Mithun Das

[Company name]  [Company address]

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

**Thesis Structure:**

1. Abstract
2. Introduction Skill and crossbones
3. Background Skill and crossbones
4. Related Works Skill and crossbones
5. Preliminary
6. Method
7. Experiments
8. Results
9. Datasets
10. Comparisons
    1. Limitations and Future work
11. Discussion and Conclusions
12. Acknowledgments
13. References

Additional Baseline method details

**Introduction**

Electron tomography, often known as ET, is now the method of choice for determining the three-dimensional ultrastructure of organelles and cells at nanoscale resolutions(Neumüller 2018). 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 +/- 60° to 80°), and then computationally recombining the images(Chreifi et al. 2019). However, the low electron doses that are applicable to biological samples (which are typically less than 100 e−/Å2) result in extremely poor signal-to-noise ratios (SNR) in the tomograms that are produced as a result (Frangakis 2021). The leading causes of the noise are the stochastic character of the events involving electron scattering and the constraints imposed by electron detection (Joy 2008). 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(Ellis and Cohen-Gould 1927). This leads to very tiny structural features, which are essential for interpreting the sophisticated cellular processes and chemical interactions, to get confused and distorted. Therefore, reducing the amount of noise in tomograms is a critical step in the preprocessing phase that comes before extracting information that has biological significance.

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(Fernandez 2009; Frangakis and Hegerl 2001). More enhanced regularization approaches such as total variation (TV) minimization (Goris et al. 2012)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 DnCNN (K. Zhang et al., 2016) 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 (González-Ruiz and Fernández 2023). Other methods involve training on simulated data, which might not translate well to tomograms taken from actual data(Martinez-Sanchez et al. 2023). In addition, the majority of currently available deep learning algorithms lack interpretability into the newly acquired features and have difficulty denoising non-uniform noise distributions, which are frequently seen in practice.

Recently, neural radiance fields (NeRF) (Mildenhall et al. 2020)have demonstrated unprecedented ability to synthesize photorealistic novel views of complex 3D scenes using a continuous volumetric Representations use a typical multilayer perceptron to learn a 5D radiance field in which each 3D coordinates (X,Y,Z) is correlated to an emission color (R,G,B) and volume density (MLP). The use of NeRF to noisy TEM tilt series data is a considerable difficulty, despite the fact that it has demonstrated impressive results for the new view synthesis of uncontaminated images. 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, I 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. Additionally, we study various training methods to properly condition the model on the noise characteristics of actual TEM data. We hope to overcome the constraints of regular NeRF when it is used directly out of the box to electron tomography volumes by customizing it to accommodate noisy inputs in this manner. Our noise aware NeRF model has the potential to open up new avenues for the high-fidelity 3D denoising and analysis of ET reconstructions.

**Background Information**

**Neural radiance field:** A fully connected neural network called a neural radiance field (NeRF) may provide inventive renderings of intricate 3D scenes from a sparse collection of 2D photos. It has been trained to replicate input views of a scene using a rendering loss. It functions by interpolating between input photos of a scene to create a single rendered scene. NeRF is a very efficient method for creating images from synthetic data (Mildenhall et al. 2020).

To render new views, a NeRF (Neural Radiance Field (NeRF): A Gentle Introduction n.d.)network is trained to directly map from viewing direction and spatial location (5D input) to opacity and color (4D output). NeRF is a computationally demanding technique, and it might take hours or even days to process complex scenes. New algorithms, nevertheless, are readily available and significantly boost performance.

**Camera Parameters:**

**Denoising:** (Mildenhall et al. 2021; Pearl, Treibitz, and Korman 2022)

**Noise Modeling:** (Kniesel et al. n.d.)

**Implicit reconstruction:** (Kniesel et al. n.d.)

**NeRF Math:** (Bian et al. 2022)

**Joint Optimization of Poses and NeRF:** (Bian et al. 2022)

**View synthesis and image-based rendering:** (Mildenhall et al. 2020)

**TEM**

A beam of electrons is used in transmission electron microscopy (TEM), which generates images of specimens with a resolution that is far higher than that of optical microscopes(Egerton, Li, and Malac 2004; Tang and Yang 2017). 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) (Gault et al. 2008). Electromagnetic lenses concentrate the electron beam such that it is directed toward the extremely thin sample. Electrons, when they go through the sample, 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(Tang and Yang 2017)

Additional lenses concentrate the transmitted electrons so that they may be captured as an image on a detector or camera(Gault et al. 2008). The transmission electron microscope (TEM) may provide magnifications of up to 2 million times (Gault et al. 2008), which enables the viewing of structures and details on a scale as tiny as a nanometer or an angstrom (Egerton, Li, and Malac 2004). Because of this, it is an extremely useful instrument for study in the fields of materials science, cell biology, molecular structure analysis, and semiconductors(Egerton, Li, and Malac 2004).

