 129–144. DOI: 10.1007/978-0-387-72972-5_4. URL: https://link.springer.com/chapter/10.1007/978-0-387-72972-5_4.
- [3] Achilleas S. Frangakis and Reiner Hegerl. “Noise reduction in electron tomographic reconstructions using nonlinear anisotropic diffusion”. In: *Journal of structural biology* 135 (3 2001), pp. 239–250. ISSN: 1047-8477. DOI: 10.1006/JSBI.2001.4406. URL: <https://pubmed.ncbi.nlm.nih.gov/11722164/>.
- [4] Vicente González-Ruiz and Jose Jesus Fernández. “FlowDenoising: Structure-preserving denoising in 3D electron microscopy (3DEM)”. In: *SoftwareX* 23 (July 2023), p. 101413. ISSN: 2352-7110. DOI: 10.1016/J.SOFTX.2023.101413.
- [5] Yunjin Chen and Thomas Pock. “Trainable Nonlinear Reaction Diffusion: A Flexible Framework for Fast and Effective Image Restoration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6 June 2017), pp. 1256–1272. ISSN: 01628828. DOI: 10.1109/TPAMI.2016.2596743.
- [6] Xin Deng and Pier Luigi Dragotti. “Deep Convolutional Neural Network for Multi-modal Image Restoration and Fusion”. In: () .
- [7] Ben Mildenhall et al. “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 12346 LNCS (Mar. 2020), pp. 405–421. ISSN: 16113349. DOI: 10.1007/978-3-030-58452-8_24. URL: <https://arxiv.org/abs/2003.08934v2>.
- [8] R. F. Egerton, P. Li, and M. Malac. “Radiation damage in the TEM and SEM”. In: *Micron* 35 (6 Aug. 2004), pp. 399–409. ISSN: 0968-4328. DOI: 10.1016/J.MICRON.2004.02.003.
- [9] Baptiste Gault et al. “Estimation of the Reconstruction Parameters for Atom Probe Tomography”. In: *Microscopy and Microanalysis* 14 (4 Aug. 2008), pp. 296–305. ISSN: 1431-9276. DOI: 10.1017/S1431927608080690. URL: <https://dx.doi.org/10.1017/S1431927608080690>.
- [10] C. Y. Tang and Z. Yang. “Transmission Electron Microscopy (TEM)”. In: *Membrane Characterization* (Jan. 2017), pp. 145–159. DOI: 10.1016/B978-0-444-63776-5.00008-5.

- [11] Marc Adrian et al. “Cryo-electron microscopy of viruses”. In: *Nature* 1984 308:5954 308 (5954 1984), pp. 32–36. ISSN: 1476-4687. DOI: 10.1038/308032a0. URL: <https://www.nature.com/articles/308032a0>.
- [12] Alexis Loiseau et al. “Silver-Based Plasmonic Nanoparticles for and Their Use in Biosensing”. In: (2019). DOI: 10.3390/bios9020078. URL: www.mdpi.com/journal/biosensors.
- [13] Jeong Joon Park et al. “DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019-June (Jan. 2019), pp. 165–174. ISSN: 10636919. DOI: 10.1109/CVPR.2019.00025. URL: <https://arxiv.org/abs/1901.05103v1>.
- [14] Lars Mescheder et al. “Occupancy Networks: Learning 3D Reconstruction in Function Space”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2019-June (Dec. 2018), pp. 4455–4465. ISSN: 10636919. DOI: 10.1109/CVPR.2019.00459. URL: <https://arxiv.org/abs/1812.03828v2>.
- [15] Vincent Sitzmann et al. “Implicit Neural Representations with Periodic Activation Functions”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 7462–7473.
- [16] Zhirong Wu et al. “3D ShapeNets: A deep representation for volumetric shapes”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 07-12-June-2015 (Oct. 2015), pp. 1912–1920. ISSN: 10636919. DOI: 10.1109/CVPR.2015.7298801.
- [17] Weihao Xia and Jing-Hao Xue. “A Survey on 3D-aware Image Synthesis”. In: (). URL: <https://weihaox.github.io>.
- [18] Steven M. Seitz et al. “A comparison and evaluation of multi-view stereo reconstruction algorithms”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 1 (2006), pp. 519–526. ISSN: 10636919. DOI: 10.1109/CVPR.2006.19.
- [19] Shenchang Eric Chen and Lance Williams. “View Interpolation for Image Synthesis”. In: *Seminal Graphics Papers: Pushing the Boundaries, Volume 2* (Aug. 2023), pp. 423–432. DOI: 10.1145/3596711.3596757. URL: <https://dl.acm.org/doi/10.1145/3596711.3596757>.
- [20] Richard Hartley and Andrew Zisserman. “Multiple View Geometry in Computer Vision, Second Edition”. In: (2000). URL: www.cambridge.org/9780521540513.
- [21] Jiemin Fang et al. “Fast Dynamic Radiance Fields with Time-Aware Neural Voxels TiNeuVox (ours) 1 min 4 mins 8 mins Sparse Time-View Input Images Time Synthesis View Synthesis Fast Training for Time-View Synthesis D-NeRF Training Time”. In: (). DOI: 10.1145/3550469.3555383. URL: <https://doi.org/10.1145/3550469.3555383>.
- [22] Janne Heikkila and Olli Silven. “Four-step camera calibration procedure with implicit image correction”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1997), pp. 1106–1112. ISSN: 10636919. DOI: 10.1109/CVPR.1997.609468.

