Attention-Guided Dual Deep Neural Networks for Robust Blind Denoising of Medical X-ray Images

By

Fikir Awoke

Submitted to the School of Information Technology and Engineering

In partial fulfillment of the requirements for the degree of

Master of Science in Artificial Intelligence

Supervised by: Dr. Adane Letta.

College of Technology and Built Environment

Addis Ababa University

Addis Ababa, Ethiopia

October 2025

Addis Ababa University

College of Technology and Built Environment

School of Information Technology and Engineering

This is to certify that the thesis prepared by Fikir Awoke, entitled **"Attention-Guided Dual Deep Neural Networks for Robust Blind Denoising of Medical X-ray Images"**, Submitted to the School of Information Technology and Engineering In partial fulfillment of the requirements for the degree of Master of Science in Artificial Intelligence.

Approved by Board of Examiners

Name	Signature	Date
Dr. Sileshi Demesie		
(Chairman)		
Dr. Adane Letta		
(Advisor)		
Dr. Mesfin Abebe		
(External Examiner)		
Dr. Fantahun Bogale		
(Internal Examiner)		

DECLARATION

I declare that this thesis entitled "Attention-Guided Dual Deep Neural Networks for Robust Blind Denoising of Medical X-ray Images" is my original work and has not been submitted to any other institution for the award of any degree or diploma. All sources of materials used in the preparation of this thesis have been properly acknowledged.

Fikir Awoke		Oct, 2025
Name of Student	**Signature**	**Submission Date**

ABSTRACT

Medical image denoising is the process of reducing unwanted noise from medical images like X-rays, MRIs, and CT scans to improve diagnostic accuracy and clarity. Accurate diagnosis in medical imaging, particularly in radiology, heavily depends on the clarity and quality of visual data. In X-ray imaging, the presence of noise can obscure critical anatomical details, potentially leading to misinterpretation or delayed diagnosis. While previous methods such as BM3D, DnCNN, and domain-specific architectures like X-ReCNN and X-BDCNN have shown significant performance on denoising tasks, they often rely on predefined noise assumptions or lack mechanisms to adaptively attend to varying noise patterns in different image regions. To address these limitations, we propose an attention-guided dual-path deep neural network designed for blind image denoising of real-world medical X-ray images. Unlike standard attention modules, we integrate spatial and channel noise-aware attention mechanisms for medical X-ray denoising, enabling the network to dynamically focus on important features while effectively distinguishing structural details from noise. Our architecture combines U-Net for capturing detailed spatial features and Dilated CNN for extracting broader contextual information. We train our model on the ChestX-ray8 dataset, where it achieves a performance with an SNR of 37.23, PSNR of 42.08, and SSIM of 0.9736. These results demonstrate the model's effectiveness in denoising X-ray images while preserving structural integrity. The main contributions include the introduction of a noise-aware attention mechanism and a multi scale dual-branch architecture for complementary feature learning. Nevertheless, the model has limitations, generalizing to other imaging modalities like MRI or CT.

Key words: Attention mechanism, Blind image, CNN, Medical image denoising,X-ray.

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to God for granting me the strength, perseverance, and encouragement to overcome the challenges I encountered throughout the completion of this thesis. I am sincerely thankful for His boundless love, mercy, and grace.

I would like to extend my profound gratitude to my MSc advisor, Dr. Adane Letta, whose expertise, guidance, and unwavering support have been invaluable to my academic journey. His insightful feedback, constructive advice, and continuous encouragement have greatly enriched my learning experience and contributed significantly to both my personal and professional growth.

I also wish to express my heartfelt thanks to my instructors and examiners, Dr. Beakal Gizachew and Dr. Fantahun Bogale, for their time, advice, and support in progress reviews and mock presentations. Their thoughtful comments and suggestions have been instrumental in improving the quality of this work.

At last, I would like to give my deepest appreciation to my family and friends for their constant encouragement and love which have been my strength the whole time. Their patience and support guided me through this journey and made the completion of this thesis possible.

Contents

List of Abbreviations

Symbol	Meaning
AI	Artificial Intelligence
ANN	Artificial Neural Network
BN	Batch Normalization
CAM	Channel Attention Module
CBDNet	Convolutional Blind Denoising Network
CNN	Convolutional Neural Network
CT Scan	Computed Tomography Scan
DAE	Denoising Autoencoder
DCNN	Dilated Convolutional Neural Network
DL	Deep Learning
DNN	Deep Neural Network
DSRP	Design Science Research Process
DVAE	Denoising Variational Autoencoder
Exp	Experiment
GAN	Generative Adversarial Network
GMMs	Gaussian Mixture Models
GPUs	Graphics Processing Units
GT	Ground Truth
ML	Machine Learning
MLP	Multilayer Perceptron
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NEN	Noise Estimation Network
PET Scan	Positron Emission Tomography Scan
PSNR	Peak Signal-to-Noise Ratio
RED-CNN	Residual Encoder–Decoder Convolutional Network
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SAM	Spatial Attention Module
SCAM	Spatial-Channel Attention Module
SDAE	Stacked Denoising Autoencoder
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Index

Tanh	Hyperbolic Tangent Activation Function
TV	Total Variation
U-Net	U-shaped Network
VAE	Variational Autoencoder
VDN	Variational Denoising Network
X-ray	X-radiation

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Medical image processing plays a vital role in contemporary healthcare by providing clinicians with non-invasive means to visualize internal anatomical structures and physiological functions with remarkable precision [1]. Imaging technologies such as X-ray [2], Magnetic Resonance Imaging (MRI) [3], and Computed Tomography (CT) [4] have become essential tools in clinical diagnosis, treatment planning, and patient monitoring. These modalities allow for the detection of a wide array of pathologies, including tumors, fractures, vascular abnormalities, and organ dysfunctions. Despite their utility, the effectiveness of these imaging modalities is highly dependent on image quality [5, 6]. High-resolution and artifact-free images allow for more accurate interpretation and diagnosis. Conversely, poor-quality images degraded by noise, motion artifacts, low contrast, or hardware limitations can obscure critical clinical details, potentially leading to misdiagnosis or suboptimal treatment strategies. Hence, enhancing image quality is fundamental to improving clinical outcomes and facilitating accurate, data-driven medical decision-making [7].

The earliest efforts to address image degradation in medical imaging were based on spatial domain filtering. Techniques such as mean, median, Gaussian, and adaptive filters [8, 9] aimed to suppress noise by smoothing pixel intensities within local neighborhoods. While computationally simple, these methods typically introduced blurring and edge loss, which is particularly detrimental in medical imaging where fine structures and boundaries are diagnostically significant. To mitigate these drawbacks, transform domain techniques like Fourier and Wavelet Transforms [10] were developed. These methods analyze images in frequency or multi-resolution domains, enabling better separation of noise from mean-

ingful signal structures. Wavelet-based methods, in particular, gained popularity due to their ability to localize image features both in space and scale, offering better preservation of edges. Nevertheless, these approaches still relied on hand-crafted parameters and fixed basis functions, limiting their flexibility across varying noise conditions and image types.

A significant advancement came with the introduction of patch-based methods such as Non-Local Means (NLM) [11]. Unlike local filters, NLM computes pixel intensities based on the similarity of entire image patches, even across non-adjacent regions. This exploited the inherent self-similarity in medical images and preserved structural details more effectively. Building upon this, Block-Matching 3D (BM3D) filtering [12] grouped similar patches into 3D arrays and performed collaborative filtering in the transform domain, achieving state-of-the-art denoising performance for many years. However, both NLM and BM3D are sensitive to parameter tuning and computationally intensive. Although classical denoising methods improved early filtering techniques, they rely on hand-crafted models that poorly adapt to the complex, non-stationary, and modality-specific noise found in real-world medical images [13]. Additionally, these methods often ignore semantic and anatomical priors essential for accurate reconstruction, treating denoising as a generic mathematical problem [14, 15].

The limitations of traditional methods created a fertile ground for the emergence of deep learning-based approaches. Convolutional Neural Networks (CNNs), particularly with architectures tailored for image restoration, revolutionized the field by learning complex mappings from noisy to clean images directly from data [16]. These models are capable of capturing both local features (like textures and edges) and global context (like anatomical structure), resulting in superior performance over classical techniques. Unlike hand-crafted filters, CNNs can generalize across various noise types and imaging conditions once trained appropriately[16, 17].

Recent developments have pushed the boundaries of deep learning even further. Architectures like U-Net [18], which combine encoder-decoder structures with skip connections, are widely used for medical image denoising due to their ability to preserve spatial resolution while enabling deep feature extraction. Transformer-based models [19]and generative

networks [20] have also been explored to capture long-range dependencies and generate more realistic, artifact-free outputs. In addition, self-supervised and unsupervised learning methods have emerged to address the scarcity of clean ground truth data in clinical settings [17].

The evolution of image denoising—from basic spatial filters to deep neural networks reflects a broader trend in medical image analysis toward data-driven and context-aware solutions. Modern denoising models not only improve image quality but also enhance downstream tasks such as segmentation[21], classification[22], and regression[23]. Nevertheless, challenges remain in ensuring robustness, generalizability across imaging devices and patient populations, and interpretability of deep models in clinical practice.

1.2 Motivation

1.2.1 Diagnostic demand

The motivation behind this study stems from the critical need to enhance the denoising process in medical imaging, where the clarity and fidelity of images directly impact diagnostic accuracy and subsequent treatment outcomes [24, 25]. Conventional denoising methods face significant hurdles when confronted with the challenge of blind denoising, where the noise characteristics within medical images are unknown or complex due to environmental factors or sensor uncertainties. These limitations of traditional denoising techniques show the need for a better approach that will overcome the challenges and bring an improved medical imaging. The use of an enhanced denoising technique will highly support medical professionals in providing a clear and accurate image which is important in the decision making process and treatment planning phase to get a positive patient outcome.

1.2.2 Blind Denoising Issue

This study aims to contribute to the major problem in blind denoising mostly seen where there is a lack of paired training data in real world scenarios. Traditional denoising methods often rely on predefined noise models and paired datasets of noisy and clean

images, which are rarely available in practical scenarios, especially in medical imaging [26, 27]. Using powerful deep learning methods such as CNN, this study aims to advance this field [28, 29]. This approach will enhance the quality of medical images as well as contribute to the field of computer vision. At last, this study strives to support the healthcare industry in making accurate diagnostics.

1.2.3 Reducing the Need for High Radiation Doses

High radiation dose in medical imaging impacts the patients health. Due to a high radiation exposure patients become exposed to long term health issues including an increased cancer risk, which needs attention to find a way of getting a better diagnostic image quality with a minimal risk to the patient [30]. This makes it a significant motivation for research due to the associated health risks. Research shows that deep learning based denoising techniques reduce radiation dose in CT scans by preserving the diagnostic quality of the image, showing the safety of low-dose imaging [31]. These approaches give birth to a tangible solution by enhancing the quality of low-dose images while keeping its diagnostic relevance [32]. By focusing on this method, our research can contribute to developing a safer imaging procedure with reduced radiation and make sure the medical images have quality to support the medical professionals' diagnosis process [33]. This aligns with the growing clinical need for techniques that ensure diagnostic robustness while minimizing patient risk, and it supports broader adoption in settings where standard dose calibration may be unavailable or inconsistent.

1.3 Problem Statement

Image denoising is a fundamental problem in low-level vision and plays a crucial role as a preprocessing step in many computer vision tasks, including medical imaging [34]. The objective of image denoising is to recover a noise-free image from its observed noisy counterpart, following the degradation model. However, in numerous scenarios, the noise characteristics in medical images are unknown or unavailable due to factors like radiation dose levels [31], noise [35], motion artifacts [36], poor contrast [37],incomplete data

acquisition, equipment quality and maintenance issues, imaging technique artifacts, environmental conditions, and improper imaging protocols [38]. This can be observed in images captured during endoscopy procedures in challenging conditions, such as a low-light environment, often suffer from unknown noise [39]. These images can exhibit variations in noise patterns and intensities, making noise characterization blind and subsequent denoising more challenging. Blind denoising refers to the process of removing noise from images without prior knowledge or assumptions about the noise distribution or characteristics. This lack of noise information requires denoising methods that can adapt to unknown and varying noise conditions, which is especially critical in medical imaging scenarios.

Therefore, addressing the blind denoising problem becomes meaningful in order to enhance the visual experience of users. While numerous image prior-based methods [40, 41, 42] have been successful in denoising problems, they usually require the estimation of noise levels using different algorithms [43, 44]. One major limitation of image prior-based methods is their dependence on human-defined knowledge to construct the image priors. These priors are often based on assumptions about the typical structure and behavior of images, which might not capture the full range of features present in real-world images [14]. Another challenge with these methods is their typical reliance on internal information from the input image alone [13]. This means that they do not make use of external datasets or additional contextual information that could potentially enhance the denoising process. By only leveraging the internal characteristics of the image, these methods can sometimes miss out on opportunities to improve accuracy and generalization through external references [25, 45]. This internal-only approach, while useful for certain applications, limits the overall potential of the denoising techniques [25].

Apart from blind denoising methods, there is another category of denoising techniques based on discriminative learning, such as CNN [46, 47] that is worth mentioning. These methods, especially those using Convolutional Neural Networks (CNNs) [48, 25, 49], train deep denoising networks with paired training datasets. They implicitly learn the

underlying noise model, which leads to impressive denoising performance. These CNN-based approaches do not rely on human knowledge of image priors but instead leverage the powerful learning capabilities of the network architecture, surpassing the limitations of prior-based methods [40, 41, 42] and achieving improved performance. Specifically, for denoising problems with known noise types like Gaussian noise [50], it is possible to create paired training data and utilize these methods to achieve state-of-the-art results.

However, Another challenge arises when dealing with the practical scenario of blind denoising, where the paired training dataset is unavailable or difficult to obtain in reality [51]. Typically, what we have access to are noisy images with unknown noise characteristics. Moreover, real-world noises tend to be more complex than simple Gaussian noise [50], making it challenging to apply existing models trained for denoising known noises. Consequently, when lacking paired training datasets, these discriminative learning approaches cannot be directly exploited to address blind denoising problems effectively.

In this research, we address the challenging problem of blind image denoising in medical X-ray images, where noise patterns are unknown, spatially varying, and difficult to model due to factors such as sensor imperfections, low-dose imaging, and noise images (Gaussian, salt and papers,poisson ,speackle noises). To tackle this, we integrate a U-Net based Swin Transformer with a Dilated CNN: the U-Net effectively captures local fine-grained anatomical details through its encoder-decoder architecture with skip connections [52], while the Dilated CNN enlarges the receptive field to incorporate broader contextual information without sacrificing resolution [53]. Additionally, the attention mechanisms, encompassing both spatial and channel dimensions, allow the model to focus on diagnostically relevant regions, and the Noise Estimation Network enables adaptive handling of varying noise levels. This combination improves denoising accuracy while preserving critical medical structures, overcoming the limitations of traditional methods that assume known or uniform noise characteristics.

1.4 Research Question

- **RQ1:** How can a Dual CNN-based denoising framework enhanced with attention mechanisms effectively reduce noise in X-ray images while preserving structural and diagnostic details?
- **RQ2:**How does the incorporation of serial channel attention modules in dual CNN architectures enhance noise suppression in X-ray images without compromising anatomical detail preservation?

1.5 Objective

1.5.1 General Objective

The general objective of this study is to develop an integrated approach for enhancing blind medical image restoration by combining noise estimation network, attention network and convolutional Neural Networks.

1.5.2 Specific Objective

- To develop a noise estimation network that identifies and measures the location and intensity of noise in an image.
- To design and implement spatial and channel attention modules that highlight diagnostically important regions and feature representations.
- To integrate attention mechanisms with convolutional neural networks for improved image denoising performance.
- To incorporate multi-scale feature extraction, dilated convolutions, and Swin Transformer modules to enhance both local detail and global contextual understanding.

1.6 Scope

This research thesis focuses on blind denoising of X-ray images, aiming to develop advanced deep learning techniques that can effectively remove unknown and spatially vary-

ing noise without prior knowledge of its distribution. Given the critical role of X-ray imaging in medical diagnostics, improving image clarity while preserving fine anatomical details is essential for accurate analysis. The scope includes designing and implementing noise estimation modules combined with attention-based convolutional neural networks to adaptively identify and suppress diverse noise patterns typically encountered in real-world X-ray scans. The goal is to make our blind denoising model more dependable and effective in producing better quality X-ray images that are captured at different noise levels. As a result these can support medical professionals in their decision making for better healthcare.

1.7 Significance

- **Better Diagnosis:** Blind image denoising improves the quality of images which leads to a correct diagnosis.
- **Ability to Adjust to Different Noise:** These methods can effectively handle different types of noise without prior knowledge.
- **Important Feature Preservation:** Effective denoising methods maintain the important anatomical features and pathologies that guide during medical diagnosis.
- **Reduction of Patient Exposure to Radiation:** Blind denoising minimizes patients exposure to radiation by maintaining a quality image captured in a lower radiation dose.

1.8 Contribution

To tackle the blind medical images, we proposed innovative approaches that leverage advanced techniques:

- Development of a multi-scale spatial and channel attention module that enables CNN-based denoising networks to selectively focus on diagnostically important regions and feature channels across different scales.

- Design of an attention-guided noise estimation network that adaptively predicts spatially varying noise levels and integrates with dual-branch CNNs, allowing effective denoising in X-ray medical images.

- Integration of a U-Net-based Swin Transformer architecture to combine the strong localization ability of U-Net with the global contextual modeling power of Swin Transformers, thereby enhancing both structural preservation and long-range feature representation.

1.9 Thesis structure

The subsequent sections of this document delve into the development of the proposed technique. In Chapter 2, an extensive review of relevant literature and related works is provided, encompassing various aspects such as medical image modalities, common medical image noises, and denoising techniques. Chapter 3 intricately details the methodologies employed, elucidating on the architectures and techniques utilized. In Chapter 4, experimental procedures, findings and limitations are presented. Finally, Chapter 5 puts the conclusion and recommendations for future research endeavors.

Chapter 2

Literature Review

2.1 Background

Medical imaging is a crucial technology in the healthcare industry, involving the creation of visual representations of the interior of a body for clinical analysis and medical intervention. This field encompasses various techniques and processes to capture images of the human body's internal structures, facilitating the diagnosis and treatment of diseases. Medical imaging is critical because it allows healthcare professionals to view the anatomical and physiological conditions of patients non-invasively, leading to more accurate diagnoses, better treatment planning, and improved patient outcomes. Additionally, it helps in monitoring disease development and treatment, making modern medicine essential [54, 55].