Imaging mode and diffraction mode are the major modes of operation for the transmission electron microscope (TEM) (Adrian et al. 1984). 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(Adrian et al. 1984).

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)(Adrian et al. 1984). Staining with substantial amounts of heavy metal salts is required for biological and polymer materials to produce contrast(Adrian et al. 1984). Imaging of hydrated materials is possible because to specialized methods such as cryo-TEM, which vitrifies the samples(Adrian et al. 1984).There is a possibility that radiation will destroy sensitive specimens, which is one of the TEM's limitations(Egerton, Li, and Malac 2004). Imaging of living biological samples is likewise not possible due to the vacuum environment(Egerton, Li, and Malac 2004). Nevertheless, transmission electron microscopy continues to be an essential instrument for high-resolution structural characterization in both the physical and biological sciences(Egerton, Li, and Malac 2004).

In this study, transmission electron microscopy (TEM) was used to examine Janus-like particles that were created from block copolymers. Transmission electron microscopy (TEM) gives the resolution and contrast necessary to clearly examine the nanostructure morphology and surface topology of the Janus particles(Walther and Müller 2013) (Tang and Yang 2017).

**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(Mildenhall et al. 2020) . NeRF represents a scene as a continuous 5D radiance field (3D position + 2D viewing direction) using a multilayer perceptron (MLP).

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, the 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 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.

We investigate training NeRF models on real noisy TEM data in a self-supervised manner to learn specialized priors for electron microscopy. The coordinate-based modeling may also better capture 3D contextual relationships compared to 2D/3D CNNs. This could enable high fidelity 3D reconstruction from extremely sparse and noisy TEM tilt series.

**Novel view Synthesis**

The term **View synthesis** refers to the process of generating new photographic viewpoints of a subject from one or more input photographs. This may be done with either a single image or many images. This allows to create unique synthetic viewpoints using only a little 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(Xia and Xue n.d.).

For view synthesis, a wide range of methods have been utilized. 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 (Seitz et al. 2006; Xia and Xue n.d.). Then, this model may be displayed from any perspectives. Image-based rendering distorts and interpolates pixels depending on the original inputs to infer new viewpoints (Chen and Williams 2023). These methods concentrate on identifying correspondences between different pictures.

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 learnt (Tewari et al. n.d.). Neural radiance fields (NeRF) are 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(Mildenhall et al. 2020). 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 (Fang et al. n.d.).

**Neural 3D shape representations:**

Major innovations in deep learning have made it possible for neural networks to automatically represent and display 3D forms. This was previously not possible. Effectively mapping 3D coordinates to shape attributes such as occupancy, signed distance, or radiance is something that neural implicit models do as opposed to explicit mesh or voxel representations (Park et al. 2019).

Early experiments were on discovering verified distance functions as a means of representing 3D surfaces for synthetic datasets.(Mescheder et al. 2018; Wu et al. 2015). In later techniques, an attempt was made to reduce the need for ground truth 3D surveillance by defining distinguishable rendering targets that could be improved with just 2D images(Sitzmann et al., 2020; Wu et al., 2015). These techniques generate a feature vector at every three-dimensional place, which is then represented as an RGB color. They have, however, been restricted to simple, smooth forms up until this point.

Neural radiance fields (NeRF), which were developed very recently, have lately shown considerable gains in modeling complicated real-world scene shape and view-dependent presentation(Mildenhall et al. 2020). Using multilayer perceptron’s, NeRF describes the radiance of the scene as well as the volume density of the scene as continuous 5D functions (3D position + 2D view direction)(Mildenhall et al. 2020). The main benefits, when compared to earlier neuronal representations of 3D space, are as follows:

1. MLPs that are based on coordinates can more accurately capture local spatial connections.
2. The production of high-quality novel perspectives is made possible by continuous scene representation.
3. The view-dependent effects such as highlights are encoded by the 5D radiance field.

In this study, we propose utilizing the capabilities of NeRF to represent and denoise 3D volumes that have been reconstructed from TEM tilt series. This will be accomplished by leveraging the strengths of NeRF. Teaching the MLP to successfully encode 3D structural priors that are crucial for biomolecular imaging might be accomplished by teaching the MLP to map noisy TEM inputs to clearer targets. It's possible that the coordinate-based volumetric modeling will be able to pick up on important local context that other 3D denoising networks overlook. The interpretability of structural features from TEM tomograms might be greatly improved because of this.

**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(Hartley and Zisserman 2000). 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.