- [23] Zhengyou Zhang. “A flexible new technique for camera calibration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (11 Nov. 2000), pp. 1330–1334. ISSN: 01628828. DOI: 10.1109/34.888718.
- [24] Zhengyou Zhang et al. “A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry”. In: *Artificial Intelligence* 78 (1-2 Oct. 1995), pp. 87–119. ISSN: 0004-3702. DOI: 10.1016/0004-3702(95)00022-4.
- [25] Johannes L Schönberger and Jan-Michael Frahm. “Structure-from-Motion Revisited”. In: (). URL: <https://github.com/colmap/colmap..>
- [26] David G. Lowe. “Object recognition from local scale-invariant features”. In: *Proceedings of the IEEE International Conference on Computer Vision* 2 (1999), pp. 1150–1157. DOI: 10.1109/ICCV.1999.790410.
- [27] David G. Lowe. “Distinctive image features from scale-invariant keypoints”. In: *International Journal of Computer Vision* 60 (2 Nov. 2004), pp. 91–110. ISSN: 09205691. DOI: 10.1023/B:VISI.0000029664.99615.94/METRICS. URL: <https://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94>.
- [28] Martin A. Fischler and Robert C. Bolles. “Random sample consensus”. In: *Communications of the ACM* 24 (6 June 1981), pp. 381–395. ISSN: 15577317. DOI: 10.1145/358669.358692. URL: <https://dl.acm.org/doi/10.1145/358669.358692>.
- [29] Noah Snavely et al. “Photo Tourism: Exploring Photo Collections in 3D”. In: (). URL: www.cc.gatech.edu/4d-cities.
- [30] Jared Heinly et al. “Reconstructing the World* in Six Days *(As Captured by the Yahoo 100 Million Image Dataset)”. In: () .
- [31] Bill Triggs et al. “Bundle Adjustment-A Modern Synthesis”. In: () .
- [32] Silvano Galliani et al. “Massively Parallel Multiview Stereopsis by Surface Normal Diffusion”. In: () .
- [33] Gabriele Facciolo, Carlo de Franchis, and Enric Meinhardt. “MGM: A Significantly More Global Matching for Stereovision”. In: (Dec. 2015), pp. 90.1–90.12. DOI: 10.5244/C.29.90.
- [34] Michael Kazhdan, Johns Hopkins University, and Hugues Hoppe. “Screened Poisson Surface Reconstruction”. In: *ACM Trans. Graph. NN, N, Article NN (Month YYYY)* (). DOI: 10.1145/XXXXXXX.YYYYYYY. URL: <http://doi.acm.org/10.1145/XXXXXXX.YYYYYYY>.
- [35] Alexander Kolesnikov et al. *Revisiting Self-Supervised Visual Representation Learning*. 2019. URL: <https://github.com/google/revisiting-self-supervised>.
- [36] Chiyu Max Jiang et al. “Local Implicit Grid Representations for 3D Scenes”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Mar. 2020), pp. 6000–6009. ISSN: 10636919. DOI: 10.1109/CVPR42600.2020.00604. URL: <https://arxiv.org/abs/2003.08981v1>.
- [37] Xintao Wang et al. “ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks”. In: (). URL: <https://github.com/xinntao/ESRGAN..>