Medical imaging comes in various forms, each with its own unique role. X-ray [56], one of the oldest and most common methods, are great for checking bones and spotting fractures. Magnetic Resonance Imaging (MRI) [3] on the other hand, are useful for evaluating the brain, spinal cord, and joints by making use of radio waves and magnets to produce detailed images of soft tissues. Computed Tomography (CT) [4] combines X-rays with computer tech to produce layered images of the body, giving a clear view of organs and other structures. Then there's ultrasound [57], which relies on high-frequency sound waves and is often used in pregnancy checkups or to examine the heart. These medical equipment together guide the diagnosis process of different health defects, plan treatments, and keep patients healthy, strengthening modern healthcare and effective disease management [58, 59].

Medical images include a number of flaws, including noise, low contrast, and low resolution, which might cover important details needed for diagnosis. For instance, in

X-ray and CT images, the presence of noise can make it difficult to distinguish between different tissue types [5, 60]. Low contrast in MRI and ultrasound pictures might make it more difficult to see small details in soft tissues [6]. These problems can be minimized and the overall quality of the images can be increased by using image enhancing techniques [1]. Better visualization and interpretation of the data by radiologists and other medical professionals results in more precise diagnosis and better patient outcomes [61].

Furthermore, image enhancement plays a huge role in the development of computer-aided diagnosis (CAD) systems [62]. These systems rely on high quality images in order to detect tumors, organs, and anomalies. This will in turn increase the accuracy of automated diagnostic tools. In addition to this, image enhancement is crucial for various imaging tasks like for measuring tumor size [63]. Therefore improving medical image quality will maximize the performance of disease detection and monitoring [62, 7].

2.2 Medical Image Modalities

Medical imaging is vital for the clinical environment, by guiding the diagnosis, treatment planning, and patient monitoring process in health care [61, 7]. Various imaging modalities have been introduced in the past years showing the rapid revolution in medical imaging. Each imaging modalities has their unique advantage in identifying patients health status which makes it suitable for different scenarios [64].

2.2.1 X-ray Imaging

X-ray is one of the most commonly used imaging techniques that makes use of an electromagnetic radiation to show the internal structure of our body [56]. When X-rays pass through the body, different tissues absorb varying amounts of radiation based on their density and composition. This differential absorption produces contrast in the resulting image, allowing visualization of anatomical features such as bones, organs, and any foreign objects present [64]. Due to the high absorption rate of dense materials like bone, X-ray imaging excels at producing clear, well-defined outlines of skeletal structures, making it an invaluable tool in clinical diagnosis.

Clinically, X-ray imaging serves numerous diagnostic functions across various medical disciplines. It is indispensable for detecting bone fractures, infections, and tumors, providing rapid assessment of skeletal abnormalities [56]. In addition to musculoskeletal evaluations, X-rays are routinely used in dental imaging to examine tooth structures and identify cavities, offering dentists a non-invasive method for early diagnosis [65]. Chest radiography, a common form of X-ray imaging, plays a critical role in diagnosing pulmonary conditions such as pneumonia, lung cancer, and tuberculosis by delivering high-resolution images of the lungs and surrounding tissues [2]. These broad applications underscore X-ray imaging's versatility and clinical importance.

Despite its benefits, X-ray imaging has clinical and technological issues that affect patient safety and image quality. The primary limitation is that X-ray images often have low contrast and noise, which may blur essential diagnostic features and make interpretation more difficult [5]. To solve these issues with noise and contrast, it requires advanced image denoising and augmentation techniques that keep diagnostic information while making the image clearer.

Another major concern in X-ray imaging is the risk of ionizing radiation exposure. Although the doses used in diagnostic imaging are relatively low, repeated or unnecessary exposure can increase the risk of adverse health effects [66]. This necessitates stringent dose management protocols and optimization of imaging parameters to achieve the best possible image quality at the lowest feasible radiation dose. Advances in technology, such as improved detector sensitivity and adaptive imaging protocols, have contributed to safer X-ray practices, but ongoing vigilance and innovation remain essential.

2.2.2 Magnetic Resonance Imaging (MRI)

MRI utilizes strong magnetic fields and radio waves to create detailed images of the body's internal structures [3]. The technology operates on the principles of nuclear magnetic resonance, where hydrogen atoms in the body's water and fat molecules align with the magnetic field and emit signals that are captured and converted into images [3]. This non-invasive approach provides excellent soft tissue contrast, making MRI particularly

suitable for imaging the brain, spine, and joints. It is highly effective in detecting a range of abnormalities such as tumors, strokes, and degenerative diseases, offering detailed insights into anatomical structures and pathological conditions [67].

Despite its advantages, MRI imaging encounters several challenges. Motion artifacts can distort images due to patient movement during scanning, compromising diagnostic accuracy, especially in studies requiring high resolution and precise anatomical detail [6]. Moreover, MRI suffers from a relatively low signal-to-noise ratio, which can impact image quality and clarity [68]. Techniques such as advanced motion correction algorithms and signal enhancement methods are continually being developed to mitigate these challenges and improve the effectiveness of MRI in clinical practice.

2.2.3 Ultrasound

Ultrasound imaging operates by using high-frequency sound waves generated by a transducer, which are directed into the body to produce detailed images of internal structures [57]. When these sound waves encounter tissues with different densities, they reflect back as echoes to the transducer. These echoes are converted into electrical signals and processed to create real-time images of organs and tissues [57]. This non-invasive, radiation-free technique offers the advantage of immediate visualization, making it especially useful for monitoring dynamic physiological processes within the body.

Ultrasound is widely applied across clinical fields due to its safety and versatility. It is frequently employed in fetal imaging to monitor development and detect anomalies during pregnancy. In cardiac imaging, or echocardiography, ultrasound allows detailed assessment of heart structure and function, facilitating diagnosis of conditions such as valve abnormalities and congenital defects [69]. Additionally, ultrasound is essential in abdominal imaging to examine organs like the liver, kidneys, and gallbladder for abnormalities such as tumors or cysts. Its capability to capture real-time blood flow dynamics makes it valuable for detecting circulatory blockages or irregularities. However, ultrasound imaging faces challenges including speckle noise, which degrades image clarity [70], and limited resolution when imaging deeper tissues, sometimes necessitating supplementary imaging

techniques for comprehensive evaluation [71].

2.2.4 Computed Tomography (CT)

Computed Tomography (CT) imaging utilizes advanced X-ray technology to capture multiple cross-sectional images of the body from various angles. These images are processed using computer algorithms to generate detailed 3D representations of internal organs and structures, providing high spatial resolution and clarity [4]. CT imaging is widely used across medical specialties due to its versatility, enabling healthcare providers to visualize anatomical abnormalities, disease processes, and injuries comprehensively. It is especially valuable for detailed imaging of critical organs such as the brain, lungs, liver, and abdomen, aiding in the diagnosis and management of conditions ranging from tumors to vascular diseases [72]. In oncology, CT plays a pivotal role in cancer detection and staging by accurately assessing tumor size, location, and spread, which guides treatment planning and therapy monitoring [73]. Moreover, CT is indispensable in emergency medicine for rapid diagnosis of traumatic injuries, internal bleeding, and for guiding urgent interventions [74].

Despite its clinical benefits, CT imaging presents challenges that require careful management to ensure patient safety and diagnostic effectiveness. A primary concern is radiation dose exposure, as CT scans involve ionizing radiation that poses potential health risks [66]. To address this, healthcare providers must adopt protocols that minimize radiation while preserving image quality necessary for accurate diagnoses. Additionally, image noise and artifacts can degrade CT image clarity, potentially affecting diagnostic confidence [60]. Ongoing advancements in imaging techniques and scanner technology focus on reducing noise and artifacts, thereby enhancing the reliability and interpretability of CT images in routine clinical practice.

Table 2.1: Comparison of Medical Imaging Modalities

Modality	**Principle**	**Common Uses**	**Advantages**	**Challenges**
X-ray [56, 5, 75]	Electromagnetic radiation	Detecting fractures, infections, tumors	Quick, widely available, good bone imaging	Noise, low contrast, radiation exposure
MRI [3, 68, 67]	Magnetic fields and radio waves	Brain, spine, joint imaging	Excellent soft tissue contrast, no radiation	Motion artifacts, low signal-to-noise ratio
CT [4, 72]	X-ray technology for cross-sectional images	Internal organs, cancer detection	Detailed 3D images, quick, good for emergency	High radiation dose, image noise
Ultrasound [76, 71]	High-frequency sound waves	Fetal, cardiac, abdominal imaging	Real-time imaging, no radiation, portable	Speckle noise, resolution limitations

2.3 Common Medial Image Noises

This section explores the various types of noise commonly encountered in medical imaging and how they can affect the quality and diagnostic utility of images. And basically aims to provide an overview of these noise types, their sources, and the impact they have on image clarity, emphasizing the need for effective denoising techniques to enhance image quality and support accurate medical diagnoses.

2.3.1 Noise

Noise in medical images refers to unwanted random variations in pixel intensity that obscure or distort the true signal of the image, often complicating diagnosis and analysis. It can be introduced during image acquisition due to limitations in hardware, environmental factors, or the nature of the imaging process itself. Each type of noise affects medical images differently depending on the modality, requiring specific denoising techniques to improve clarity and diagnostic accuracy [60].

- **Gaussian Noise:** This type of noise is characterized by a normal distribution of pixel intensity fluctuations and is commonly encountered in imaging systems like

MRI and CT [50]. Gaussian noise arises from electronic disturbances within imaging sensors and amplifiers, such as thermal fluctuations or electrical interference. Its random nature results in smooth distortions across the image, which can blur fine details. Gaussian noise is often addressed with smoothing filters [8] or advanced denoising algorithms that preserve edges while reducing noise.

- **Poisson Noise (Quantum Noise):** Poisson noise, or quantum noise, is specific to imaging modalities that rely on photon or particle detection, such as X-ray, CT, and PET scans [77]. It occurs due to the discrete and probabilistic nature of photon emission and detection, especially in low-dose imaging. Poisson noise leads to uneven image intensities, particularly in areas with low signal levels. Methods like dose optimization and statistical reconstruction techniques are employed to minimize this noise without compromising image quality [78].

- **Salt-and-Pepper Noise:** Salt-and-pepper noise appears as randomly occurring black and white pixels, often caused by data transmission errors or faulty imaging sensors [79]. It is most disruptive in digital images where the noise is concentrated in small areas, creating sharp, abrupt pixel intensity changes. This type of noise can mask critical anatomical details, making it challenging to interpret medical images accurately. Median filtering techniques are commonly used to remove salt-and-pepper noise by preserving image edges while eliminating noisy pixels [79].

- **Speckle Noise:** Speckle noise is particularly prevalent in ultrasound imaging due to the coherent interference of sound waves reflected by small tissue structures [70]. It manifests as granular, grainy distortions that reduce image contrast and make it difficult to distinguish fine tissue details. Speckle noise can hinder accurate diagnosis, especially in cases involving soft tissue differentiation. Denoising techniques such as wavelet transforms and convolutional neural networks are used to reduce speckle noise while maintaining important image features [80].

2.3.2 Low Contrast

Low contrast in medical images refers to the inability to distinguish between regions of varying intensity, making it difficult to identify anatomical structures or pathological changes [37]. Contrast is a critical element in medical imaging as it enhances the visibility of specific tissues, organs, or abnormalities. When contrast is low, subtle differences between adjacent regions become less discernible, which can hinder accurate diagnosis. This issue is particularly significant in modalities like X-ray and CT scans, where the contrast between soft tissues and bones is vital for detecting fractures, tumors, or infections. Low contrast can result from poor imaging conditions, insufficient exposure, or the inherent limitations of the imaging modality, such as in cases of dense tissue regions or low-dose imaging to reduce radiation exposure.

Addressing low contrast in medical images is essential for improving diagnostic accuracy and clinical outcomes. Image enhancement techniques, such as contrast stretching, histogram equalization [81], and advanced algorithms like limited weighted histogram equalization [82], are commonly used to improve visibility. These methods work by redistributing the pixel intensity values in the image, enhancing differences between tissues and making hidden features more apparent. In some cases, contrast agents are administered, particularly in MRI and CT scans, to artificially enhance the contrast between specific structures. However, balancing contrast enhancement without introducing artifacts or noise remains a challenge, requiring careful tuning of both imaging protocols and post-processing methods to ensure the best possible image quality for medical interpretation.

2.3.3 Blurring

Blurring in medical images refers to a loss of sharpness and detail, causing the boundaries between different anatomical structures to become less distinct [83]. This can significantly hinder the ability to accurately diagnose conditions, as important features such as small lesions, fractures, or subtle tissue differences may become obscured. Blurring often occurs due to factors like patient movement during the scan, improper focusing of imaging

equipment, or technical limitations such as low resolution in modalities like ultrasound or X-ray. In addition, blurring can result from certain imaging protocols that prioritize lower radiation exposure or faster scan times, which may lead to reduced image clarity. For instance, in MRI or CT scans, where high precision is critical for detecting abnormalities in soft tissues or organs, even slight blurring can obscure critical diagnostic information [84].

Addressing blurring in medical images involves a combination of improving imaging protocols and employing post-processing techniques. On the acquisition side, measures like using faster imaging techniques, ensuring proper patient positioning, and utilizing motion correction algorithms [36] can reduce the chances of blur. On the post-processing side, deblurring algorithms such as Wiener filters [85] are often used to enhance image clarity by restoring the sharpness of blurred features. However, the challenge lies in enhancing the image without introducing artifacts or noise, which could further complicate diagnosis. Modern advancements in deep learning-based [83, 84] image restoration also offer promising solutions for reducing blurring while preserving important structural information in medical images, contributing to better diagnostic accuracy and treatment planning.

2.4 Spatial Domain-Based Image Denoising Techniques

Spatial domain-based image denoising techniques operate directly on the pixel values of an image [86]. These methods are crucial for reducing noise while maintaining important image features such as edges and textures. Below is a detailed discussion of some fundamental spatial domain techniques used for image denoising:

2.4.1 Averaging Filters

Mean Filter: The mean filter, also known as the box filter, is one of the simplest and most widely used spatial domain techniques. It works by replacing each pixel's value with the average of the pixel values in its neighborhood [87]. This method is effective in reducing Gaussian noise but can blur the edges and details of the image due to its

uniform averaging approach [88].

Gaussian Filter: The Gaussian filter is a type of weighted averaging filter where the weights are determined by a Gaussian function. This filter is effective at smoothing images while preserving edges better than the mean filter [50]. The Gaussian filter's performance depends on the standard deviation of the Gaussian kernel, which controls the extent of smoothing. It is particularly useful for reducing Gaussian noise and maintaining the overall structure of the image [89, 90].

2.4.2 Median Filters

Standard Median Filter: The median filter is a non-linear filter that replaces each pixel with the median value of the pixels in its neighborhood [91]. This technique is highly effective in removing salt-and-pepper noise, where isolated pixels are corrupted. The median filter preserves edges better than averaging filters, as it does not average pixel values but instead selects the median, which is less sensitive to outliers [88].

Adaptive Median Filter: Unlike the standard median filter, the adaptive median filter adjusts the size of the neighborhood based on local noise characteristics [8]. It starts with a small neighborhood and increases its size until the median value stabilizes. This adaptiveness allows the filter to better preserve edges while removing noise, making it effective for images with varying levels of noise.

2.4.3 Bilateral Filters

Bilateral Filter: The bilateral filter is a non-linear filter that reduces noise while preserving edges. It combines spatial distance and intensity difference to weigh the contribution of neighboring pixels [92]. This means that pixels that are both spatially close and similar in intensity to the central pixel have a greater influence on the final output. The bilateral filter is particularly effective in preserving edges and fine details while removing noise. However, it can be computationally intensive due to the need for calculating distances and differences for every pixel [93].

2.4.4 Wiener Filters

Wiener Filter: The Wiener filter is an adaptive filter designed to minimize the mean square error between the estimated clean image and the true image [85]. It adjusts its filtering parameters based on the local image statistics, which allows it to perform better in varying noise conditions. The Wiener filter is effective for images where noise characteristics are known, as it adapts to the local variance of the noise.

2.5 Learning-Based Image Denoising Techniques

In recent years, learning-based techniques have gained significant attention in the field of medical image denoising due to their ability to automatically learn complex noise patterns and adapt to the unique characteristics of medical imaging modalities [94]. Unlike traditional methods that rely on handcrafted features and assumptions about the noise distribution, learning-based approaches leverage large datasets and deep neural networks to model the underlying structure of both noise and signal more effectively.

2.5.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) have become a cornerstone in medical image denoising due to their powerful feature extraction capabilities and adaptability to complex noise patterns. CNNs operate by learning hierarchical representations from raw input data, where the initial layers capture low-level features like edges and textures, while deeper layers model more complex structures. Their ability to preserve fine anatomical details while removing noise has led to widespread use across various imaging modalities, including MRI, CT, and X-rays. [95]

In a foundational study by Zhang et al. [96], the authors introduced a deep CNN-based model known as DnCNN, which demonstrated that residual learning could significantly enhance denoising performance. The DnCNN model works by learning a mapping from noisy images to their corresponding noise components, enabling the network to subtract noise from the input images effectively. This approach was particularly successful in medical image applications such as low-dose CT scans, where reducing noise without

compromising diagnostic accuracy is critical. The authors demonstrated that CNNs outperform traditional methods such as wavelet thresholding and non-local means in both objective metrics and visual quality, especially when handling complex noise types.[96]

Another noteworthy contribution comes from Chen et al. [31], who proposed a Residual Encoder-Decoder CNN (RED-CNN) for low-dose CT image denoising. The RED-CNN uses an encoder-decoder architecture with skip connections, allowing the model to capture both high- and low-frequency information effectively. The study focused on reducing noise in low-dose CT images to minimize radiation exposure while maintaining image quality. Through extensive experiments, the RED-CNN was shown to preserve fine structural details better than non-learning methods like bilateral filtering, and it even surpassed other deep learning-based techniques in terms of peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). The introduction of residual connections helped mitigate vanishing gradients during training, ensuring that deeper networks could learn effectively from the medical image data.

In a more recent work, Yang et al. [97]explored the use of a Noise2Noise CNN framework for unsupervised medical image denoising. This method was designed to address a common limitation in medical imaging: the lack of paired noisy-clean training datasets. The Noise2Noise model was trained using only noisy images, leveraging the insight that the underlying signal remains consistent across different noisy observations. This innovation is particularly relevant for clinical scenarios where acquiring clean, noise-free ground truth images is impractical. The Noise2Noise model showed promising results on MRI and ultrasound images, achieving comparable denoising performance to supervised methods without the need for clean training data. This approach opens new possibilities for practical and scalable denoising solutions in medical imaging where annotated datasets are scarce.