**Intrinsic** parameters are those that are unique to a camera and are not affected by the scene:

* **Focal length** - 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 (Heikkila and Silven 1997). When dealing with non-square pixels, the x and y axes may have unique values.
* **Principal point** - The coordinates of the image's center on the plane of the sensor. enables the use of lenses that are not centered (Heikkila and Silven 1997). It is dependent on how the lens is aligned.
* **Skew coefficient** - A rotation of the axis between the pixel grid and the sensor that considers non-rectangular pixel shapes (Heikkila and Silven 1997). 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 (Heikkila and Silven 1997).

The **extrinsic** parameters are determined by the position of the camera in relation to the world:

* **Rotation matrix** - Orientation of the camera's coordinate frame in 3 dimensions with respect to a fixed world frame (Z. Zhang 2000). A representation of a sequence of rotations based on the Euler angle.
* **Translation vector** - The 3D origin point of the camera center in the space for world coordinates (Z. Zhang 2000).

Intrinsic and extrinsic parameters collaborate to completely define the camera projection matrix (Lepetit, Moreno-Noguer, and Fua 2009), which is responsible for mapping 3D world points into 2D picture coordinates. To perform computer vision tasks such as 3D reconstruction, posture estimation, and new view synthesis, it is necessary to have an accurate assessment of these parameters.

Applications such as augmented reality (Z. Zhang et al. 1995), autonomous navigation (Z. Zhang et al. 1995), and computational photography (Z. Zhang et al. 1995) rely heavily on accurate camera calibration to perform optimally. Adapting camera models to new modalities such as light field imaging is still a research challenge that is being actively worked on.

**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 (Schönberger and Frahm n.d.). 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. Based on local gradients, SIFT is frequently used to locate scale- and rotation-invariant key points (Lowe 1999). Each key point has a high-dimensional descriptor vector that is insensitive to noise, perspective, and illumination (Lowe 1999).

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 (Lowe 1999). Outlier matches that are inconsistent with a single 3D point are eliminated by geometric verification using RANSAC (Lowe 2004).

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

In an incremental SfM method, the registered 2D-2D matches create initial sparse 3D point clouds (Fischler and Bolles 1981). An initial point cloud is plotted using an initial image pair. Which views to update next are efficiently chosen by robust visibility constraints (Fischler and Bolles 1981). With points recursively mapped from fresh views (Snavely et al. n.d.), camera poses are predicted using a Straight Linear Transform within a RANSAC cycle.

* **Global SfM Optimization**

Utilizing bundle adjustment, the progressive reconstruction is globally improved to simultaneously improve camera 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 (Heinly et al. n.d.). 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 normalized cross correlation between distorted picture patches, dense correspondence is created each view (Triggs et al. n.d.). Accuracy is improved by regularization using filtering such Gaussian smoothing (Galliani et al. n.d.). 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 (Facciolo, Franchis, and Meinhardt 2015). 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 (Kazhdan, Hopkins University, and Hoppe n.d.), may be utilized to improve details even further. Realism and color are added when texturing with the use of input images.

The fact that the source code for COLMAP is freely available has made it possible for several different extensions to be developed, such as semantic 3D modeling, augmented reality, and fusion with other modalities (Zhou and Koltun n.d.). The pipeline is a shining example of some of the most effective methods for visual reconstruction, including robust matching, global optimization, and volumetric fusion.

**Related Work**

**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(Lin et al. 2020). This allows the volume to be recovered from the 2D projections. Traditional algorithms, such as back projection, are incapable of fully addressing the ill-posed Ness of the problem because of its structural dimension.

Regularized iterative reconstruction approaches have been created as a means of overcoming restrictions that are unique to analytical methods (Widmer et al. 2013). 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(Basu et al. n.d.; Widmer et al. 2013), 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 (Sorzano et al. 2004). 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 (K. Zhang, Zuo, Gu, et al. 2017). 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(Fernandez 2012), 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(Fernandez 2012). 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 (Lawrence et al. 2006). In recent years, learning-based algorithms, such as dictionary learning (García-Nafría and Tate 2021) and deep convolutional networks (K. Zhang, Zuo, and Zhang 2017), 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 (Dong, Fu, and He n.d.) and neural radiance/volumetric fields (Moawad et al. 2020), 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 learnt regularization techniques, could provide a more resilient solution.

In this work, I will present the latest technology called Neural Radiant Field (Wang et al. 2021). It is designed to effectively create 3D images from 2D image input. This cutting-edge technology utilizes advanced algorithms and neural networks to enhance the quality of the images by reducing noise and improving clarity. With its ability to handle complex and intricate image data, Neural Radiant Field opens new possibilities in various fields, including materials science, biology, and nanotechnology.