- [38] Christian Ledig et al. “Photo-realistic single image super-resolution using a generative adversarial network”. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-January (Nov. 2017), pp. 105–114. DOI: 10.1109/CVPR.2017.19.
- [39] Yochai Blau et al. “The 2018 PIRM Challenge on Perceptual Image Super-resolution”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 11133 LNCS (Sept. 2018), pp. 334–355. ISSN: 16113349. DOI: 10.1007/978-3-030-11021-5_21. URL: <https://arxiv.org/abs/1809.07517v3>.
- [40] Ruoqian Lin et al. “TEMImageNet Training Library and AtomSegNet Deep-Learning Models for High-Precision Atom Segmentation, Localization, Denoising, and Super-Resolution Processing of Atomic-Resolution Images”. In: (Dec. 2020). URL: <https://arxiv.org/abs/2012.09093v2>.
- [41] Albert Lawrence et al. “Transform-based backprojection for volume reconstruction of large format electron microscope tilt series”. In: *Journal of structural biology* 154 (2 May 2006), pp. 144–167. ISSN: 1047-8477. DOI: 10.1016/J.JSB.2005.12.012. URL: <https://pubmed.ncbi.nlm.nih.gov/16542854/>.
- [42] Christian Widmer et al. “GRED: Graph-Regularized 3D Shape Reconstruction from Highly Anisotropic and Noisy Images”. In: *Signal, Image and Video Processing* 8 (1 Sept. 2013), pp. 41–48. ISSN: 18631711. DOI: 10.1007/s11760-014-0694-8. URL: <https://arxiv.org/abs/1309.4426v1>.
- [43] C. O.S. Sorzano et al. “Transfer function restoration in 3D electron microscopy via iterative data refinement”. In: *Physics in medicine and biology* 49 (4 Feb. 2004), pp. 509–522. ISSN: 0031-9155. DOI: 10.1088/0031-9155/49/4/003. URL: <https://pubmed.ncbi.nlm.nih.gov/15005161/>.
- [44] Kai Zhang et al. “Learning Deep CNN Denoiser Prior for Image Restoration”. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-January (Apr. 2017), pp. 2808–2817. DOI: 10.1109/CVPR.2017.300. URL: <https://arxiv.org/abs/1704.03264v1>.
- [45] Jose Jesus Fernandez. “Computational methods for electron tomography”. In: *Micron* 43 (10 Oct. 2012), pp. 1010–1030. ISSN: 0968-4328. DOI: 10.1016/J.MICRON.2012.05.003.
- [46] Gaby N. Moawad et al. “Augmented Realities, Artificial Intelligence, and Machine Learning: Clinical Implications and How Technology Is Shaping the Future of Medicine”. In: *Journal of Clinical Medicine* 9 (12 Dec. 2020), pp. 1–7. ISSN: 20770383. DOI: 10.3390/JCM9123811. URL: [/pmc/articles/PMC7761251/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7761251/)?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7761251/.
- [47] Kazuaki Kawahara et al. “Atomic-resolution STEM image denoising by total variation regularization”. In: *Microscopy (Oxford, England)* 71 (5 Oct. 2022), pp. 302–310. ISSN: 2050-5701. DOI: 10.1093/JMICRO/DFAC032. URL: <https://pubmed.ncbi.nlm.nih.gov/35713554/>.