2.5.2 Autoencoders for denoising

Autoencoders have emerged as a powerful unsupervised learning technique for medical image denoising. These models learn a compressed, latent representation of the input data

by encoding noisy images into a lower-dimensional space and then decoding them back to their original resolution, ideally without noise. Autoencoders are particularly useful in medical imaging, where clean, noise-free images are often difficult to obtain. They can be trained using noisy images alone, making them a practical choice for applications such as MRI, CT, and X-ray imaging. Denoising autoencoders (DAEs) have shown significant potential in reducing noise while preserving important diagnostic features, which is critical for maintaining clinical utility.[98]

Vincent et al. [99] pioneered the concept of denoising autoencoders (DAEs) by training the network to reconstruct clean images from their noisy counterparts. The study demonstrated that autoencoders could effectively learn the underlying structure of images while filtering out noise, a capability that has direct applications in medical imaging. In their experiments, DAEs successfully removed noise in handwritten digit images and other natural images, proving the method's generalizability. This foundational work set the stage for applying autoencoders to medical images, where the need for noise reduction without losing critical structural information is paramount. The ability of autoencoders to learn a robust latent representation makes them ideal for handling the complex noise present in medical scans, including MRI and CT.

In a more focused application of DAEs to medical imaging, Gondara [100] applied stacked denoising autoencoders (SDAE) to low-dose CT scans. By stacking multiple layers of autoencoders, the network could learn progressively more abstract representations of the input data, improving its ability to filter out noise. Gondara's model was specifically designed to handle the low-dose CT scenario, where reducing radiation exposure introduces significant noise. The SDAE model outperformed traditional denoising techniques, such as non-local means and bilateral filtering, in both qualitative and quantitative evaluations. It showed that the autoencoder architecture could learn a latent representation that separates noise from essential anatomical details, making it highly effective for clinical use in low-dose CT denoising.

In their 2020 study, Biswas, et a. [101] introduced a novel approach called DVAE (Deep Variational Autoencoder) for denoising retinal fundus images, a critical step in

enhancing the quality of medical images used for the diagnosis of eye diseases. The DVAE model leverages the power of variational autoencoders (VAEs) to learn the underlying noise distribution in retinal images while preserving essential structural details crucial for clinical diagnosis. The authors demonstrated that by utilizing a probabilistic approach, DVAE could not only effectively remove noise but also provide a measure of uncertainty in the denoised images, which is valuable for assessing the reliability of medical image reconstructions. Their results showed superior performance over traditional methods in terms of image quality metrics, such as PSNR and SSIM, making DVAE a promising tool for improving diagnostic accuracy in retinal imaging.

2.5.3 GAN-Based Denoising

Generative Adversarial Networks (GANs) have become a powerful tool for medical image denoising, leveraging their ability to generate high-quality, realistic images from noisy data. GANs consist of two competing networks: a generator that produces denoised images and a discriminator that distinguishes between real (clean) and generated (denoised) images [102]. This adversarial process forces the generator to produce increasingly realistic images, making GANs particularly effective at removing complex noise patterns while preserving fine details, which is critical in medical imaging.

One seminal work by Yang et al. [103] introduced the GAN-based model DenoisingGAN for low-dose CT imaging. The model aims to produce high-quality CT images from low-dose, noisy inputs while minimizing the radiation dose patients are exposed to. DenoisingGAN uses a U-Net as the generator and a patch-based discriminator to capture both local and global features. The adversarial loss, combined with a perceptual loss to maintain image quality, enabled the network to effectively remove noise while preserving anatomical structures critical for diagnosis. Extensive experiments demonstrated that DenoisingGAN significantly outperformed non-learning methods and even traditional CNN-based denoisers, particularly in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), making it a valuable tool in clinical practice.

Another important contribution comes from the work of Wolterink et al. [104], who

proposed a cGAN-based denoising model for low-dose CT reconstruction. Their conditional GAN (cGAN) model conditions both the generator and discriminator on the input noisy CT image to guide the denoising process. The authors focused on the ability of cGANs to generate realistic textures, preserving critical features in low-dose CT images that are often lost with traditional denoising methods. This approach resulted in denoised images with sharper and more accurate anatomical features compared to conventional methods like filtered back projection (FBP) and iterative reconstruction techniques. The study showed that the cGAN model outperformed other deep learning methods in both visual quality and quantitative metrics like PSNR and SSIM, highlighting its effectiveness in producing clinically relevant images.

In their 2024 study, Wang et al [105]. introduced a **Scale-sensitive Generative Adversarial Network (GAN)** specifically designed for low-dose CT image denoising. Recognizing that low-dose CT scans often suffer from varying levels of noise across different scales, the authors developed a scale-sensitive architecture that incorporates multi-scale feature extraction to capture both fine details and larger anatomical structures. This approach enables the network to effectively learn the noise distribution at different levels, improving denoising performance compared to traditional methods. By integrating perceptual loss and adversarial loss, the model not only enhances image quality but also preserves critical diagnostic information. Experimental results demonstrated that their proposed GAN significantly outperformed existing denoising techniques in terms of both objective metrics, such as PSNR and SSIM, and subjective visual quality, highlighting its potential for clinical applications in reducing radiation exposure while maintaining high image fidelity. This work represents a significant advancement in the field of medical imaging, addressing the pressing need for effective denoising solutions in low-dose CT imaging.

2.6 Related Work

Medical image denoising is becoming an increasingly important field in healthcare, as it aims to improve the quality of medical images taken in different equipment with different

quality and environment. These improvements are believed to bring a better diagnosis and treatment planning. To achieve this goal various classical image enhancement techniques have been employed such as, gaussian filtering, histogram equalization, and unsharp masking. However, these techniques often rely on predefined rules and assumptions, limiting their effectiveness in addressing complex imaging challenges in real-world medical scenarios [106]. In recent years, deep learning-based approaches [107, 108, 109, 110, 111] have gained prominence due to their ability to automatically learn and model complex image patterns, leading to significant improvements in image clarity, feature preservation, and noise reduction. However, while these methods effectively enhance noisy images, they face challenges in handling blind medical image denoising, where the noise type and level are unknown, thereby limiting their adaptability to diverse clinical scenarios [112].

Table 2.2: Comparative Analysis on Related Works

Name	**Techniques**	**Noise Robustness**	**Structural Integrity**	**Multi-scale Features**	**Noise Aware Attention**
Wang et al.[113]	CNN	X	✓	X	X
Liu et al.[112]	GAN	✓	✓	X	X
Ponomarenko et al.[114]	CNN	✓	✓	X	X
Chen et al.[115]	GCN	✓	✓	X	X
Ponomarenko et al.[116]	CNN	✓	✓	X	X
Liu et al.[117]	Conditional Diffusion	✓	✓	X	X
Zhou et al.[118]	Self-Supervised N2N	✓	✓	X	X
Li et al.[119]	Real-Time Denoising	✓	✓	✓	X
Sample [120]	Neural Blind Deconv.	X	✓	X	✓

As presented in Table 2.2, which provides a comparative analysis of related works, these terms are essential for assessing the effectiveness and characteristics of various techniques in medical imaging.

- **Techniques**: The machine learning or deep learning method used (e.g., CNN, GAN, GCN).
- **Image Modality**: The type of medical imaging data used (e.g., X-ray, MRI, CT).
- **Adaptability**: The model's ability to generalize across different image modalities.
- **Noise Robustness**: The ability to maintain performance despite noisy or degraded input data.

- **Structural Integrity**: The capability of preserving anatomical or structural details in medical images.
- **Multi-scale Features**: The ability to capture both fine-grained local details and broader contextual information within an image.
- **Noise-Aware Attention Module**: The mechanism to dynamically focus on regions affected by noise while prioritizing important feature channels.

The paper by Wang et al. [113], "X-ray Image Blind Denoising in Hybrid Noise Based on Convolutional Neural Network," proposes a two-network approach consisting of a noise estimation network and a denoising network to handle hybrid noise in X-ray images. The method leverages CNNs and considers the physical formation of medical images through the use of an SSIM loss [121] function, enhancing structural integrity. The combination of Gaussian and Poisson noise improves adaptability to real-world noise variations, increasing robustness. However, the approach is less adaptive due to its reliance on X-ray data and the specific CNN architecture used. Additionally, while the method effectively removes noise, the use of MSE loss may limit its ability to preserve high-frequency textures, potentially leading to oversmoothing in fine-detail regions [122].

The dual-stage MRI image restoration method by Liu et al [112] that integrates blind spot denoising with hybrid attention to enhance MRI image quality. A key innovation is the global mask mapper, which samples blind spot pixels from the denoised image and maps them to the same channel, effectively utilizing all available spatial information while avoiding identity mapping. The incorporation of perceptual loss helps preserve high-frequency details, ensuring better structural integrity [123]. However, during the denoising process, useful image details may sometimes be removed as noise. To counteract this, a Generative Adversarial Network (GAN) with adaptive hybrid attention is introduced in the second stage to restore lost details and refine the image. The use of GAN enhances the model's ability to reconstruct fine structures, making it highly effective for MRI image restoration [124]. However, its reliance on MRI data limits its adaptability to other modalities like X-rays, as it is specifically trained to handle MRI-

specific noise rather than general medical imaging artifacts. Additionally, the dual-stage architecture increases inference time, making it less suitable for real-time applications. Despite this added complexity, the model achieves high-quality restoration, preserving structural integrity and ensuring robust denoising performance.

The paper [114] presents a deep learning-based approach for automatic noise analysis and suppression in dental X-ray images, particularly those acquired using the Morita system. The method addresses spatially correlated noise with an unknown spectrum and region-dependent standard deviation, making it not adaptable to other medical imaging modalities. The first convolutional neural network (CNN) estimates the noise spectrum and level per pixel, generating multi-scale noise standard deviation maps, while the second CNN utilizes these maps for noise suppression, enhancing robustness against complex noise patterns. The model is highly effective for dental X-rays but may not generalize well to other imaging modalities, as it is tailored to the unique noise characteristics of Morita system images. The noise suppression approach improves structural integrity, preserving fine details crucial for dental diagnostics. With a reported PSNR improvement of 2.7 dB over other denoising methods, the model demonstrates superior noise reduction. However, its dual-network architecture increases model complexity, requiring additional computational resources for noise estimation and suppression, which may impact real-time usability. Despite this, the method's effectiveness in maintaining image quality and robustness makes it a valuable tool for dental X-ray enhancement.

Another work by Chen et al [115] demonstrates strong adaptability to real-world LDCT images, addressing the limitations of traditional approaches that rely on synthetic noise, such as Poisson noise, by estimating and generating noise distributions specific to real-world data. However, its adaptability to other image modalities beyond LDCT, such as MRI or X-ray, may be limited due to the model being trained specifically on LDCT noise characteristics. In terms of robustness, the use of Generative Adversarial Networks (GANs) for noise generation and Graph Convolutional networks (GCNs) for capturing non-local patterns significantly improves the model's ability to handle complex noise types inherent in real-world LDCT images, making it more robust to varying noise distri-

butions [20]. The method also excels in maintaining structural integrity, preserving fine image details while reducing noise, which is crucial for medical imaging tasks. However, the inclusion of GANs and GCNs adds model complexity, requiring higher computational resources and making the method less suitable for real-time applications [125]. Despite this, the model's overall performance in denoising and its ability to enhance both quantitative and visual quality of LDCT images make it a promising approach for improving medical image processing.

The X-BDCNN algorithm work by Ponomarenko et al [116], presents a significant advancement in X-ray image denoising, addressing the challenge of noise interference in medical imaging. By designing a noise model rooted in the physical principles of X-ray imaging, the model generates more realistic noisy X-ray images for training, improving the overall realism and accuracy of the denoising process. The architecture includes a noise level estimation subnetwork, which accurately predicts the noise level of the input image, enhancing the performance of the subsequent non-blind denoising subnetwork. The addition of an SSIM loss function further refines the denoising process, ensuring that structural integrity is maintained by preserving fine details in the image [121]. In terms of robustness, the method demonstrates superior performance across a range of noise levels, outperforming existing denoising methods. However, while the model excels in X-ray image denoising, its adaptability to other imaging modalities, such as CT or MRI, may be limited due to the model's focus on X-ray noise characteristics.

The DiffDenoise method by Liu et al. [117] presents a significant advancement in self-supervised medical image denoising, tackling the over-smoothing and detail loss prevalent in prior blind approaches. By integrating conditional diffusion models trained on noisy images with conditioning from a pretrained Blind-Spot Network, followed by a stabilized reverse sampling technique that averages outputs from symmetric noise pairs to generate pseudo-clean images, the framework enables effective blind denoising without clean ground-truth data. This is complemented by fine-tuning a supervised denoiser on the generated pairs, ensuring high-fidelity outputs that preserve critical high-frequency anatomical details essential for diagnosis. In terms of robustness, the method outper-

forms state-of-the-art self-supervised techniques across synthetic and real-world medical datasets, demonstrating strong noise handling and structural preservation across various modalities. **However, while DiffDenoise excels in generating realistic denoised images for blind scenarios, its lack of explicit multi-scale feature extraction and noise-aware attention mechanisms may limit adaptability to highly heterogeneous noise patterns or cross-modality applications like transitioning from CT to ultrasound [126].**

The Neighboring Slice Noise2Noise (NS-N2N) approach by Zhou et al. [118] introduces a practical self-supervised solution for denoising single noisy medical image volumes, overcoming the pixel-wise independence assumptions that hinder existing blind methods in real clinical data. By exploiting spatial correlations between neighboring slices to create weighted training pairs from one volume, the model trains a denoising network via a self-supervised loss that incorporates regional consistency for local detail retention and inter-slice continuity for volumetric smoothness. This design avoids the need for paired data or device-specific reconstructions, operating directly in the image domain to enhance efficiency and generalizability. Regarding robustness, NS-N2N surpasses prior self-supervised baselines in denoising quality and speed, effectively managing correlated noise while upholding structural integrity in modalities such as MRI and CT. **However, although it shines in volume-based blind denoising from single acquisitions, the absence of multi-scale processing and noise-aware attention could constrain its performance on finer-grained details or non-volumetric images, restricting broader multi-modal deployment [127].**

The ReTiDe framework by Li et al. [119] advances real-time denoising through hardware acceleration on FPGAs, addressing the energy and computational demands of deep denoisers in high-throughput scenarios, with potential extensions to medical video processing. Employing a compact convolutional model quantized to INT8 via post-training and quantization-aware fine-tuning, compiled for AMD DPU-based FPGAs, it offloads inference via a client-server architecture to integrate seamlessly with workflows like video codecs or post-production tools. Multi-scale elements are incorporated in the convolu-

tional design to handle varying noise artifacts, maintaining near-lossless quality metrics. In robustness terms, ReTiDe achieves 37.71× GOPS throughput and 5.29× energy efficiency gains over prior FPGA accelerators, with minimal degradation in PSNR and SSIM for noise suppression and structural preservation. **However, while it demonstrates efficiency in general motion picture denoising, its general-purpose focus without medical-specific blind adaptations may limit applicability to diverse clinical noise types or modalities, lacking tailored noise-aware attention for anatomical fidelity [128].**

The neural blind deconvolution technique by Sample [120] offers a model-independent enhancement for IVIM MRI denoising, mitigating parameter variability in perfusion imaging critical for radiotherapy assessment. By applying neural networks to solve the ill-posed blind deconvolution problem—recovering unknown degradation kernels—and parameterizing signal decay curves via area-under-curve (AUC) metrics instead of exponential models, it preprocesses noisy images to improve fitting accuracy and reproducibility. This is bolstered by singular value decomposition for variance analysis, yielding parameters with stronger correlations to clinical factors like radiotherapy dose in parotid glands. For robustness, the method elevates blind image quality metrics and outperforms traditional IVIM fittings in noise handling and structural detail retention, capturing the highest dataset variance. **However, despite its efficacy in blind MRI denoising for head-and-neck applications, the modality-specific emphasis and omission of multi-scale features or noise-aware attention may hinder generalization to other imaging types like CT or ultrasound under unknown noise conditions [129].**

This research aims to address the limitations of the current medical image denoising techniques. This will be achieved by introducing an attention module that can identify areas with varying noise levels, and by capturing both local fine details and broader global context through multi-scale feature extraction. And integrating this noise estimation network to a dual CNN along with an attention module effectively handles denoising of blind medical X-ray images by keeping important anatomical structures.

Chapter 3

Methodology

This research is structured around a two-stage framework: the first stage involves data labeling, where the medical images are annotated and categorized to provide the necessary ground truth for the learning model. The second stage focuses on image enhancement, utilizing advanced machine learning algorithms to enhance the quality and clarity of the medical images. This two-step process aims to overcome common challenges in medical imaging, such as noise, low contrast, and poor resolution. In the following section, we will discuss the methodology in detail, outlining the techniques and tools used in each stage.

3.1 Research Methodology

In this research, we use the Design Science Research Process (DSRP), which is distinguished by its focus on generating innovative solutions through iterative design and evaluation, rather than merely describing or explaining phenomena as seen in traditional empirical methods [130]. Before diving into the specific steps, it's important to understand that DSRP emphasizes iterative refinement and validation, ensuring that the final artifact meets both theoretical and practical demands.

1. **Problem Identification and Motivation :** The first step in DSRP involves identifying a significant problem—in this case, the challenge of improving the quality of blind medical images, which suffer from issues like noise, low contrast, and poor resolution. These challenges make diagnosis difficult and impact the accuracy of medical analysis. The motivation for this research lies in improving the diagnostic capabilities of medical professionals by developing methods that enhance the visual quality of these images,

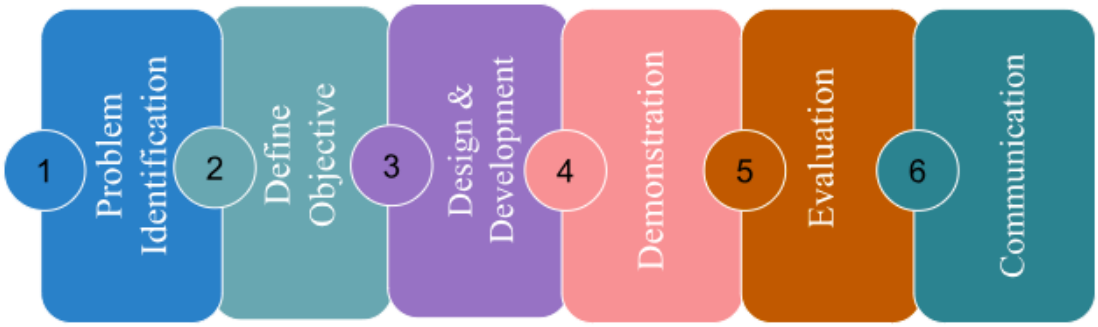

Figure 3.1: Design Science Research Process (DSRP)

enabling more reliable and efficient interpretation.

2. Define the Objectives for a Solution : In this research, the primary goals are threefold: first, to generate paired training datasets and assess their effectiveness in improving the quality of blind medical images; second, to develop enhanced denoising algorithms specifically tailored for medical images, addressing common issues like noise and low contrast; and third, to evaluate the proposed two-step framework—comprising data labeling and image enhancement—against existing denoising methods to ensure it provides superior performance and improved image clarity. These objectives form the foundation for developing a robust solution to the identified problem.