**Atomic Resolution**

Atomic-resolution transmission electron microscopy (TEM) has emerged as a revolutionary technique for materials characterization by directly imaging individual atoms (Kawahara et al. 2022). Modern TEMs use aberration correctors and monochromators that eliminate chromatic blurring and lens flaws to achieve sub-angstrom resolution(Krivanek et al. 2010). Advanced detectors and highly stable devices have created new opportunities for quantitative analysis (Wastl, Weymouth, and Giessibl 2013).

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 Z2 (Miao, Ercius, and Billinge 2016). Elemental distributions are mapped by atomically resolved energy dispersive X-ray (EDX) spectroscopy (Stevens et al. 2014). High resolution TEM can image light elements and even depict atom columns (Jones and Nellist 2013). For high precision data, scanning TEM (STEM) raster scans a focused probe (MacLaren and Ramasse 2014).

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 (Muller 2009). The fundamentally probabilistic electron-sample interactions thus mask important atomic organization details with quantum noise (Linck et al. 2016). Robust denoising techniques designed for TEM are necessary to obtain quantitative data.

Through statistical post-processing, techniques like multi-frame averaging (MacLaren and Ramasse 2014) and principal component analysis enhance signal-to-noise. Sparse regularization techniques take advantage of structural redundancy to reduce noise (Jones and Nellist 2013). Compact representations for denoising are discovered through dictionary learning (Miao, Ercius, and Billinge 2016). Convolutional neural networks have most recently demonstrated the ability to learn potent priors from atomistic image simulations(Miao, Ercius, and Billinge 2016; Stevens et al. 2014).

In general, the developments that have been made in aberration corrected TEM have effectively actualized single-atom sensitivity and precision(Linck et al. 2016). 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.

**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’ (K. Ishizuka and Uyeda 1977; Nellist and Pennycook 1999). 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 (Barthel and Thust 2010). On CCD cameras, dark current shot noise and readout noise are present (McMullan et al. 2007). Scintillator-photomultiplier detectors show signal-dependent Poisson noise characteristics (Ruskin, Yu, and Grigorieff 2013). 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 (Faruqi and McMullan 2011). Monitoring beam current during acquisition allows normalization to reduce this noise (Faruqi and McMullan 2011; Sang and LeBeau 2014). 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 (Kazuo Ishizuka 1980).

**Multiresolution Modeling**

Noise also exhibits signal-dependency and non-stationarity over spatial frequencies (Falsini et al. 2023; Jones et al. 2015). Variance stabilization using multiresolution transforms has been proposed to normalize noise over different scales (Boulanger, Kervrann, and Bouthemy 2007; Foi et al. n.d.).

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.

**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 (Sim, Teh, and Nia 2016). 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 (Sim, Teh, and Nia 2016).

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 (Elad and Aharon 2006) 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 (K. Zhang, Zuo, Chen, et al. 2017). Convolutional networks trained as discriminators between clean and noisy image patches can implicitly model complex image priors. Recurrent inference further boosts quality(K. Zhang, Zuo, Chen, et al. 2017) . Autoencoder architectures directly optimize reconstruction loss (K. Zhang, Zuo, Chen, et al. 2017). 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.

**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 (Çiçek et al. 2016).

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(Gondara n.d.). 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 (Moawad et al. 2020; K. Zhang, Zuo, Gu, et al. 2017). 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.

**View Synthesis and Image-based rendering**

View synthesis refers to rendering novel views of a scene from limited input images. Traditional image-based rendering (IBR) approaches extrapolate new perspectives by warping and interpolating input images based on estimated dense correspondence (Chen and Williams 2023). These methods rely on accurate multi-view stereo matching.

More recent learning-based strategies use neural networks to implicitly infer scene structure and appearance for high quality view synthesis (McMillan and Bishop 1995). Neural radiance fields (NeRF) have shown promising results by encoding scenes as continuous 5D radiance fields with MLPs (Mildenhall et al. 2020). The coordinate-based volumetric scene representation better captures smooth variations compared to voxels.

Applying view synthesis techniques to TEM could significantly improve 3D volume reconstruction from sparse 2D tilt series projections. The continuous modeling of NeRF may effectively represent structural priors and noise characteristics to enable generating new perspectives. However, scaling to large TEM datasets remains challenging.

Among view synthesis techniques, NeRF provides an appealing approach for TEM reconstruction by leveraging MLPs to model local relationships in a coordinate-based volumetric domain. In this work, we propose adapting NeRF to learn mappings between real noisy TEM tilt inputs and target volumes in a self-supervised manner. The noise-aware coordinate modeling could address key challenges in TEM imaging.

**Method**

**Datasets:**

**Experiments**

**Results**

**Limitation and Future work**

**Discussion and Conclusion**