- [48] Ondrej L. Krivanek et al. “Atom-by-atom structural and chemical analysis by annular dark-field electron microscopy”. In: *Nature* 464 (7288 Mar. 2010), pp. 571–574. ISSN: 1476-4687. DOI: 10.1038/NATURE08879. URL: <https://pubmed.ncbi.nlm.nih.gov/20336141/>.
- [49] Daniel S. Wastl, Alfred J. Weymouth, and Franz J. Giessibl. “Optimizing atomic resolution of force microscopy in ambient conditions”. In: *Physical Review B - Condensed Matter and Materials Physics* 87 (24 Mar. 2013). DOI: 10.1103/PhysRevB.87.245415. URL: <http://arxiv.org/abs/1303.5204%20http://dx.doi.org/10.1103/PhysRevB.87.245415>.
- [50] Jianwei Miao, Peter Ercius, and Simon J.L. Billinge. “Atomic electron tomography: 3D structures without crystals”. In: *Science* 353 (6306 Sept. 2016), aaf2157–aaf2157. ISSN: 0036-8075. DOI: 10.1126/SCIENCE.AAF2157. URL: <http://science.sciencemag.org/>.
- [51] Ian MacLaren and Quentin M. Ramasse. “Aberration-corrected scanning transmission electron microscopy for atomic-resolution studies of functional oxides”. In: *International Materials Reviews* 59 (3 Apr. 2014), pp. 115–131. ISSN: 09506608. DOI: 10.1179/1743280413Y.0000000026. URL: <https://www.tandfonline.com/doi/abs/10.1179/1743280413Y.0000000026>.
- [52] David A. Muller. “Structure and bonding at the atomic scale by scanning transmission electron microscopy”. In: *Nature Materials* 2009 8:4 8 (4 2009), pp. 263–270. ISSN: 1476-4660. DOI: 10.1038/nmat2380. URL: <https://www.nature.com/articles/nmat2380>.
- [53] Martin Linck et al. “Chromatic Aberration Correction for Atomic Resolution TEM Imaging from 20 to 80 kV”. In: *Physical Review Letters* 117 (7 Aug. 2016), p. 076101. ISSN: 10797114. DOI: 10.1103/PHYSREVLETT.117.076101/FIGURES/5/MEDIUM. URL: <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.117.076101>.
- [54] K. Ishizuka and N. Uyeda. “A new theoretical and practical approach to the multislice method”. In: *urn:issn:0567-7394* 33 (5 Sept. 1977), pp. 740–749. ISSN: 0567-7394. DOI: 10.1107/S0567739477001879. URL: <http://scripts.iucr.org/cgi-bin/paper?a14220>.
- [55] P. D. Nellist and S. J. Pennycook. “Incoherent imaging using dynamically scattered coherent electrons”. In: *Ultramicroscopy* 78 (1-4 June 1999), pp. 111–124. ISSN: 03043991. DOI: 10.1016/S0304-3991(99)00017-0.
- [56] J. Barthel and A. Thust. “Aberration measurement in HRTEM: implementation and diagnostic use of numerical procedures for the highly precise recognition of diffractogram patterns”. In: *Ultramicroscopy* 111 (1 Dec. 2010), pp. 27–46. ISSN: 1879-2723. DOI: 10.1016/J.ULTRAMIC.2010.09.007. URL: <https://pubmed.ncbi.nlm.nih.gov/21111264/>.
- [57] Adrian I. Ruskin, Zhiheng Yu, and Nikolaus Grigorieff. “Quantitative characterization of electron detectors for transmission electron microscopy”. In: *Journal of structural biology* 184 (3 Dec. 2013), pp. 385–393. ISSN: 1095-8657. DOI: 10.1016/J.JSB.2013.10.016. URL: <https://pubmed.ncbi.nlm.nih.gov/24189638/>.