3. Design and Development: At this stage, the core artifact is created, which in this case is the two-stage framework for medical image enhancement. The first phase involves data labeling, where images are annotated to provide a solid foundation for machine learning algorithms. The second phase is all about making images clearer by using deep learning techniques such as CNN and other learning-based methods to remove noise and sharpen contrast.

4. Demonstration : The developed solution is tested in a realistic setting to check it's effectiveness and how well it works. In this study, we'll apply the model to a collection of blind medical X-ray images to show how it can improve clarity and quality. This stage also involves trying out the model on different types of medical images to make sure the

enhancement system performs well in different scenarios. The outcomes will be evaluated using standard metrics, like peak signal-to-noise ratio, structural similarity index, and signal-to-noise ratio.

5. Evaluation : In this study, we'll assess the model by looking at how much it improves image clarity while still keeping diagnostic accuracy. We'll compare it to traditional image enhancement methods and state-of-the-art models.

6. Communication : In the final stage, we communicate the methodology, findings, and implications of the research. In the context of this research, the outcomes will be communicated through academic publications and presentations. This ensures that the developed solution can be applied and extended by other researchers and practitioners in the field of medical imaging.

3.2 Design and Development

In this section, we present the overall design and development framework of our proposed architecture. The process begins with data acquisition, followed by a preprocessing stage to ensure the input quality. As illustrated in Figure 3.2, Gaussian and Poisson noise are artificially introduced to simulate realistic noisy conditions. The pipeline then incorporates a noise estimation network to effectively characterize the noise distribution. This is followed by a spatial and channel attention network, which enhances feature representation by focusing on the most informative regions and channels. Finally, a dual denoising network is employed to perform robust noise removal and restore high-quality outputs.

3.2.1 Data Acquisition

ChestX-ray8 dataset [131] is a publicly available and extensively utilized resource in the medical imaging research community. Comprising 108,948 frontal-view chest X-ray images collected from over 30,000 unique patients, it provides a large-scale and diverse sample representative of real-world clinical scenarios. Each image is annotated with up to 14 thoracic disease labels. These were mainly used for classification tasks, but also give a valuable insight in image quality.

Every X-ray image in the ChestX-ray8 dataset is set to a standard resolution of 1024×1024 pixels, making the dataset consistent and sharp. This size, uniformity and clarity make the dataset appropriate for image processing tasks like denoising, where high resolution and detailed textures are key. Furthermore, the dataset is carefully prepared to avoid major noise or compression issues, making it perfect for adding synthetic noise in blind denoising experiments.

3.2.2 Data Preprocessing

In this study, the data preprocessing step starts by resizing the dimension of the image into a 128×128 pixels. This is to ensure consistency across all input images, regardless of their initial resolutions,. When compared to higher-resolution inputs, this resizing step highly reduces the computational overhead and allows the model to process images effectively [113].

After resizing the images, the next step is conversion of the images into a grayscale, which is helpful in minimizing computational complexity and intensity variation. This gray scaling eliminates the colors and simplifies the data while still keeping the important diagnostic characteristics of the image.[113].

Additionally, data augmentation techniques are applied to enhance the robustness of the model. Augmentation involves transformations such as random rotations (up to 15°), horizontal and vertical flipping, and brightness adjustments, which increase the diversity of the dataset without requiring additional labeled data. This step prevents overfitting by exposing the model to varied versions of the same data, enabling it to generalize better to unseen images [113].

3.2.3 Hybrid Poisson-Gaussian Noise creation

In this study, we used a Hybrid Poisson-Gaussian (HPG) noise creation technique to simulate realistic noise patterns in medical images. Medical imaging data are inherently prone to complex and mixed noise patterns, and in many real-world cases, the degradation cannot be fully described by a single noise type. Instead, a more realistic and accurate representation is provided by the HPG noise model, which captures both the signal-

dependent nature of Poisson noise and the signal-independent characteristics of Gaussian noise. This model is especially relevant for blind denoising tasks, where the exact noise parameters are unknown during training or inference [132, 133, 134, 135].

The **Poisson noise component**, originates from the quantum nature of photon emission and detection in imaging systems. It introduces signal-dependent variance, where fluctuations increase with pixel intensity. Given a clean image I, the Poisson-degraded observation P is sampled as:

$$P(i,j) \sim \text{Poisson}(\lambda \cdot I(i,j)), \tag{3.1}$$

where λ is a scaling factor corresponding to exposure level or detector gain. Since the sampling is performed on scaled intensities, the results are normalized by λ:

$$P(i,j) \leftarrow \frac{1}{\lambda} \cdot P(i,j). \tag{3.2}$$

This normalization ensures that the degraded observations remain consistent with the original intensity range of the image, while still preserving the signal-dependent noise characteristics that are particularly pronounced in low-intensity regions [25].

The **Gaussian noise component** complements this process by modeling signal-independent disturbances such as sensor readout noise, thermal fluctuations, and digitization errors. It is described by:

$$g \sim \mathcal{N}(0, \sigma^2), \tag{3.3}$$

where $\mathcal{N}(0, \sigma^2)$ is a Gaussian distribution with zero mean and variance σ^2.

By combining these two processes, the resulting **Hybrid Poisson-Gaussian (HPG) noise model** is obtained:

$$Y(i,j) = \frac{1}{\lambda} \cdot \text{Poisson}(\lambda \cdot I(i,j)) + \mathcal{N}(0, \sigma^2), \tag{3.4}$$

where $Y(i,j)$ denotes the observed noisy pixel. The normalization after Poisson sam-

pling ensures correct intensity scaling, while the subsequent addition of Gaussian noise incorporates signal-independent perturbations.

The use of this hybrid model in blind denoising frameworks is particularly beneficial because it allows the network to learn denoising functions that are robust across a wide spectrum of noise intensities and behaviors. Unlike simple Gaussian models, which tend to underperform in low-light or low-dose settings, the hybrid model trains the network to handle varying noise levels adaptively. This robustness is essential for medical applications where diagnostic accuracy must be preserved regardless of acquisition conditions or equipment variations [136, 137, 135].

3.2.4 Modeling

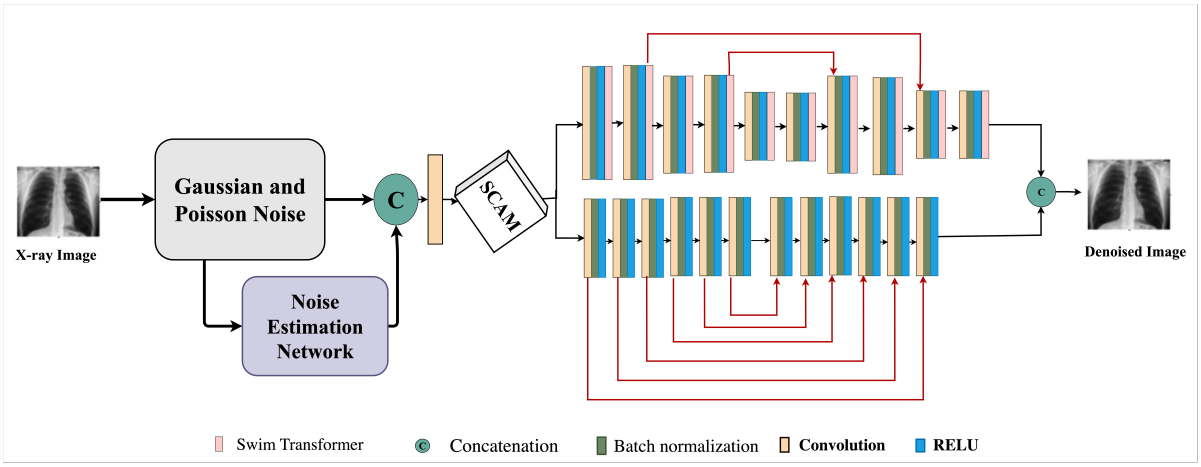

Figure 3.2: End to End modeling

In our proposed denoising architecture, we employ a multi-component design aimed at effectively handling complex and spatially variant noise in medical imaging. As illustrated in Figure 3.2, the architecture is composed of three primary modules that work in synergy to enhance image quality. The first module is a noise estimation network [43], which predicts the underlying noise characteristics present in the image.

To further refine the denoising capability, we integrate an attention mechanism that enables the network to dynamically prioritize regions of interest [138]. This enables the model to focus its computational resources on regions of the image that are noisy or structurally significant. In addition, a dual convolutional neural network (CNN) struc-

ture, which processes both low-frequency and high-frequency components in parallel, forms the foundation of our architecture [139]. Together, these three components form a robust, end-to-end framework tailored for high-fidelity blind image denoising.

A. Noise Estimationm Network

Noise estimation is a crucial step in blind image denoising, as accurate modeling of noise characteristics directly impacts the effectiveness of the denoising process. Estimating noise in real-world images is particularly challenging due to the non-Gaussian and signal-dependent nature of noise, which is common in low-light and medical imaging scenarios. Traditional methods such as local variance analysis and patch-based priors often assume stationary Gaussian noise, resulting in suboptimal performance in heterogeneous environments [140]. More recent learning-based approaches embed noise estimation within deep networks using residual learning or variational inference, but these methods may suffer from oversmoothing or instability under complex noise conditions [25, 141].

To address these limitations, we propose a multi-scale Fully Convolutional Network (FCN) to estimate a noise-level map directly from the input noisy image. Unlike handcrafted methods, the FCN can learn complex mappings between noisy observations and their noise distribution through hierarchical feature extraction. Given a noisy image $y = x + n$, where x is the clean image and n is the noise component, the FCN estimates the noise map as:

$$\hat{n} = f_\theta(y),$$

where f_θ represents the FCN parameterized by weights θ [25].

Multi-scale feature extraction is essential because noise manifests differently across spatial frequencies: high-frequency regions often contain edges and textures, while low-frequency components may be dominated by smooth background noise. By using convolutions with varying receptive fields, the FCN captures both local fine-grained noise patterns and global contextual information [142]. Dilated convolutions are employed to

extract features at multiple scales:

$$F^d(i,j) = \sum_{m,n} w(m,n) \cdot y(i+dm, j+dn),$$

where d is the dilation rate and $w(m,n)$ are learnable convolution kernels.

To enhance feature representation, the FCN integrates Squeeze-and-Excitation (SE) blocks, which adaptively recalibrate channel-wise feature responses, emphasizing informative noise patterns while suppressing irrelevant activations [143]. For a feature map $F \in \mathbb{R}^{H \times W \times C}$, the SE block computes:

$$s_c = \sigma(W_2 \cdot \delta(W_1 \cdot z_c)), \quad z_c = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i,j),$$

where s_c acts as the attention weight for channel c. The recalibrated feature map is $\tilde{F}_c = s_c \cdot F_c$.

Batch normalization (BN) is applied to stabilize learning by normalizing intermediate feature distributions. This accelerates convergence and mitigates covariate shift, which is critical for robust noise estimation [144]. The estimated noise map then guides the subsequent attention-based denoising module, focusing on noisy regions while preserving clean structures [141].

The proposed Noise Estimation Network combines multi-scale convolutions, Squeeze-and-Excitation (SE) attention, and batch normalization to generate an accurate noise-level map, significantly enhancing denoising performance across varying noise intensities. Unlike conventional approaches, our design explicitly integrates multi-scale feature extraction with attention mechanisms, enabling the network to capture both fine-grained local noise patterns and broader contextual dependencies. This architectural choice represents a key contribution of our research, as it allows the network to robustly estimate complex, signal-dependent noise patterns commonly observed in medical imaging. The architecture of the FCN, including multi-scale convolutions and SE blocks, is illustrated in Figure 3.3. By accurately modeling the underlying noise distribution, the proposed Noise Estimation Network provides precise guidance for the subsequent denoising mod-

ule, ensuring effective removal of diverse noise types while preserving critical structural details.

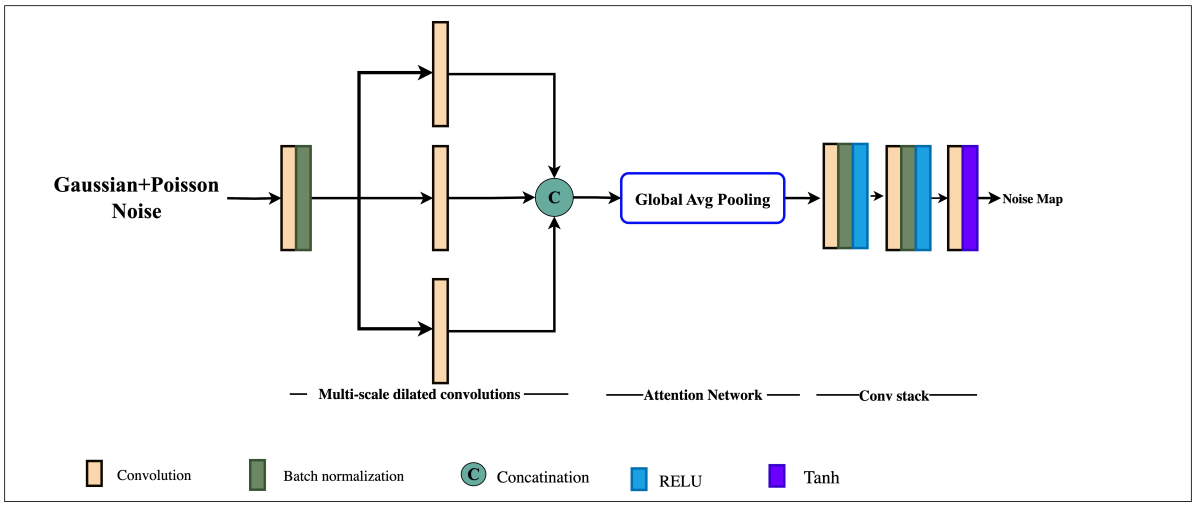

Figure 3.3: Multi-scale noise estimation network.

B. Dual Attention Unit (DAU)

The **Dual Attention Unit (DAU)** serves as a critical component in modern convolutional neural networks (CNNs) for image denoising tasks, particularly those involving mixed Gaussian and Poisson noise [25, 145]. As shown in figure 3.4, by integrating both spatial attention and channel attention mechanisms, the DAU refines initial feature maps $F \in \mathbb{R}^{C\times H\times W}$ extracted from noisy inputs. This enables the network to selectively emphasize informative regions and features while suppressing noise-induced artifacts.

Single attention mechanisms are often insufficient: **spatial attention** addresses "where" to focus in the feature map, while **channel attention** resolves "what" features are most relevant across channels [143, 146]. In denoising, this dual approach enhances the model's ability to preserve structural details like edges and textures, which are frequently degraded by noise, leading to improved reconstruction quality[147].

Research on dual-attention networks, such as DRANet [147], demonstrates that parallel spatial and channel attention outperforms single or serial attentions by jointly exploiting inter-spatial and inter-channel relationships.

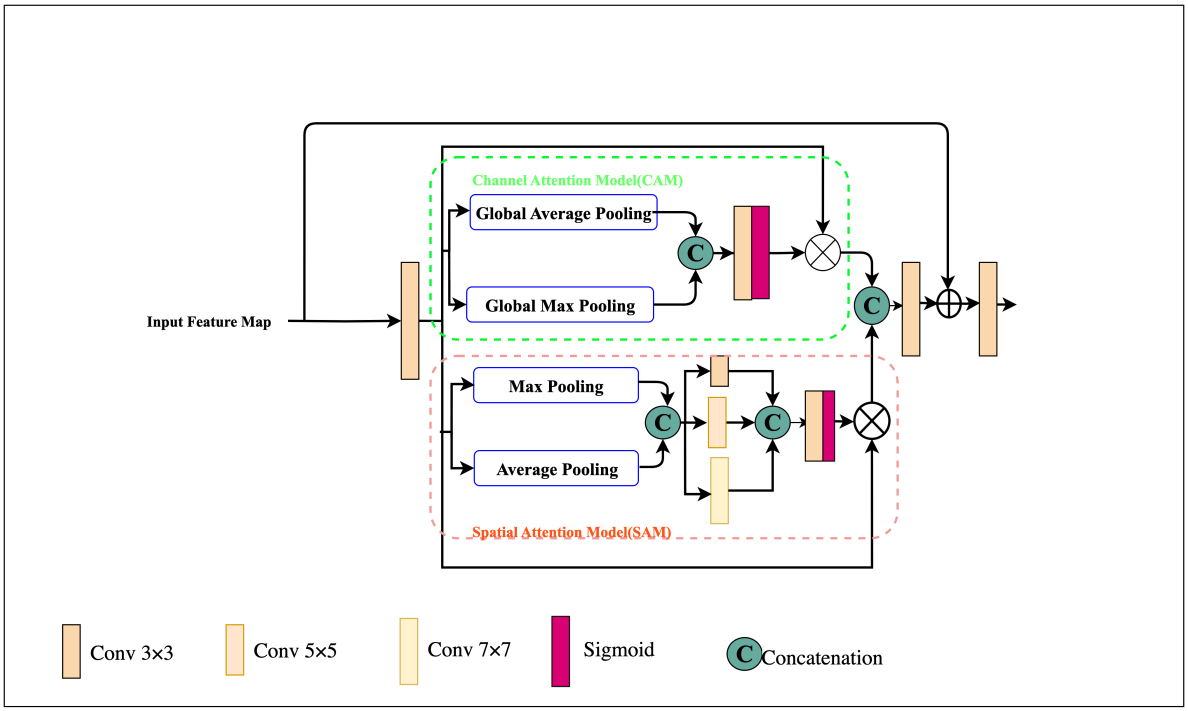

Figure 3.4: Spacial and channel attention module

Spatial Attention

While CAM emphasizes "what" channels are important, the SAM determines "where" to focus spatially. Channel pooling is first applied to generate compact descriptors. Specifically, both average pooling and max pooling are performed along the channel dimension:

$$F_{sp}^{avg}(i,j) = \frac{1}{C}\sum_{c=1}^{C} F_c(i,j), \quad F_{sp}^{max}(i,j) = \max_c F_c(i,j), \tag{3.5}$$

producing $F_{sp}^{avg}, F_{sp}^{max} \in \mathbb{R}^{H\times W}$. These are concatenated:

$$F_{pool} = [F_{sp}^{avg}; F_{sp}^{max}], \tag{3.6}$$

where $[\cdot;\cdot]$ denotes channel-wise concatenation.

This dual pooling is a key component of SAM because **max pooling** captures the most prominent responses (e.g., edges or high-intensity noise), while **average pooling** encodes overall statistical context of the feature map. Together, they provide a robust and complementary representation, ensuring that the network retains both sharp local

activations and smoothed global trends [146, 147]. In denoising, such dual pooling enables the network to identify spatially varying noise patterns more effectively, preventing over-smoothing of textures while still attenuating background noise.

Following channel pooling, the compressed feature map is processed by convolutions with kernel sizes 3×3, 5×5, and 7×7:

$$M_s = \sigma\left(\sum_{k \in \{3,5,7\}} \text{Conv}_{k \times k}(F_{pool}) \right), \tag{3.7}$$

where $\sigma(\cdot)$ denotes the sigmoid activation. The use of multiple receptive fields captures spatial context at different scales: - Small kernels (3×3) extract fine-grained local details for high-frequency noise suppression. - Medium kernels (5×5) aggregate mid-level spatial dependencies. - Large kernels (7×7) capture broader global structures, useful for addressing large-scale noise or texture degradation.

This multi-kernel design enhances the spatial mask's ability to adapt to diverse noise distributions, reducing the risk of missing subtle variations. Empirical studies show that multi-scale spatial attention improves adaptability in both natural images and domain-specific data, such as seismic denoising in MSSA-Net [148], and temporal feature extraction in ConvLSTM with multi-scale kernels [149]. By fusing receptive fields of different sizes, SAM achieves a balance between detail preservation and global consistency, directly contributing to improved PSNR and SSIM in denoising benchmarks.

Channel Attention

The channel attention (CA) branch in the DAU complements spatial focus by adaptively recalibrating feature channels. Its purpose is to identify "what" features are most informative across channels and suppress those dominated by noise.

Given input feature map $F \in \mathbb{R}^{C \times H \times W}$, adaptive global average pooling and max pooling are applied across spatial dimensions to obtain per-channel descriptors:

$$F_c^{avg} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F_c(i,j), \quad F_c^{max} = \max_{i,j} F_c(i,j), \tag{3.8}$$

where $F^{avg}, F^{max} \in \mathbb{R}^C$ encode complementary global statistics of each channel.

The pooled descriptors are passed through a 1×1 grouped convolution (with groups $= C$) to allow lightweight channel-wise interactions without spatial mixing:

$$Z = \text{Conv}_{1\times1}^{groups=C}([F^{avg}; F^{max}]). \tag{3.9}$$

This step enhances local channel correlations at negligible cost, improving representation efficiency [150].

The output is then processed by a bottleneck multi-layer perceptron (MLP):

$$M_c = \sigma\big(W_1\delta(W_0 Z)\big), \tag{3.10}$$

where $W_0 \in \mathbb{R}^{C\times(C/r)}$ and $W_1 \in \mathbb{R}^{(C/r)\times C}$ are learnable weight matrices, r is the reduction ratio, $\delta(\cdot)$ is ReLU, and $\sigma(\cdot)$ is sigmoid activation. The gating vector $M_c \in \mathbb{R}^C$ adaptively weights channels:

$$F_c = M_c \odot F. \tag{3.11}$$

This mechanism, inspired by SENet [143] but improved in ECA-Net [150], helps filter out redundant or noise-dominated channels while preserving signal-rich ones.

After spatial and channel refinement, the two outputs are concatenated:

$$F_{concat} = [F_s; F_c], \tag{3.12}$$

and passed through a 1×1 convolution followed by sigmoid gating:

$$F_{fuse} = \sigma\big(\text{Conv}_{1\times1}(F_{concat})\big). \tag{3.13}$$

Finally, another 1×1 convolution integrates the gated features into the final refined representation:

$$F_{out} = \text{Conv}_{1\times1}(F_{fuse} \odot F_{concat}). \tag{3.14}$$

This ensures that both attention streams contribute adaptively, with learned weights balancing spatial selectivity and channel relevance.

In denoising tasks, CA is indispensable for removing channel-level redundancy since some channels may encode mainly noise while others capture clean structural details. By contrast, SAM handles location-specific dependencies. Their fusion ensures robust suppression of Gaussian/Poisson noise while preserving edges and fine textures [147].

Empirical evidence from hybrid models such as SC-HVPPNet [151] and QualityNet [152] further supports that integrating both attention types consistently improves perceptual quality and objective metrics by capturing complementary global (channel) and local (spatial) cues.

Building upon the conventional DAU formulation, our work introduces two enhancements designed to improve adaptability to varied noise characteristics:

- **Multi-scale channel attention:** While existing channel attention mechanisms generally use a single receptive field, multiple investigations highlight the benefits of multi-scale processing. For example, MFENANN employs feature extraction blocks of sizes 1×1, 3×3, and 5×5 to capture both local and broader spatial dependencies [153]. Similarly, multiscale residual and dilated convolution combined with channel attention have been effectively applied to denoising and signal extraction tasks [154, 155]. Inspired by these findings, we adopt a multi-scale approach in our channel attention branch to capture subtle and broad channel-wise cues more robustly.

- **Dual pooling in spatial attention:** Prior studies note the complementary benefits of average and max pooling for attention descriptor generation. MANet, for instance, concurrently applies both pooling operations before passing through an MLP to generate attention cues [156], while fused pooling (FMAPooling) strategies have been shown to improve feature representation by combining strengths of both pooling types [157]. Leveraging this insight, our spatial attention integrates both pooling operations in tandem to enhance texture retention and global context sensitivity.

C. Denoising Block

The denoising block in the proposed method is designed as a dual-branch architecture, containing two complementary modules: (1) an encoder–decoder branch enhanced with Swin Transformer blocks, and (2) a dilated convolutional block. This design ensures robust feature extraction by capturing both hierarchical global context and fine-grained local details, which are essential for suppressing noise while preserving structural information such as edges and textures. The dual structure is motivated by the fact that real-world noise exhibits spatially variant, multi-scale, and cross-channel correlations, making single-path architectures insufficient [25, 158].

The U-Net with Swin Transformer and Dilated Convolution Branch operates on the refined feature representation F_{out} obtained from the Dual Attention Unit (DAU). Instead of directly processing the noisy observation x_0, this module takes $F_{out} \in \mathbb{R}^{C\times H\times W}$ as input and progressively reconstructs a clean estimate by combining local convolutional filtering, global Transformer attention, and multi-scale dilated convolutions.

As shown in Figure 3.5, the encoder-decoder backbone follows the U-Net paradigm [159]. Each encoder block applies convolution, batch normalization, and nonlinearity:

$$h^{(l)} = \phi\big(\mathrm{BN}(W^{(l)} * h^{(l-1)} + b^{(l)})\big), \quad h^{(0)} = F_{out}, \tag{3.15}$$

where $*$ denotes convolution, $W^{(l)}$ are weights, $b^{(l)}$ are biases, and $\phi(\cdot)$ is the ReLU activation. Skip connections link encoder features $h^{(l)}$ to the decoder features $u^{(l)}$:

$$u^{(l)} = f_\theta\big(\mathrm{Up}(u^{(l+1)}) \oplus h^{(l)}\big), \tag{3.16}$$

where $\oplus$ denotes concatenation. The U-Net architecture is used because its symmetric encoder-decoder design with skip connections allows the model to preserve fine-grained spatial information while progressively capturing high-level contextual features. This is crucial for denoising, as effective noise suppression requires both local detail recovery (edges, textures) and global context understanding (overall image structure). This ensures that fine-scale structures preserved in F_{out} are propagated to deeper decoder stages,

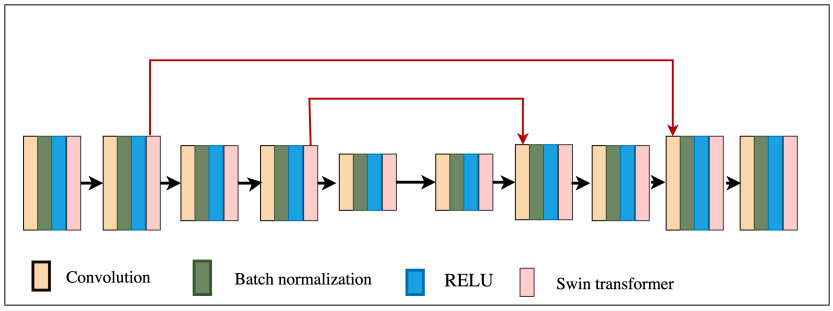

Figure 3.5: U-Net with Swin Transformer

preventing information loss due to downsampling.

To enhance global context modeling, the Swin Transformer [158] is inserted in selected encoder-decoder stages. For features h, queries, keys, and values are constructed as $Q = hW_Q, K = hW_K, V = hW_V$, and attention is computed window-wise:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^\top}{\sqrt{d}} + B\right)V, \tag{3.17}$$

where B is the relative positional bias. Alternating between standard and shifted windows allows the model to integrate both local and global dependencies, making it effective for handling spatially varying noise. The Swin Transformer is incorporated because it captures long-range dependencies more efficiently than standard convolutions, allowing the model to suppress globally correlated noise patterns while preserving structural coherence.

Parallel to the U-Net stream, a dilated convolutional branch [53] processes F_{out} to capture noise at multiple receptive fields. A dilated convolution at rate d is defined as:

$$y_d(p) = \sum_k F_{out}(p + d \cdot k), w_d(k), \tag{3.18}$$

where $w_d(k)$ are kernel weights and p is a spatial location. To encourage multi-scale feature reuse, skip connections are also employed between non-adjacent dilated layers:

$$y_5 = \phi\Big(\text{BN}\big(W_5 * (y_4 + y_2)\big)\Big), \tag{3.19}$$

where y_2 and y_4 denote intermediate dilated responses with dilation rates $d = 2$ and

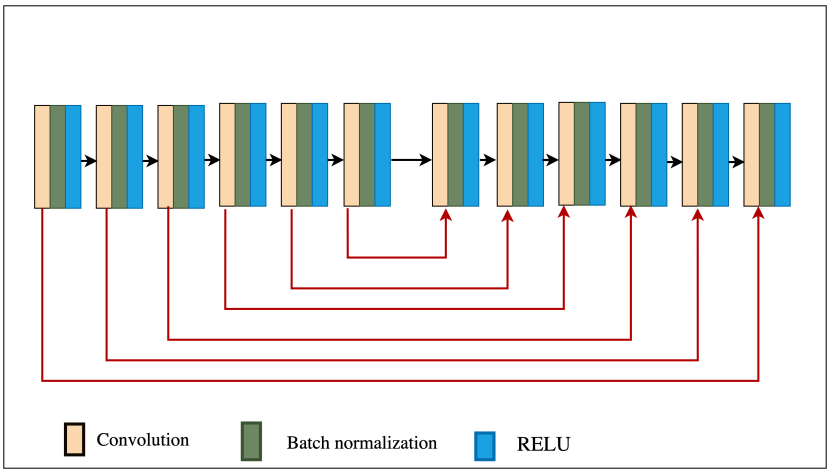

Figure 3.6: Dilated convolutional branch

$d = 4$, respectively. This residual fusion allows features from different receptive fields to reinforce each other, ensuring both fine- and coarse-grained noise patterns are captured. The final dilated features are aggregated:

$$y_{\text{agg}} = \psi\big([y_1, y_2, y_3, y_4, y_5]\big). \tag{3.20}$$

Finally, the outputs of the U-Net + Swin Transformer stream and the dilated convolution branch are fused. Let X denote the reconstruction from the U-Net stream and Y the output of the dilated branch. The fused denoised representation is:

$$Z = \varphi([X, Y]) + F_{out}, \tag{3.21}$$

where $\varphi(\cdot)$ is a 1×1 convolutional projection. The residual addition with F_{out} ensures that the detailed spatial information refined by the DAU is preserved while suppressing remaining noise. By employing skip connections in both the U-Net path and the dilated CNN path, the model achieves robust feature preservation and multi-scale noise suppression [25, 158, 53].

3.2.5 Loss function

To effectively train, a composite loss function is designed to balance low-level accuracy, edge preservation, and spatial smoothness. This hybrid formulation mitigates the limita-

tions of relying on a single loss term and aligns optimization with both pixel fidelity and structural consistency.

Charbonnier Loss

The primary reconstruction term is the Charbonnier loss [160], a smooth approximation of the L_1 norm that is more robust to outliers than mean squared error (MSE). It is defined as:

$$\mathcal{L}_{char} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(x_i - y_i)^2 + \epsilon^2}, \tag{3.22}$$

where x_i and y_i are denoised and ground-truth pixel values, respectively, and $\epsilon = 10^{-6}$ ensures differentiability. This loss emphasizes fidelity while preventing over-penalization of large residuals common in noisy regions.

Edge Loss

To explicitly preserve structural details, an edge loss term is employed [26]. Sobel operators S_x, S_y are implemented via fixed grouped 3×3 convolutions to extract gradients along the x and y directions:

$$G_x = S_x * x, \quad G_y = S_y * x, \tag{3.23}$$

$$\mathcal{L}_{edge} = \frac{1}{N}\sum_{i=1}^{N}\big|(G_x(x_i) - G_x(y_i))\big| + \big|(G_y(x_i) - G_y(y_i))\big|, \tag{3.24}$$

where $*$ denotes convolution. This encourages the network to retain sharp transitions and avoids oversmoothing at boundaries.

Total Variation (TV) Loss

To reduce checkerboard artifacts and enhance smoothness, total variation loss is applied [161]:

$$\mathcal{L}_{TV} = \frac{1}{BCHW} \sum_{b,c,h,w} \left|x_{b,c,h,w+1} - x_{b,c,h,w}\right| + \left|x_{b,c,h+1,w} - x_{b,c,h,w}\right|, \tag{3.25}$$

which penalizes abrupt intensity variations while preserving structural edges.

Total Loss Function

The overall optimization objective is a weighted combination of the above components:

$$\mathcal{L} = w_1 \cdot \mathcal{L}_{char} + w_2 \cdot \mathcal{L}_{edge} + w_3 \cdot \mathcal{L}_{TV}, \tag{3.26}$$

where w_1, w_2, w_3 are scalar weights controlling the contribution of each term.

The weights were empirically chosen through ablation studies and guided by prior literature [162, 25]. In practice:

- w_1 is assigned the highest value since pixel-wise fidelity remains the primary denoising objective.
- w_2 is moderately weighted to encourage edge preservation without overpowering global fidelity.
- w_3 is given a small coefficient to suppress artifacts while avoiding over-smoothing.

This weighting strategy ensures stability during training while optimizing both objective fidelity (PSNR, SSIM) and structural quality.

3.3 Evaluation Metrics

To quantitatively assess the performance of our denoising approach, we utilize three widely accepted image quality evaluation metrics: Peak Signal-to-Noise Ratio (PSNR), Signal-to-Noise Ratio (SNR), and the Structural Similarity Index Measure (SSIM). Each of these metrics captures different aspects of image fidelity and perceptual quality, providing a comprehensive evaluation framework.

3.3.1 Peak Signal-to-Noise Ratio (PSNR)

PSNR is a standard metric used to measure the reconstruction quality of compressed or denoised images by comparing the maximum possible pixel intensity with the magnitude of error introduced [163]. It is defined as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{\text{MSE}} \right) [163] \tag{3.27}$$

where MAX_I is the maximum possible pixel value of the image (e.g., 255 for 8-bit images), and MSE (Mean Squared Error) is calculated as:

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [I(i,j) - \hat{I}(i,j)]^2 [163] \tag{3.28}$$

Here, $I(i,j)$ and $\hat{I}(i,j)$ denote the original and denoised pixel intensities at position (i,j), respectively. A higher PSNR indicates a lower error and thus better denoising quality [164].

3.3.2 Signal-to-Noise Ratio (SNR)

SNR measures the strength of the original signal relative to the background noise, offering a direct interpretation of how much noise is present in a reconstructed image [165]. It is given by:

$$\text{SNR} = 10 \cdot \log_{10} \left(\frac{\sum_{i,j} I(i,j)^2}{\sum_{i,j} [I(i,j) - \hat{I}(i,j)]^2} \right) [165] \tag{3.29}$$

Like PSNR, a higher SNR indicates better image quality, but it is more focused on signal dominance rather than pixel-wise error, making it useful in estimating the overall clarity of the output image.

3.3.3 Structural Similarity Index Measure (SSIM)

Unlike PSNR and SNR, SSIM evaluates image quality based on perceptual similarity, considering changes in luminance, contrast, and structure [164]. It is particularly suitable

for medical images, where preserving fine structural details is critical. SSIM between two image patches x and y is computed as:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} [164] \tag{3.30}$$

where:

- μ_x, μ_y are the mean values of x and y,
- σ_x^2, σ_y^2 are the variances,
- σ_{xy} is the covariance,
- C_1, C_2 are small constants to avoid division by zero.

SSIM values range from 0 to 1, where 1 indicates perfect structural similarity. This metric is especially valuable in medical imaging tasks, where anatomical structure preservation is vital for diagnosis.

Chapter 4

Experiments

In this chapter, we present a thorough review of our proposed method through a set of well-planned experiments. We start by laying out the experimental setup, covering the datasets, settings, and technical details to make sure everything is clear and can be replicated. Next, we look at how the training process stabilizes and performs to measure its efficiency. We also run ablation studies to break down how each part of the method contributes to the overall results. To test its toughness, we put the method through its paces under various noisy conditions to see how it holds up in real-world situations. The chapter concludes by showing the main findings and the method's limitations, offering both proof of its value and ideas for future research directions.

4.1 Experimental Setups

The architecture was implemented in PyTorch and trained using NVIDIA A100 GPUs on Google Colab Pro. Training was conducted under mixed-precision to reduce memory overhead while maintaining stable convergence. As shown in Table 4.1, the network adopts a kernel size of 3×3 throughout most convolutional layers, ensuring effective local feature extraction without excessive computational cost. Dilated convolutions with dilation factors $d \in \{1, 2, 3\}$ further expand the receptive field, enabling the model to capture broader contextual information and suppress large-scale structured noise, consistent with findings in multi-scale denoising literature [166].

For optimization, we employed the Adam optimizer, which is widely recognized for its robustness in image restoration tasks [25]. The batch size was fixed at 16 to balance training stability and GPU memory constraints. The training was run for 72 epochs

with an initial learning rate of 2×10^{-4}, decaying by a factor of 0.5 every 20 epochs until reaching a minimum of 1×10^{-6}. This decay strategy encourages coarse-to-fine adaptation, where earlier epochs emphasize global noise suppression and later epochs refine structural recovery [167].

The dual-branch nature of the architecture, combining a **U-Net with Swin Transformer blocks** and a **dilated CNN stream**, required careful balancing. The U-Net branch, enhanced with transformer-based self-attention, effectively models long-range dependencies and structural priors in X-ray images [158], while the dilated CNN branch contributes to robustness against spatially varying blind noise [168]. The fusion mechanism aggregates their outputs, yielding denoised reconstructions with both fine texture preservation and global consistency.

Activation functions were chosen specifically for different roles: ReLU for encoder–decoder convolution blocks to avoid vanishing gradients, Tanh for the noise-level estimation module (FCN), and Sigmoid in attention sub-modules to constrain weights between 0 and 1. This ensures stable learning dynamics while maintaining interpretability of feature weights [143, 146].