- [58] A. R. Faruqi and G. McMullan. “Electronic detectors for electron microscopy”. In: *Quarterly reviews of biophysics* 44 (3 Aug. 2011), pp. 357–390. ISSN: 1469-8994. DOI: 10.1017/S0033583511000035. URL: <https://pubmed.ncbi.nlm.nih.gov/21524337/>.
- [59] Xiahua Sang and James M. LeBeau. “Revolving scanning transmission electron microscopy: Correcting sample drift distortion without prior knowledge”. In: *Ultramicroscopy* 138 (Mar. 2014), pp. 28–35. ISSN: 0304-3991. DOI: 10.1016/J.ULTRAMIC.2013.12.004.
- [60] Kazuo Ishizuka. “Contrast transfer of crystal images in TEM”. In: *Ultramicroscopy* 5 (1-3 1980), pp. 55–65. ISSN: 0304-3991. URL: https://www.academia.edu/18916287/Contrast_transfer_of_crystal_images_in_TEM.
- [61] Naomi Falsini et al. “Halide Perovskites Films for Ionizing Radiation Detection: An Overview of Novel Solid-State Devices”. In: *Sensors 2023, Vol. 23, Page 4930* 23 (10 May 2023), p. 4930. ISSN: 1424-8220. DOI: 10.3390/S23104930. URL: <https://www.mdpi.com/1424-8220/23/10/4930/htm%20https://www.mdpi.com/1424-8220/23/10/4930>.
- [62] Lewys Jones et al. “Smart Align—a new tool for robust non-rigid registration of scanning microscope data”. In: *Advanced Structural and Chemical Imaging* 1 (1 Dec. 2015), pp. 1–16. ISSN: 21980926. DOI: 10.1186/S40679-015-0008-4/FIGURES/11. URL: <https://ascimaging.springeropen.com/articles/10.1186/s40679-015-0008-4>.
- [63] Jérôme Boulanger, Charles Kervrann, and Patrick Bouthemy. “Space-time adaptation for patch-based image sequence restoration”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29 (6 June 2007), pp. 1096–1102. ISSN: 01628828. DOI: 10.1109/TPAMI.2007.1064. URL: https://www.researchgate.net/publication/6397823_Space-Time_Adaptation_for_Patch-Based_Image_Sequence_Restoration.
- [64] Özgün Çiçek et al. “3D U-net: Learning dense volumetric segmentation from sparse annotation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9901 LNCS (2016), pp. 424–432. ISSN: 16113349. DOI: 10.1007/978-3-319-46723-8_49/TABLES/3. URL: https://link.springer.com/chapter/10.1007/978-3-319-46723-8_49.
- [65] Lovedeep Gondara. “Medical image denoising using convolutional denoising autoencoders”. In: () .
- [66] Himmat S. Kushwaha et al. “De-noising Filters for TEM (Transmission Electron Microscopy) Image of Nanomaterials”. In: *2012 Second International Conference on Advanced Computing Communication Technologies*. 2012, pp. 276–281. DOI: 10.1109/ACCT.2012.41.
- [67] Michael Elad and Michal Aharon. “Image denoising via sparse and redundant representations over learned dictionaries”. In: *IEEE Transactions on Image Processing* 15 (12 Dec. 2006), pp. 3736–3745. ISSN: 10577149. DOI: 10.1109/TIP.2006.881969.
- [68] Rakibuzzaman Mahmud. “Neural Radiance Fields for TEM Tomography Images”. Unpublished master’s thesis. MA thesis. Kiel, Germany: University of Kiel, 2024, pp. 40–44.

- [69] Zirui Wang et al. “NeRF–: Neural Radiance Fields Without Known Camera Parameters”. In: (Feb. 2021). URL: <https://arxiv.org/abs/2102.07064v4>.
- [70] Ben Mildenhall et al. “NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images”. In: (Nov. 2021). URL: <http://arxiv.org/abs/2111.13679>.
- [71] Naama Pearl, Tali Treibitz, and Simon Korman. “NAN: Noise-Aware NeRFs for Burst-Denoising”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2022-June (Apr. 2022), pp. 12662–12671. ISSN: 10636919. DOI: 10.1109/CVPR52688.2022.01234. URL: <https://arxiv.org/abs/2204.04668v2>.

List of Figures

1.1	Comparative Analysis of Original TEM Images and Enhanced Outputs.	3
2.1	Schematic diagram of Transmission Electron Microscope	5
2.2	TEM images of a mesoporous material(MCM-41) at different magnifications [12]	6
2.3	DeepSDF’s 3D bunny shape representation [13]	7
2.4	Multi-view to Novel view synthesis[20]	8
2.5	Schematic Representation of Camera Parameters	9
2.6	COLMAP camera position extracting from set of 2D images	11
2.7	Conceptual Illustration of Neural Radiance Fields (NeRF) [7]	12
2.8	ESRGAN Architecture [37]	14
3.1	Comparative Reconstruction of a 10 nm Diameter Gold Particle: (A) Using Linear Backprojection in IMOD Software, showing deviation from spherical shape. (B) Using Curvilinear Backprojection in TxBR, accurately representing a spherical particle [41].	15
3.2	Atomic Resolution 3D Imaging of Platinum Nanoparticle: (A) STEM image showing flat twin boundaries. (B) AET reconstruction reveals atomic steps at twin boundaries, with a grain boundary and stacking fault [50]. .	17
3.3	First row - Original real images, Second row - Noisier versions, Third row - Images after denoising with CNN DAE [64].	19
3.4	TEM Imaging and Denoising of Cadmium Sulfide Nanoparticles: A. Original Image, B. Image with Gaussian Noise, C. Denoised with Average Filter, D. Denoised with Median Filter, E. Denoised with Weiner Filter [66].	20
4.1	Single image from TEM Dataset 1	21
4.2	TEM Dataset 1	22
4.3	Single image from TEM Dataset 2	23
4.4	TEM Dataset 2	23
4.5	Single image from TEM Dataset 3	24