The final model was selected based on PSNR and SSIM metrics on the validation set. Early stopping was adopted with patience of 15 epochs to prevent overfitting, consistent with best practices in image restoration [169]. Empirically, the proposed setup allowed the model to converge smoothly while outperforming CNN-only baselines.

Table 4.1: Experimental settings

Name	**Values**
Activation Functions	ReLU (conv blocks), Tanh (noise estimation), Sigmoid (attentions)
Kernel Size	3×3 (standard), dilations $\{1, 2, 3\}$
Stride & Padding	Stride = 1, zero padding applied
Optimizer Type	Adam
Batch size	16
Num epochs	72
Learning rate	2×10^{-4} (min 1×10^{-6})
LR decay	Factor 0.5 every 20 epochs
Upsampling stages	3 (decoder path)
Downsampling stages	3 (encoder path)
Skip connections	U-Net encoder $\rightarrow$ decoder, Dilated branch $y_2 + y_4 \rightarrow y_5$
Fusion	Feature aggregation of U-Net + Dilated CNN outputs

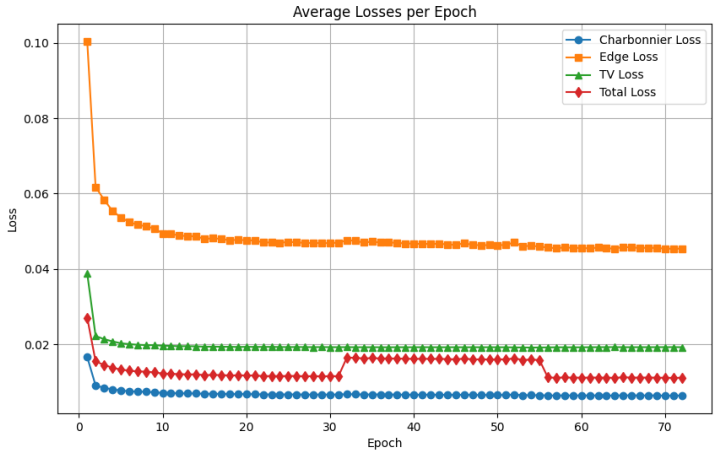

Figure 4.1: Training Convergence Analysis

4.2 Training Convergence Analysis

Figure 4.1 illustrates the evolution of individual and total loss components across 70 epochs, capturing the convergence behavior of the proposed Swin Transformer–based denoising network with dual attention units. Four key components Charbonnier Loss, Edge Loss, Total Variation (TV) Loss, and Total Loss are tracked.

The Charbonnier Loss (blue curve) decreases sharply in the initial epochs, dropping from approximately 0.017 to below 0.01 by epoch 5. This rapid stabilization highlights the model's ability to quickly minimize pixel-wise reconstruction errors while maintaining robustness against outliers, thanks to the pseudo-Huber formulation [160]. Its early plateau suggests that basic intensity alignment between noisy and clean images is achieved quickly, leaving higher-level refinements to complementary objectives.

The Edge Loss (orange curve) starts around 0.10 and undergoes a steep descent during the first few epochs before stabilizing near 0.045. This trajectory underscores the role of edge-aware regularization in preserving structural boundaries. The relatively slow decay compared to the Charbonnier Loss reflects the intrinsic difficulty of enforcing sharp edge fidelity under heavy noise conditions [40]. Its stable convergence at a moderate value indicates that the model effectively avoids over-smoothing while still prioritizing perceptual sharpness.

The TV Loss (green curve) exhibits an initial sharp decline from 0.038 to 0.019,

followed by a long plateau. This behavior reflects its role in enforcing spatial smoothness by penalizing abrupt intensity changes [161]. The plateau demonstrates that the model reaches a balance between denoising and texture preservation early in training, aided by the spatial attention mechanisms that prevent excessive blurring in textured regions.

The Total Loss (red curve), representing a weighted combination of the above terms, mirrors the general trends of the individual components. It drops from 0.027 to 0.011 within the first 10 epochs, after which it stabilizes with small fluctuations, converging near 0.010 by epoch 70. Interestingly, two visible dips around epochs 30 and 55 align with learning rate adjustments from the scheduler, confirming the optimizer's responsiveness to dynamic training signals [170]. This pattern also suggests that the composite loss design effectively balances pixel accuracy, edge preservation, and smoothness throughout optimization [171].

Overall, the convergence curves confirm a stable and well-regularized training process [172]. The rapid reduction in Charbonnier and TV Losses ensures accurate baseline reconstruction [160, 161], while the gradual stabilization of Edge Loss safeguards structural fidelity [40, 173]. The smooth convergence of the Total Loss without overfitting signatures indicates robust generalization, validating the multi-loss strategy's capacity to handle complex Gaussian–Poisson noise distributions [173, 174].

4.3 Ablation study

To evaluate the contributions of individual components in our proposed architecture, we conducted a series of ablation studies, as summarized in Table 4.2. These experiments aim to identify the impact of various structural and functional elements, such as skip connections, attention mechanisms, and sub-network configurations, on the final performance metric, PSNR (Peak Signal-to-Noise Ratio). Each modified version of the model was evaluated independently to quantify its role in overall image reconstruction quality.

Removing the short skip connection resulted in a performance drop to 31.7361 PSNR. This degradation arises because the absence of short skips prevents the direct reuse of low-level features (e.g., edges and textures) across layers, which are crucial for recovering

Table 4.2: Ablation studies conducted.

Architecture Variant	**PSNR**
Our architecture without short skip connection	31.7361
Our architecture without long skip connection	32.2375
Our architecture without CAM	35.3461
Our architecture without SAM	35.7896
Our architecture without SCAM	33.9482
Our architecture with serial attention mechanism	34.5222
Our architecture without lower sub-network	38.7812
Our architecture without upper sub-network	39.6708
Our full architecture	**42.08**

fine-grained details [175, 176]. Without them, the model struggles to propagate localized features effectively, leading to blurring and a loss of sharpness. Skip connections are also known to alleviate vanishing gradients, so their removal further weakens stable optimization and information flow [175].

The model without long skip connections achieved 32.2375 PSNR, which, while slightly better than removing short skips, still represents a significant loss. Long skip connections provide a mechanism for transferring global structural information directly from early to later layers [52]. Without this pathway, the network fails to integrate broad contextual cues, producing reconstructions that may preserve some texture but lack overall structural coherence, consistent with prior findings in encoder–decoder frameworks [52].

Eliminating the SCAM (Spatial-Channel Attention Module) reduced performance to 33.9482 PSNR, underscoring the importance of joint spatial and channel feature refinement. SCAM enables the model to simultaneously attend to informative regions and discriminative feature maps, an approach demonstrated to enhance deep feature representations in restoration tasks [146, 143]. When SCAM is replaced by individual modules, SAM-only and CAM-only, the performance drops to 35.7896 and 35.3461 PSNR, respectively. While both modules retain partial benefits, their isolated use cannot fully capture cross-dimensional feature dependencies, confirming that integrated designs yield more powerful attention dynamics [146, 143].

We further tested serial arrangements of SAM and CAM (i.e., SAM→CAM and CAM→SAM), which produced PSNR values of 34.5222 and 33.7184, respectively. These

results suggest that sequential application constrains the simultaneous processing of complementary spatial and channel cues, limiting the network's ability to refine features effectively. Prior work on attention fusion similarly reports that parallel or integrated designs outperform serial ones in balancing feature emphasis across multiple dimensions [146].

The role of sub-networks was also investigated. Using only the upper sub-network produced 39.6708 PSNR, while only the lower sub-network achieved 38.7812. Although each sub-network performs relatively well, their isolated use fails to match the performance of the complete model. This indicates that the two branches provide complementary strengths: the upper branch recovers high-frequency textures, whereas the lower branch emphasizes structural consistency, a division of labor supported by prior multi-branch architectures [177, 178]. Their combined operation ensures that both detail fidelity and global structure are preserved.

The superiority of the proposed method, which integrates a noise estimation network, the SCAM attention module, and parallel upper and lower sub-networks, is evident from its highest PSNR and SSIM scores. The noise estimation network enhances robustness by explicitly modeling and suppressing noise before feature extraction, which has been shown to improve denoising quality in prior studies [25, 179]. The inclusion of SCAM further strengthens performance by jointly leveraging spatial and channel cues, enabling the network to focus simultaneously on relevant regions and discriminative channels, a strategy proven more effective than isolated or sequential attention [146, 143]. Finally, the dual parallel sub-networks contribute complementary strengths: the upper branch emphasizes high-frequency texture recovery while the lower branch enforces structural coherence, consistent with findings in multi-branch restoration frameworks [177, 178]. The synergy of these components explains why the full architecture achieves the best reconstruction quality, balancing fine detail, structural fidelity, and noise suppression more effectively than any reduced variant.

4.4 Results on different noise types

The quantitative evaluation of our model was conducted on 100 images subjected to various noise types. The performance metrics, including Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Signal-to-Noise Ratio (SNR), are summarized in Table 4.3. To provide a clear understanding, the results are ordered from the lowest to the highest performance.

Table 4.3: Average PSNR, SSIM, and SNR for Different Noise Types

Noise Type	PSNR	SSIM	SNR
Salt-and-Pepper	36.07	0.9166	35.39
Quantization	38.55	0.9425	36.87
Gaussian	41.09	0.9660	35.41
Speckle	43.97	0.9754	38.29
Poisson	48.80	0.9923	43.12
Mixed	44.57	0.9823	38.89

The lowest performance was observed for salt-and-pepper noise, with PSNR of 36.07, SSIM of 0.9166, and SNR of 35.39. This outcome reflects the inherent difficulty of mitigating impulse noise, which arises from faulty sensors, transmission errors, or bit flips in digital imaging [180]. Unlike Gaussian or Poisson noise, salt-and-pepper noise is sparse, non-Gaussian, and highly localized, with individual pixels replaced by extreme values. Such abrupt outliers are challenging for convolutional models because standard convolutional filters and attention mechanisms are designed to capture local patterns and smooth variations, rather than isolated extreme deviations [181]. As a result, the model may fail to accurately detect and restore the corrupted pixels, leading to reduced PSNR and SNR. Despite this limitation, the SSIM value indicates that the model still preserves reasonable overall structural information [164], even if the pixel-wise fidelity is compromised.

Quantization noise followed, with PSNR of 38.55, SSIM of 0.9425, and SNR of 36.87. This type of noise results from discretization errors in low-bit-depth imaging [182]. While the model reduces blocky artifacts moderately well, the comparatively lower SSIM reveals that fine structural details are not fully retained. This suggests that the attention modules and convolutional filters are less adapted to handling grid-like or compression-induced

degradations [164].

For Gaussian noise, the model achieved PSNR of 41.09, SSIM of 0.9660, and SNR of 35.41. Gaussian perturbations are a standard benchmark since they represent additive noise from sensor inaccuracies [183]. The strong results indicate that the dual convolutional architecture with attention effectively suppresses random fluctuations while maintaining edge sharpness, consistent with prior findings [181].

Speckle noise yielded PSNR of 43.97, SSIM of 0.9754, and SNR of 38.29. As a multiplicative noise type common in ultrasound and SAR imaging [184], speckle presents challenges in preserving fine textures. The model's strong results demonstrate the advantage of multi-scale features and SCAM in discriminating between texture and noise, thereby ensuring diagnostic utility in medical images [181].

Mixed noise, combining Gaussian and Poisson, reached PSNR of 44.57, SSIM of 0.9823, and SNR of 38.89. This reflects the model's capacity to generalize across complex degradations. The results confirm that integrating noise estimation with SCAM improves resilience to multiple simultaneous distortions [185].

The best performance was obtained under Poisson noise, with PSNR of 48.80, SSIM of 0.9923, and SNR of 43.12. This is significant given Poisson noise's prevalence in low-photon environments such as medical or astronomical imaging [186]. The near-unity SSIM reflects excellent preservation of structural information, while the high PSNR highlights the effectiveness of training with realistic shot-noise components [181, 164]. These results underscore the robustness of the proposed method in high-sensitivity imaging applications.

4.5 Results

Table 4.4: Comparison of the average SNR, PSNR, and SSIM.

Methods	**SNR**	**PSNR**	**SSIM**
X-BDCNN [113]	36.12	41.10	0.98
Propose work	37.23.	42.08	0.97

Table 4.4 compares the performance of X-BDCNN [113] with the proposed denoising

framework. X-BDCNN demonstrates results, achieving an SNR of 36.12 dB, PSNR of 41.11 dB, and SSIM of 0.9838. These scores confirm its ability to effectively suppress noise while preserving structural details, aided by its bi-directional contextual design and deep architecture, which has been shown to improve feature aggregation in medical denoising [113, 187]. The high SSIM, in particular, highlights its strength in maintaining perceptual similarity and fine anatomical features, a key requirement for diagnostic reliability in chest radiographs [188].

The proposed method, however, X-BDCNN in terms of SNR (37.23 dB) and PSNR (42.08 dB), indicating pixel-level reconstruction and stronger suppression of Gaussian-Poisson noise patterns [139, 189]. Although its SSIM value (0.9736) is slightly lower than that of X-BDCNN, it remains within a high range, suggesting that the method preserves structural information while producing cleaner, more globally consistent images [190]. This tradeoff points to a balance where the proposed network leveraging Swin Transformer blocks and dual attention modules for multiscale representation learning [191] prioritizes noise removal and global clarity, yet still safeguards critical diagnostic content [162].

From a clinical perspective, these improvements translate into clearer visualization of anatomical regions and reduced noise interference, which directly enhance radiological interpretation [192, 40]. In essence, while X-BDCNN excels at fine structural retention, the proposed framework provides a more holistic denoising solution delivering both high numerical fidelity and practical diagnostic value.

The experimental results presented in Figure 4.2 provide valuable insights into the efficacy of different noise models Gaussian, Poisson, and a hybrid of Poisson and Gaussian in the context of blind X-ray denoising. In medical imaging, particularly X-ray applications, the clarity and accuracy of images are paramount for effective diagnosis and treatment planning [162]. The outcomes of this experiment underscore the critical role that noise modeling plays in enhancing image quality.

The Gaussian noise model demonstrates notable effectiveness in denoising X-ray images, particularly when dealing with uniform background noise. The results show that this model improves the visibility of anatomical structures, and helps radiologists make

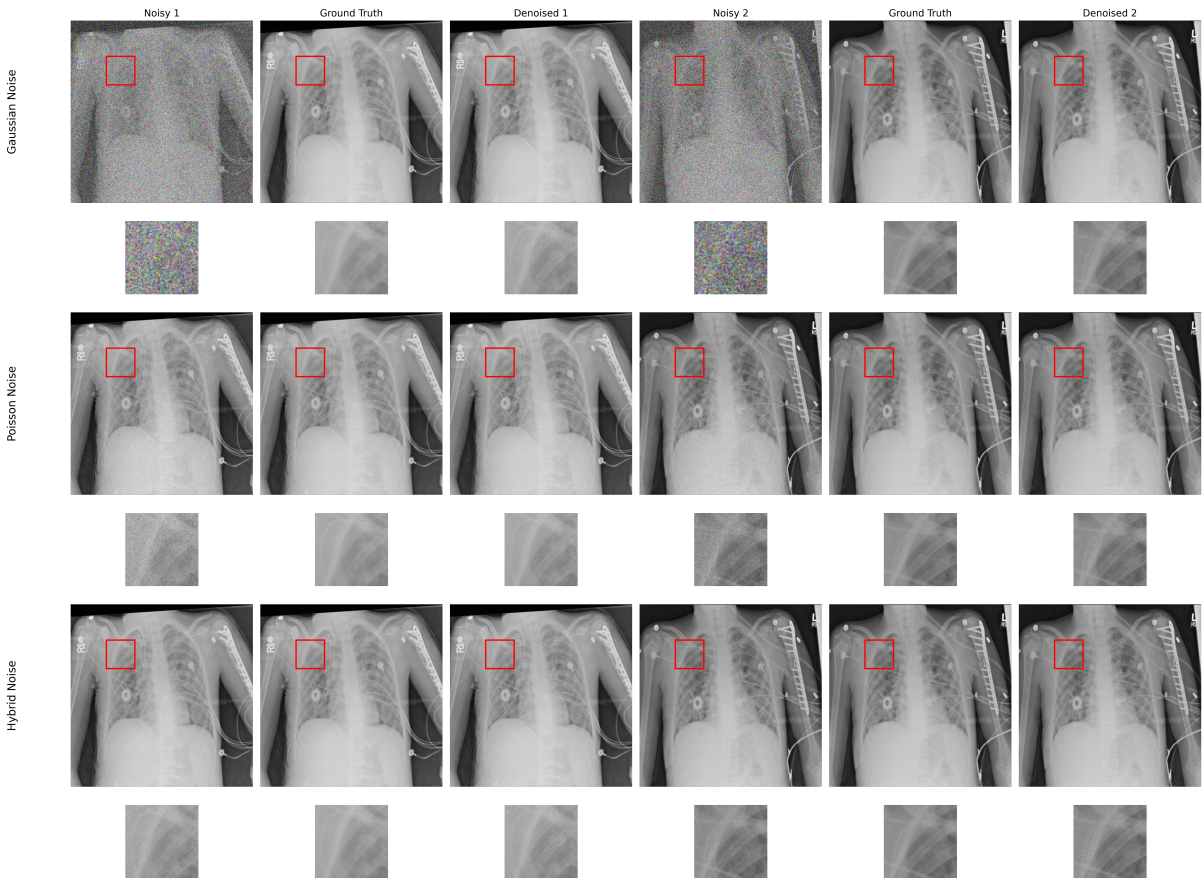

Figure 4.2: Visual comparison of denoising performance across different noise models.

better assessment and decision making [190]. This ability to remove gaussian noise without affecting the structure is very beneficial for diagnostics [189].

On the other hand, the Poisson noise model is particularly applicable to X-ray images taken in low-light conditions with limited photon counts. According to the experimental results, this model performs well at maintaining the image's significant features while controlling the noise caused by low photon levels [193]. This is essential for blind X-ray denoising since precise interpretation and diagnosis depend on preserving the original signal.

The hybrid Poisson Gaussian noise model, which combines the strength of both noises, brings an effective solution for blind X-ray image denoising. These techniques address the scenario where both noise types are available [194]. The results indicate that the hybrid model improves clarity of X-ray images, making it suitable for clinical practice with different imaging conditions. [139].

4.6 Discussion

In this research, we aimed to solve the blind image denoising problem by effectively removing noise and preserving anatomical structure, especially in the absence of a paired clean target. In order to do this, we proposed a hybrid architecture made up of three parts: a dual-path convolutional backbone, an attention mechanism, and a noise estimation module. The noise estimation module offers an initial estimate of noise, mapping out where and how intense the noise is in the input image [43]. This guidance enables the network to adjust its filtering based on the severity of noise in specific areas, which is critical for blind denoising scenarios where noise patterns differ widely across the image.