4.6	TEM Dataset 3	24
4.7	Single image from TEM Dataset 4	25
4.8	TEM Dataset 4	25
4.9	Single image from STEM Dataset 1	26
4.10	STEM Dataset 1	26
4.11	Single image from STEM Dataset 2	27
4.12	STEM Dataset 2	27
4.13	Single image from Synthetic Dataset [68]	28
4.14	Synthetic Dataset [68]	28
5.1	NeRF Overview: Image creation involves sampling 5D coordinates along camera rays, using a multilayer perceptron for color and density, and employing volume rendering for image compilation [7].	29
5.2	Optimization of Camera Pose in NeRF-MM. The purple path represents the camera trajectory derived from COLMAP, while the blue path illustrates the camera trajectory as determined by NeRF-MM. [69]	30
5.3	Comprehensive overview of our framework: Integrating NeRF and advanced Denoising Techniques, ESRGAN for Enhanced TEM Image Analysis	35
7.1	Sample Image from the Clock Dataset.	42
7.2	Dataset of 2D Images from Various Camera Positions.	42
7.3	Still Frames from Video Generated by NeRFStudio Using 2D Images.	43
7.4	COLMAP Error During Training.	43
7.5	Dataset 3 with 200 epochs.	46
7.6	Dataset 3 with 10,000 epochs.	47
7.7	Progression of PSNR values over 10,000 epochs for Dataset 3	47
7.8	Progression of camera position optimization from 0 to 10,000 epochs in Dataset 3	48
7.9	Dataset 3 with Tanh Activation function (200 epochs).	49
7.10	Dataset 3 with Tanh Activation function (10,000 epochs).	50
7.11	Dataset 3 with Sigmoid Activation function (200 epochs).	50
7.12	Dataset 3 with Leaky ReLU Activation function (200 epochs).	51
7.13	Dataset 3 with Leaky ReLU Activation function (10,000 epochs).	51
7.14	Progression of PSNR values over 10,000 epochs after adjusting hyperparameters for Dataset 3	52
7.15	Dataset 3 with 8 Layer Architecture.	52
7.16	Remarkable Enhancements in Image Quality Through NeRF-Dark and NAN-NeRF Integration.	53
7.17	Image Postprocessed with Bilateral Filtering.	54
7.18	Image Postprocessed with Gaussian Blur.	54
7.19	Image Postprocessed with Median Filtering.	55
7.20	Image Postprocessed with Non-Local Means Denoising.	55
7.21	Image Postprocessed with Wavelet Denoising.	56
7.22	Image Postprocessed with ESRGAN.	57
7.23	Experimental Results for TEM Dataset 1.	58
7.24	Experimental Results for TEM Dataset 2	59
7.25	Experimental Results for TEM Dataset 4	60
7.26	Experimental Results for STEM Dataset 1	61
7.27	Experimental Results for STEM Dataset 2	62

7.28 Experimental Results for Synthetic Dataset	63
---	----

List of Tables

4.1 Table Summary: Characteristics of TEM Dataset 1	22
4.2 Table Summary: Characteristics of TEM Dataset 2	23
4.3 Table Summary: Characteristics of TEM Dataset 3	24
4.4 Table Summary: Characteristics of TEM Dataset 4	25
4.5 Table Summary: Characteristics of STEM Dataset 1	26
4.6 Table Summary: Characteristics of STEM Dataset 2	27
4.7 Table Summary: Characteristics of Synthetic Dataset	28
7.1 Summary of Nerf Models Experimented with TEM Images.	44
7.2 Summary of COLMAP Alternatives Experimented with TEM Images. . .	45
7.3 Comparative Analysis of Image Denoising Techniques Across TEM and STEM Datasets Measured by Peak Signal-to-Noise Ratio (PSNR) . .	64
7.4 Evaluation of Image Denoising Efficacy on Various Datasets Using the Structural Similarity Index Measure (SSIM)	64