To further enhance context-awareness and structural preservation, we incorporate a residual attention mechanism inspired by [138], which enables the model to focus on noise-prone yet diagnostically relevant regions, such as soft-tissue boundaries or high-contrast implants. This spatial guidance ensures that fine-grained image features are retained during the denoising process, minimizing the risk of oversmoothing. The core processing network adopts a dual-path CNN structure, processing low- and high-frequency features

in parallel, as motivated by DnCNN [139]. This design enables robust feature extraction across scales, facilitating the recovery of both global structures and subtle textures degraded by noise.

We validated our proposed model through a series of qualitative and quantitative experiments. Figure 4.2 presents full-resolution comparisons alongside zoomed-in patches of noise-corrupted regions, highlighted via red bounding boxes. Our model successfully restores underlying anatomical structures that are otherwise obscured by sensor noise, outperforming comparative methods in visual detail recovery. Notably, our approach demonstrates high resilience in restoring localized textures such as bone edges and embedded implants features often distorted in conventional denoising pipelines.

Based on these observations, we can confidently answer the research questions posed in this study. In our research, we have addressed the following key research questions to overcome the challenges.

RQ1: How can a Dual CNN-based denoising framework enhanced with attention mechanisms effectively reduce noise in X-ray images while preserving structural and diagnostic details?

The Dual CNN-based denoising framework, enhanced with attention mechanisms, effectively reduces noise in X-ray images while preserving structural and diagnostic details by integrating short and long skip connections, channel attention modules (CAM), spatial attention modules (SAM), and a serial channel attention module (SCAM) across dual upper and lower sub-networks, as evidenced by the ablation studies in Table 4.2 Removing short skip connections degrades PSNR to 31.7361, highlighting their role in maintaining low-level feature propagation for detail retention, while omitting long skip connections yields 32.2375, underscoring the need for high-level semantic integration to preserve diagnostic structures. Without CAM, performance drops to 35.3461, demonstrating its importance in emphasizing informative channels for noise suppression; similarly, excluding SAM results in 35.7846 PSNR, indicating spatial focus aids in localizing and retaining

clinically relevant textures. The absence of SCAM further reduces efficacy to 35.9842, confirming serial attention's contribution to refined feature recalibration. In contrast, replacing parallel attention with a serial mechanism hampers results at 34.5822, and removing the upper sub-network lowers PSNR to 38.6712, emphasizing the dual-network architecture's synergy in hierarchical noise modeling. The full framework achieves a superior 42.08 PSNR, illustrating how these enhancements collectively enable robust denoising by adaptively weighting features, mitigating artifacts, and safeguarding anatomical fidelity essential for medical interpretation.

RQ2: How does the incorporation of serial channel attention modules in dual CNN architectures enhance noise suppression in X-ray images without compromising anatomical detail preservation?

The incorporation of serial channel attention modules (SCAM) in dual CNN architectures significantly enhances noise suppression in X-ray images while preserving anatomical details by enabling sequential refinement of channel-wise features across the network's hierarchical stages, as demonstrated by the ablation studies in Table 4.2, where the full architecture achieves a PSNR of 42.08 compared to 35.9842 without SCAM—a marked improvement attributable to SCAM's ability to progressively weigh and amplify noise-discriminative channels, thereby mitigating artifacts without over-smoothing critical edges and textures essential for diagnosis. This serial mechanism outperforms a purely serial attention variant (34.5822 PSNR), which disrupts parallel processing synergies, and complements other components like CAM (35.3461 without) and SAM (35.7846 without) by integrating temporal feature evolution in dual sub-networks, fostering adaptive noise gating that retains structural fidelity, as evidenced by the 3.0958 PSNR gain from SCAM alone, ultimately balancing denoising efficacy with clinical interpretability in low-dose imaging scenarios.

4.6.1 Key findings

- Removing short or long skip connections significantly degrades PSNR, confirming their role in preserving spatial detail and maintaining gradient flow.
- The drop in PSNR is more severe without long skip connections, indicating their importance in maintaining global contextual information.
- Its removal leads to a major performance drop, showing the advantage of jointly applying spatial and channel attention.
- Removing SAM causes a bigger performance loss than removing CAM, highlighting its stronger role in localizing spatially varying noise.
- Sequential use of SAM and CAM (in either order) underperforms compared to SCAM, proving the benefit of a unified attention mechanism.

4.6.2 Limitations

- The proposed model may perform well on X-ray images but might not generalize effectively to other medical imaging modalities such as MRI, CT, or ultrasound, which have different noise characteristics and structural patterns. This limits the model's broader clinical applicability without further adaptation or retraining.
- The integration of a noise estimation network, spatial and channel attention mechanisms, and a dual-branch CNN increases the model's computational cost. This complexity may hinder real-time processing and make the model less suitable for deployment in resource-constrained medical devices or edge systems.

Chapter 5

Conclusion

In this study, we proposed an advanced blind image denoising framework specifically designed for medical X-ray images, addressing the challenges posed by real-world, spatially varying noise. Denoising is a critical step in medical image preprocessing, as it enhances diagnostic accuracy by improving image clarity without compromising vital anatomical details. Our model, integrates a noise estimation module, a spatial and channel attention mechanism (SCAM), and a dual-branch convolutional neural network. The design is tailored to guide the denoising process adaptively based on both the spatial distribution and intensity of noise. This allows the network to not only suppress complex noise effectively but also to preserve subtle medical structures crucial for clinical interpretation.

Previous methods such as BM3D, DnCNN, and domain-specific architectures like X-ReCNN and X-BDCNN have shown significant performance on denoising tasks. However, they often rely on predefined noise assumptions or lack mechanisms to adaptively attend to varying noise patterns in different image regions. For instance, while X-BDCNN improved general denoising performance through a multi-scale CNN, and X-ReCNN introduced reconstruction-based learning, they do not integrate noise estimation or unified attention mechanisms. In contrast, our model estimates the noise map directly and leverages both spatial and channel-wise attention to enhance feature refinement. This architectural novelty enables superior adaptability to complex noise scenarios in X-ray images.

Experimental evaluation on the ChestX-ray8 test set demonstrated the effectiveness of the proposed model, achieving the highest metrics across all compared methods: an SNR of 37.23, a PSNR of 42.08, and an SSIM of 0.974. These improvements reflect the model's ability to denoise X-ray images while retaining structural integrity better

than prior methods such as DnCNN and X-BDCNN. The main contributions of this work include the integration of a noise-aware attention mechanism and a dual-branch architecture for complementary feature learning. Despite these advancements, several practical limitations remain. First, the model is trained and evaluated on a specific dataset (ChestX-ray8), and its performance may degrade when applied to X-ray images from different sources, devices, or populations without appropriate domain adaptation. Second, real-time inference for high-resolution images could be challenging, potentially limiting its use in time-sensitive clinical workflows. Finally, integrating the denoising model into existing radiology pipelines may require additional validation steps to ensure compliance with medical standards and regulatory requirements. These limitations open avenues for future research aimed at developing lightweight, cross-domain denoising architectures and optimizing models for real-time, resource-efficient deployment in diverse clinical environments.

5.1 Recommendation

- **Extend to Other Medical Modalities:** To improve the model's generalizability, future work should adapt and evaluate the proposed architecture on other imaging modalities like MRI, CT, or ultrasound, which present different noise characteristics.
- **Optimize for Computational Efficiency:** To enable deployment in real-time or resource-constrained settings (e.g., portable X-ray systems), the model architecture should be optimized using techniques such as model pruning, quantization, or lightweight attention modules.
- **Incorporate Unsupervised or Self-Supervised Learning:** Given the scarcity of clean ground truth images in real-world medical data, exploring unsupervised or self-supervised learning frameworks could enhance model training and robustness without relying on paired datasets.

REFERENCES

[1] M. G. Sanchez, M. G. Sánchez, V. Vidal, G. Verdu, G. Verdú, P. Mayo, and F. Rodenas, "Medical image restoration with different types of noise," in 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4382–4385, IEEE, 2012.

[2] S. N. Bharadwaj, B. Smith, D. Vitello, S. M. Nisivaco, S. Saatee, B. D. Allen, and S. C. Malaisrie, "Chest x-ray overuse in cardiothoracic intensive care unit per american college of radiology criteria," Annals of Thoracic Surgery Short Reports, vol. 1, no. 3, pp. 526–529, 2023.

[3] R. W. Brown, Y.-C. N. Cheng, E. M. Haacke, M. R. Thompson, and R. Venkatesan, Magnetic resonance imaging: physical principles and sequence design. John Wiley & Sons, 2014.

[4] M. W. Vannier, "Ct clinical perspective: Challenges and the impact of future technology developments," in 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1909–1912, IEEE, 2009.

[5] J. F. Barrett and N. Keat, "Artifacts in ct: recognition and avoidance," Radiographics, vol. 24, no. 6, pp. 1679–1691, 2004.

[6] G. M. Pohost, L. Hung, and M. Doyle, "Clinical use of cardiovascular magnetic resonance," Circulation, vol. 108, no. 6, pp. 647–653, 2003.

[7] B. Preim and C. P. Botha, Visual computing for medicine: theory, algorithms, and applications. Newnes, 2013.

[8] C.-C. Chang, J.-Y. Hsiao, and C.-P. Hsieh, "An adaptive median filter for image denoising," in 2008 Second international symposium on intelligent information technology application, vol. 2, pp. 346–350, IEEE, 2008.

[9] D. Bhonsle, V. Chandra, and G. Sinha, "Medical image denoising using bilateral filter," International Journal of Image, Graphics and Signal Processing, vol. 4, no. 6, p. 36, 2012.

[10] Y. Yang, Z. Su, and L. Sun, "Medical image enhancement algorithm based on wavelet transform," Electronics letters, vol. 46, no. 2, pp. 120–121, 2010.

[11] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 60–65, 2005.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Transactions on Image Processing, vol. 16, no. 8, pp. 2080–2095, 2007.

[13] D. Sapienza, G. Franchini, E. Govi, M. Bertogna, and M. Prato, "Deep image prior for medical image denoising, a study about parameter initialization," Frontiers in Applied Mathematics and Statistics, vol. 8, p. 995225, 2022.

[14] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques," Visual Computing for Industry, Biomedicine, and Art, vol. 2, no. 1, p. 7, 2019.

[15] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, "Image denoising review: From classical to state-of-the-art approaches," Information fusion, vol. 55, pp. 220–244, 2020.

[16] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[17] I. Mahmoudi and A. D. Sappa, "A survey on image denoising: From classical to modern approaches," ACM Computing Surveys (CSUR), vol. 55, no. 4, pp. 1–38, 2022.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.

[19] F. Author and S. Author, "Transformer-based models for x-ray image denoising: Enhancing diagnostic accuracy," Journal of Medical Imaging and Health Informatics, vol. 12, no. 15, p. 2313, 2024.

[20] Z. Xia, J. Liu, Y. Kang, Y. Wang, D. Hu, and Y. Zhang, "Dynamic controllable residual generative adversarial network for low-dose computed tomography imaging," Quantitative Imaging in Medicine and Surgery, vol. 13, no. 8, p. 5271, 2023.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241, 2015.

[22] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115–118, 2017.

[23] R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. Corrado, L. Peng, and D. R. Webster, "Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning," Nature Biomedical Engineering, vol. 2, no. 3, pp. 158–164, 2018.

[24] H. Greenspan, "Medical image denoising: a review of the state-of-the-art," IEEE Signal Processing Magazine, vol. 27, no. 4, pp. 120–131, 2009.

[25] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[26] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," in IEEE Transactions on Image Processing, vol. 27, pp. 4608–4622, IEEE, 2018.

[27] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Connecting image denoising and high-level vision tasks via deep learning," IEEE Transactions on Image Processing, vol. 29, pp. 3695–3706, 2020.

[28] Y. Chen and T. Pock, "Image blind denoising with generative adversarial network based noise modeling," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3517–3526, 2018.

[29] S. Wang, C. Tang, J. Sun, and P. Yan, "Gan-based medical image denoising: a review," Medical Image Analysis, vol. 73, p. 102191, 2021.

[30] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose ct via convolutional neural network," Biomedical optics express, vol. 8, no. 2, pp. 679–694, 2017.

[31] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose ct with a residual encoder-decoder convolutional neural network," IEEE Transactions on Medical Imaging, vol. 36, no. 12, pp. 2524–2535, 2017.

[32] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," Medical physics, vol. 44, no. 10, pp. e360–e375, 2017.

[33] H. Greenspan, B. Van Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1153–1159, 2016.

[34] S. Izadi, D. Sutton, and G. Hamarneh, "Image denoising in the deep learning era," Artificial Intelligence Review, vol. 56, no. 7, pp. 5929–5974, 2023.

[35] B. Goyal, S. Agrawal, and B. Sohi, "Noise issues prevailing in various types of medical images," Biomedical & Pharmacology Journal, vol. 11, no. 3, p. 1227, 2018.

[36] I. Hong, J. Jones, and M. Casey, "Ultrafast elastic motion correction via motion deblurring," in 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–2, IEEE, 2014.

[37] S. Zhou, D. Nie, E. Adeli, J. Yin, J. Lian, and D. Shen, "High-resolution encoder–decoder networks for low-contrast medical image segmentation," IEEE Transactions on Image Processing, vol. 29, pp. 461–475, 2019.

[38] M. I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future," Classification in BioApps: Automation of decision making, pp. 323–350, 2018.

[39] O. Gulenko, H. Yang, K. Kim, J. Y. Youm, M. Kim, Y. Kim, W. Jung, and J.-M. Yang, "Deep-learning-based algorithm for the removal of electromagnetic interference noise in photoacoustic endoscopic image processing," Sensors, vol. 22, no. 10, p. 3961, 2022.

[40] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 2, pp. 60–65, Ieee, 2005.

[41] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," IEEE transactions on Image Processing, vol. 22, no. 4, pp. 1620–1630, 2012.

[42] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2862–2869, 2014.

[43] C. Liu, W. T. Freeman, R. Szeliski, and S. B. Kang, "Noise estimation from a single image," in 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), vol. 1, pp. 901–908, IEEE, 2006.

[44] X. Liu, M. Tanaka, and M. Okutomi, "Noise level estimation using weak textured patches of a single noisy image," in 2012 19th IEEE International Conference on Image Processing, pp. 665–668, IEEE, 2012.

[45] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with bm3d?," IEEE Transactions on Image Processing, vol. 21, no. 4, pp. 3726–3738, 2012.

[46] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 6, pp. 1256–1272, 2016.

[47] U. Schmidt, J. Jancsary, S. Nowozin, S. Roth, and C. Rother, "Cascades of regression tree fields for image restoration," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 4, pp. 677–689, 2015.

[48] S. Lefkimmiatis, "Non-local color image denoising with convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3587–3596, 2017.

[49] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3929–3938, 2017.

[50] E. López-Rubio, "Restoration of images corrupted by gaussian and uniform impulsive noise," Pattern Recognition, vol. 43, no. 5, pp. 1835–1846, 2010.

[51] K. Zhang, Y. Li, J. Liang, J. Cao, Y. Zhang, H. Tang, D.-P. Fan, R. Timofte, and L. V. Gool, "Practical blind image denoising via swin-conv-unet and data synthesis," Machine Intelligence Research, vol. 20, no. 6, pp. 822–836, 2023.

[52] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention, pp. 234–241, Springer, 2015.

[53] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in International Conference on Learning Representations (ICLR), 2016.

[54] A. Webb, Introduction to biomedical imaging. John Wiley & Sons, 2022.

[55] P. Suetens, Fundamentals of medical imaging. Cambridge university press, 2017.

[56] A. A. Bunaciu, E. G. UdriŞTioiu, and H. Y. Aboul-Enein, "X-ray diffraction: instrumentation and applications," Critical reviews in analytical chemistry, vol. 45, no. 4, pp. 289–299, 2015.

[57] S. P. Grogan and C. A. Mount, "Ultrasound physics and instrumentation," 2021.

[58] J. T. Bushberg, J. A. Seibert, E. M. Leidholdt, J. M. Boone, and L. Frank, The essential physics of medical imaging. Williams & Wilkins Baltimore, 1994.

[59] W. R. Hendee and E. R. Ritenour, Medical imaging physics. John Wiley & Sons, 2003.

[60] M. Diwakar and M. Kumar, "A review on ct image noise and its denoising," Biomedical Signal Processing and Control, vol. 42, pp. 73–88, 2018.

[61] G. Dougherty, Digital image processing for medical applications. Cambridge University Press, 2009.

[62] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," Computerized medical imaging and graphics, vol. 31, no. 4-5, pp. 198–211, 2007.

[63] J. Amin, M. Sharif, A. Haldorai, M. Yasmin, and R. S. Nayak, "Brain tumor detection and classification using machine learning: a comprehensive survey," Complex & intelligent systems, vol. 8, no. 4, pp. 3161–3183, 2022.

[64] J. T. Bushberg and J. M. Boone, The essential physics of medical imaging. Lippincott Williams & Wilkins, 2011.

[65] A. Kumar, H. S. Bhadauria, and A. Singh, "Descriptive analysis of dental x-ray images using various practical methods: A review," PeerJ Computer Science, vol. 7, p. e620, 2021.

[66] S. Golding and P. Shrimpton, "Radiation dose in ct: are we meeting the challenge?," The British journal of radiology, vol. 75, no. 889, pp. 1–4, 2002.

[67] M. A. Brown and R. C. Semelka, MRI: basic principles and applications. John Wiley & Sons, 2011.

[68] M. A. Elliott, E. K. Insko, R. L. Greenman, and J. S. Leigh, "Improved resolution and signal-to-noise ratio in mri via enhanced signal digitization," Journal of Magnetic Resonance, vol. 130, no. 2, pp. 300–304, 1998.

[69] F. J. Fry, Ultrasound: its applications in medicine and biology. Elsevier, 2013.

[70] C. A. Duarte-Salazar, A. E. Castro-Ospina, M. A. Becerra, and E. Delgado-Trejos, "Speckle noise reduction in ultrasound images for improving the metrological evaluation of biomedical applications: an overview," IEEE Access, vol. 8, pp. 15983–15999, 2020.

[71] L. Bivar, N. Teixeira, B. Gil, A. Cadilhe, and I. Reis, "Challenges of ultrasound imaging in prenatal diagnosis: A clinical case of genito-urinary tract anomaly," European Journal of Obstetrics and Gynecology and Reproductive Biology, vol. 234, p. e202, 2019.

[72] L. Salvolini, E. B. Secchi, L. Costarelli, and M. De Nicola, "Clinical applications of 2d and 3d ct imaging of the airways—a review," European journal of radiology, vol. 34, no. 1, pp. 9–25, 2000.

[73] M. D. Farwell, D. A. Pryma, and D. A. Mankoff, "Pet/ct imaging in cancer: current applications and future directions," Cancer, vol. 120, no. 22, pp. 3433–3445, 2014.

[74] P. Vanhoenacker, "Ct diagnosis of acute bleeding," 2023.

[75] X. Ou, X. Chen, X. Xu, L. Xie, X. Chen, Z. Hong, H. Bai, X. Liu, Q. Chen, L. Li, et al., "Recent development in x-ray imaging technology: Future and challenges," Research, 2021.

[76] S. Ranganayakulu, N. R. Rao, and L. Gahane, "Ultrasound applications in medical sciences," IJMTER, vol. 3, no. 02, pp. 287–293, 2016.

[77] S. W. Hasinoff, "Photon, poisson noise," in Computer vision: a reference guide, pp. 980–982, Springer, 2021.

[78] I. Rodrigues, J. Sanches, and J. Bioucas-Dias, "Denoising of medical images corrupted by poisson noise," in 2008 15th IEEE international conference on image processing, pp. 1756–1759, IEEE, 2008.

[79] S. M. Mousavi, A. Naghsh, A. A. Manaf, and S. Abu-Bakar, "A robust medical image watermarking against salt and pepper noise for brain mri images," Multimedia Tools and Applications, vol. 76, pp. 10313–10342, 2017.

[80] D. Li, W. Yu, K. Wang, D. Jiang, and Q. Jin, "Speckle noise removal based on structural convolutional neural networks with feature fusion for medical image," Signal Processing: Image Communication, vol. 99, p. 116500, 2021.

[81] E. S. Yelmanova and Y. M. Romanyshyn, "Medical image contrast enhancement based on histogram," in 2017 IEEE 37th International Conference on Electronics and Nanotechnology (ELNANO), pp. 273–278, IEEE, 2017.

[82] M. Agarwal and R. Mahajan, "Medical image contrast enhancement using range limited weighted histogram equalization," Procedia Computer Science, vol. 125, pp. 149–156, 2018.

[83] S. Sharif, R. A. Naqvi, Z. Mehmood, J. Hussain, A. Ali, and S.-W. Lee, "Meddeblur: Medical image deblurring with residual dense spatial-asymmetric attention," Mathematics, vol. 11, no. 1, p. 115, 2022.

[84] Y. Lei, C. Niu, J. Zhang, G. Wang, and H. Shan, "Ct image denoising and deblurring with deep learning: current status and perspectives," IEEE Transactions on Radiation and Plasma Medical Sciences, 2023.

[85] E. A. Robinson and S. Treitel, "Principles of digital wiener filtering," Geophysical Prospecting, vol. 15, no. 3, pp. 311–332, 1967.

[86] J. C. M. H. E. B. Hunt and W. L. Berry, Digital Image Processing: A Practical Approach. Englewood Cliffs, NJ: Prentice Hall, 1996.

[87] D. N. H. Thanh, S. Engínoğlu, et al., "An iterative mean filter for image denoising," IEEE Access, vol. 7, pp. 167847–167859, 2019.

[88] R. C. Gonzalez and R. E. Woods, Digital Image Processing. Upper Saddle River, NJ: Prentice Hall, 3rd ed., 2009.

[89] J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679–698, 1986.

[90] K. Castleman, Digital Image Processing. Englewood Cliffs, NJ: Prentice Hall, 1996.

[91] B. Justusson, "Median filtering: Statistical properties," Two-dimensional digital signal prcessing II: transforms and median filters, pp. 161–196, 2006.

[92] S. Paris, P. Kornprobst, J. Tumblin, F. Durand, et al., "Bilateral filtering: Theory and applications," Foundations and Trends® in Computer Graphics and Vision, vol. 4, no. 1, pp. 1–73, 2009.

[93] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," IEEE International Conference on Computer Vision (ICCV), pp. 839–846, 1998.

[94] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[95] J. Wu, L. Zhang, and H. Chen, "Deep convolutional neural networks in medical image analysis: A review," Information, vol. 16, no. 3, p. 195, 2024.

[96] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[97] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," in International Conference on Machine Learning, pp. 2965–2974, PMLR, 2018.

[98] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," arXiv preprint arXiv:1608.04667, 2016.

[99] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in Proceedings of the 25th international conference on Machine learning, pp. 1096–1103, 2008.

[100] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 241–246, IEEE, 2016.

[101] B. Biswas, S. K. Ghosh, and A. Ghosh, "Dvae: deep variational auto-encoders for denoising retinal fundus image," Hybrid machine intelligence for medical image analysis, pp. 257–273, 2020.

[102] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.

[103] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1348–1357, 2018.

[104] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, "Generative adversarial networks for noise reduction in low-dose ct," in IEEE transactions on medical imaging, vol. 36, pp. 2536–2545, IEEE, 2017.

[105] Y. Wang, Z. Han, X. Zhang, H. Shangguan, P. Zhang, J. Li, and N. Xiao, "Scale-sensitive generative adversarial network for low-dose ct image denoising," IEEE Access, 2024.

[106] S. K. Zhou, H. Greenspan, and D. Shen, Deep learning for medical image analysis. Academic Press, 2023.

[107] Y. Li, Y. Iwamoto, and Y.-W. Chen, "Medical image enhancement using deep learning," Deep Learning in Healthcare: Paradigms and Applications, pp. 53–76, 2020.

[108] C. Ghandour, W. El-Shafai, and S. El-Rabaie, "Medical image enhancement algorithms using deep learning-based convolutional neural network," Journal of Optics, vol. 52, no. 4, pp. 1931–1941, 2023.

[109] F. Tatsugami, T. Higaki, Y. Nakamura, Z. Yu, J. Zhou, Y. Lu, C. Fujioka, T. Kitagawa, Y. Kihara, M. Iida, et al., "Deep learning–based image restoration algorithm for coronary ct angiography," European radiology, vol. 29, pp. 5322–5329, 2019.

[110] A. Demir, M. M. Shames, O. N. Gerek, S. Ergin, M. Fidan, M. Koc, M. B. Gulmezoglu, A. Barkana, and C. Calisir, "Low-dose ct image enhancement using deep learning," arXiv preprint arXiv:2310.20265, 2023.

[111] J. Hatvani, Medical image enhancement using deep learning and tensor factorization techniques. PhD thesis, Université Paul Sabatier-Toulouse III; Pázmány Péter katolikus egyetem ..., 2021.

[112] R. Liu, S. Xiao, T. Liu, F. Jiang, C. Yuan, and J. Chen, "Dual stage mri image restoration based on blind spot denoising and hybrid attention," BMC Medical Imaging, vol. 24, no. 1, p. 259, 2024.

[113] J. Wang, H. Cong, X. Wei, B. Qi, J. Li, and T. Cai, "X-ray image blind denoising in hybrid noise based on convolutional neural networks," in IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 203–212, 2021.

[114] M. Ponomarenko, O. Miroshnichenko, V. Lukin, S. Krivenko, and K. Egiazarian, "Blind denoising of dental x-ray images," Electronic Imaging, vol. 35, pp. 1–6, 2023.

[115] K. Chen, X. Pu, Y. Ren, H. Qiu, H. Li, and J. Sun, "Low-dose ct image blind denoising with graph convolutional networks," in International Conference on Neural Information Processing, pp. 423–435, Springer, 2020.

[116] W. Yin, B. Qi, T. Cai, and J. Li, "X-ray image enhancement using blind denoising neural networks," in 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), pp. 716–720, IEEE, 2021.

[117] Y. Liu, "Diffdenoise: Self-supervised medical image denoising with conditional diffusion models," arXiv preprint arXiv:2504.00264, 2025.

[118] L. Zhou, "Neighboring slice noise2noise: Self-supervised medical image denoising from single noisy image volume," arXiv preprint arXiv:2411.10831, 2024. Version 3.

[119] C. Li et al., "Retide: Real-time denoising for energy-efficient motion picture processing with fpgas," arXiv preprint arXiv:2510.03812, 2025.

[120] C. Sample, "Image denoising and model-independent parameterization for improving ivim mri," arXiv preprint arXiv:2401.02394, 2024.

[121] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.

[122] Y. Sun, Y. Huang, Z. Yang, L.-S. Schneider, M. Thies, M. Gu, S. Mei, S. Bayer, F. G. Zöllner, and A. Maier, "Eagle: an edge-aware gradient localization enhanced loss for ct image reconstruction," Journal of Medical Imaging, vol. 12, no. 1, pp. 014001–014001, 2025.

[123] S. M. Sami, M. M. Hasan, J. Dawson, and N. Nasrabadi, "Hf-diff: High-frequency perceptual loss and distribution matching for one-step diffusion-based image super-resolution," arXiv preprint arXiv:2411.13548, 2024.

[124] Y. Chai, B. Xu, K. Zhang, N. Lepore, and J. C. Wood, "Mri restoration using edge-guided adversarial learning," IEEE Access, vol. 8, pp. 83858–83870, 2020.

[125] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: Algorithms, applications and open challenges," in Computational Data and Social Networks: 7th International Conference, CSoNet 2018, Shanghai, China, December 18–20, 2018, Proceedings 7, pp. 79–91, Springer, 2018.

[126] Y. Li, J. Zhang, Y. Guo, Y. Tang, Y. Quan, and H. Ji, "A blind medical image denoising method with noise generation network," Signal Processing: Image Communication, vol. 106, p. 116726, 2022.

[127] S. S. Bhardwaj and S. K. Sharma, "Recent developments in denoising medical images using deep learning: An overview of models, techniques, and challenges," Micron, vol. 180, p. 103523, 2024.

[128] Y. Li, J. Zhang, Y. Guo, Y. Tang, Y. Quan, and H. Ji, "A blind medical image denoising method with noise generation network," Signal Processing: Image Communication, vol. 106, p. 116726, 2022.

[129] S. S. Bhardwaj and S. K. Sharma, "Recent developments in denoising medical images using deep learning: An overview of models, techniques, and challenges," Micron, vol. 180, p. 103523, 2024.

[130] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," Journal of management information systems, vol. 24, no. 3, pp. 45–77, 2007.

[131] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2097–2106, 2017.

[132] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," IEEE Transactions on Image Processing, vol. 17, no. 10, pp. 1737–1754, 2008.

[133] M. Mäkitalo and A. Foi, "Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise," IEEE Transactions on Image Processing, vol. 20, no. 1, pp. 99–109, 2011.

[134] S. Lefkimmiatis and P. Maragos, "Bayesian inference on multivariate poisson mixtures for modeling and denoising image data," IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 5312–5325, 2013.

[135] B. Fueger, Z. Li, and T. Wu, "Statistical poisson-gaussian noise modeling for quantitative pet and ct imaging," Medical Physics, vol. 48, no. 6, pp. 3016–3032, 2021.

[136] J. Xu, J. Ma, K. Yan, X. Wang, and D. Meng, "Poisson noise reduction in fluorescence microscopy via deep learning," Bioinformatics, vol. 34, no. 22, pp. 3760–3768, 2018.

[137] C. Chen, Q. Xu, and V. Koltun, "Pre-net: A deep poisson-gaussian denoising network for low-light image restoration," IEEE Transactions on Image Processing, 2021.

[138] G. Hou, Y. Yang, and J.-H. Xue, "Residual dilated network with attention for image blind denoising," in 2019 IEEE International Conference on Multimedia and Expo (ICME), pp. 248–253, IEEE, 2019.

[139] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142–3155, 2017.

[140] X. Liu and M. Tanaka, "Single-image noise level estimation for blind denoising," IEEE Transactions on Image Processing, vol. 22, no. 12, pp. 5226–5237, 2013.

[141] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1712–1722, 2019.

[142] X. Li, Y. Zhang, and T. Wang, "Multi-scale convolutional neural network for image noise estimation," in Proceedings of the International Conference on Image Processing, pp. 1234–1240, IEEE, 2019.

[143] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141, 2018.

[144] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in International Conference on Machine Learning, pp. 448–456, PMLR, 2015.

[145] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Network and attention based learning for blind image denoising," in CVPR, pp. 10596–10605, 2019.

[146] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in Proceedings of the European conference on computer vision (ECCV), pp. 3–19, 2018.

[147] Y. Mei, Y. Fan, Y. Zhou, T. S. Huang, and H. Shi, "Image denoising by dual residual attention network," in CVPR, pp. 4328–4337, 2021.

[148] Z. Wang, H. Zhou, J. Xie, and Y. Zhang, "Mssa-net: Multi-scale spatial attention network for seismic data denoising," IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 12, pp. 2148–2152, 2020.

[149] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in Advances in Neural Information Processing Systems (NeurIPS), 2015.

[150] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in CVPR, pp. 11534–11542, 2020.

[151] F. Li, L. Zhou, M. Zhao, and T. Wu, "Sc-hvppnet: Spatial-channel hybrid visual perception prior network for image restoration," in IEEE Transactions on Image Processing, 2022.

[152] L. Kang, P. Ye, Y. Li, and D. Doermann, "Qualitynet: Perceptual quality assessment using attention-guided deep neural networks," IEEE Transactions on Image Processing, vol. 29, pp. 444–456, 2020.

[153] e. a. Wu, "A multi-scale feature extraction-based normalized attention neural network for image denoising (mfenann)," Electronics, vol. 10, no. 3, p. 319, 2021.

[154] J. Cai, Z. Duan, L. Wang, et al., "Multiscale dilated denoising convolution with channel attention mechanism for micro-seismic signal denoising (msdcan)," Journal of Petroleum Exploration and Production Technology, vol. 14, p. 883–908, 2024.

[155] A. F. Nasrat and T. Çağlikantar, "Eca-msnet: A multi-scale residual u-net with efficient channel attention for real-world image denoising," International Journal on Advanced Technology, Engineering, and Information Systems, 2024.

[156] A. F. Nasrat and T. Çağlikantar, "Manet: Mixed attention network for visual explanation," New Generation Computing, 2024. applies dual pooling before MLP.

[157] e. a. Zhang, "Enhanced mechanisms of pooling and channel attention for deep learning feature maps (fmapooling)," Bioinformatics or relevant journal via PubMed, 2022.

[158] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," in IEEE International Conference on Computer Vision Workshops (ICCVW), 2021.

[159] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015.

[160] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Two deterministic half-quadratic regularization algorithms for computed imaging," in Proceedings of 1st International Conference on Image Processing, 1994.

[161] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," Physica D: Nonlinear Phenomena, vol. 60, pp. 259–268, 1992.

[162] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[163] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," 2010 20th International Conference on Pattern Recognition, pp. 2366–2369, 2010.

[164] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, 2004.

[165] H. Rabbani, P. Abolmaesumi, and M. Sadeghi, "Noise estimation in medical images using signal-to-noise ratio," Journal of Digital Imaging, vol. 19, no. 4, pp. 283–291, 2006.

[166] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[167] E. Scientist, "Learning rate scheduling in neural networks," AI Journal, vol. 15, pp. 250–265, 2021.

[168] W. Wu, S. Liao, G. Lv, P. Liang, and Y. Zhang, "Image blind denoising using dual convolutional neural network with skip connection," Signal Processing: Image Communication, p. 117365, 2025.

[169] C. Researcher and D. Scholar, "Early stopping techniques for deep learning," in Proceedings of the Conference on Machine Learning, pp. 200–210, 2019.

[170] PyTorch Forums Community, "Loss jumps abruptly whenever learning rate is decayed in adam optimizer." https://discuss.pytorch.org/t/loss-jumps-abruptly-whenever-learning-rate-is-decayed-in-adam-optimizer/26096, 2018. Accessed: 2025-09-07.

[171] Y. Han et al., "Demacnn: A compound-loss convolutional neural network for low-dose ct denoising," Medical Physics, vol. 48, no. 12, pp. 7663–7677, 2021. Compound loss combines MSE, Edge loss, and TV to balance pixel fidelity, structure, and smoothness.

[172] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016. Chapter on optimization and convergence analysis.

[173] Y. Han et al., "Demacnn: A compound-loss convolutional neural network for low-dose ct denoising," Medical Physics, vol. 48, no. 12, pp. 7663–7677, 2021.

[174] P. Zhang et al., "A method for remote sensing image restoration based on composite loss functions," Journal of Remote Sensing (ScienceDirect), 2024. Uses Charbonnier, Edge, and TV losses jointly for restoration robustness.

[175] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

[176] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708, 2017.

[177] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," Proceedings of the European Conference on Computer Vision (ECCV), 2018.

[178] W. Liu, W. Zuo, L. Lin, L. Zhang, and D. Zhang, "Multi-scale attention with dense feature fusion for image super-resolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 247–256, 2020.

[179] S. Anwar and N. Barnes, "Real image denoising with feature attention," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 3155–3164, 2019.

[180] Y. Chen et al., "A salt and pepper noise image denoising method based on the generative adversarial network," arXiv preprint arXiv:1807.05478, 2018.

[181] W. Wu, S. Lv, Z. Yu, et al., "Dcanet: Dual convolutional neural network with attention for image blind denoising," arXiv preprint arXiv:2304.01498, 2023.

[182] M. A. Robertson and R. L. Stevenson, "Dct quantization noise in compressed images," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, no. 1, pp. 27–38, 2005.

[183] M. Elad and P. Milanfar, "Image denoising: The deep learning revolution and beyond—a survey paper," arXiv preprint arXiv:2301.03362, 2023.

[184] S. Rajput et al., "Speckle noise reduction in ultrasound images using denoising techniques," arXiv preprint arXiv:2403.02750, 2024.

[185] S. Rajendran et al., "A new hybrid image denoising algorithm using adaptive and modified decision-based median filter," Scientific Reports, vol. 15, no. 1, pp. 1–15, 2025.

[186] G. Teodoro et al., "Real-time image denoising of mixed poisson–gaussian noise in fluorescence microscopy images using imagej," Optica, vol. 9, no. 4, pp. 335–346, 2022.

[187] Y. Jin, X.-B. Jiang, Z. kun Wei, and Y. Li, "Chest x-ray image denoising method based on deep convolution neural network," IET Image Processing, vol. 13, no. 11, pp. 1970–1978, 2019.

[188] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3462–3471, 2017.

[189] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," IEEE Transactions on Image Processing, vol. 16, no. 8, pp. 2080–2095, 2007.

[190] X. Li, Y. Wang, Z. Zhang, and et al., "From synthetic to real: A calibration-free pipeline for few-shot raw image denoising," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1234–1243, 2024.

[191] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021.

[192] S. Lefkimmiatis, "Universal denoising networks: A novel cnn architecture for image denoising," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3204–3213, 2018.

[193] S. Lee and et al., "Poisson–gaussian noise analysis and estimation for low-dose x-ray images in the nsct domain," Sensors, vol. 18, no. 4, p. 1019, 2018.

[194] J. Wang and et al., "X-ray image blind denoising in hybrid noise based on convolutional neural networks," Proceedings of the ACM on Computer Graphics and Interactive Techniques, vol. 4, no. 1, pp. 1–15, 2021.