HSLU Hochschule Luzern

LUCERNE UNIVERSITY OF APPLIED SCIENCES AND ARTS

BACHELOR THESIS PROPOSAL

Effectiveness of deep learning architectures for metal artifact reduction (MAR) in CT imaging

---

Author:
Marcos COSTA

Supervisor:
Javier Montoya

A thesis submitted in fulfillment of the requirements
for the degree of Bachelor in International IT Management

May 2, 2025

Declaration of Authorship

I, Marcos COSTA, declare that this thesis titled, "Effectiveness of deep learning architectures for metal artifact reduction (MAR) in CT imaging" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a bachelor degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

Acknowledgements

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

LUCERNE UNIVERSITY OF APPLIED SCIENCES AND ARTS

Abstract

Computer Science and Information Technology

Bachelor in International IT Management

Effectiveness of deep learning architectures for metal artifact reduction (MAR) in CT imaging

by Marcos COSTA

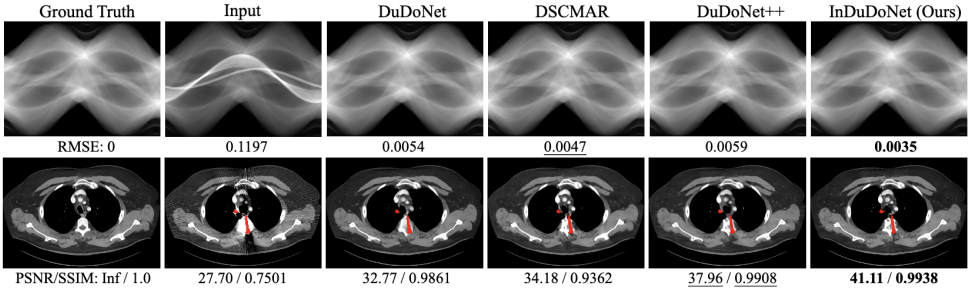


FIGURE 1: Comparison of various dual-domain enhancement-based MAR methods. Image from (H. Wang et al., 2021).

Metal artifacts in computed tomography (CT) imaging significantly impact diagnostic accuracy and treatment planning, especially in patients with metallic implants such as orthopedic screws or prostheses. This thesis will investigate advanced computational methods for metal artifact reduction (MAR), exploring supervised and unsupervised techniques to improve image quality. Successful validation of these methods could surpass traditional MAR approaches by effectively reducing streak artifacts and preserving anatomical details. Achieving these improvements would meaningfully enhance diagnostic precision and radiotherapy planning efficacy.

Keywords: Metal Artifact Reduction (MAR), Computed Tomography (CT), Deep Learning, Medical Imaging, Image Reconstruction, Imaging Geometry, Clinical Diagnosis, Unsupervised Learning, Supervised Learning, Semi-Supervised Learning, Neural Networks

Contents

List of Figures

List of Tables

List of Abbreviations

BN	**B**atch **N**ormalization
BHC	**B**eam **H**ardening **C**orrection
CNN	**C**onvolutional **N**eural **N**etwork
CRF	**C**onditional **R**andom **F**ield
CT	**C**omputed **T**omography
CV	**C**omputer **V**ision
DL	**D**eep **L**earning
FBP	**F**iltered **B**ack **P**rojection
FCN	**F**ully **Co**nvolutional **N**etwork
FOV	**F**ield **O**f **V**iew
GAN	**G**enerative **A**dversarial **N**etwork
GELU	**G**aussian **E**rror **L**inear **U**nit
GlaS	**Gla**nd **S**egmentation
GP	**G**radient **P**enalty
H&E	**H**ematoxylin and **E**osin (staining)
HU	**H**ounsfield **V**alue
IN	**I**nstance **N**orm
LN	**L**ayer**n**orm
MAR	**M**etal **A**rtifact **R**eduction
MLP	**M**ulti **L**ayer **P**erceptron
MSA	**M**ultihead **S**elf **A**ttention
NLP	**N**atural **L**anguage **P**rocessing
ReLU	**R**ectified **L**inear **U**nit
ResNet	**R**esidual **N**eural **N**etwork
ROI	**R**egion **o**f **I**nterest
SA	**S**elf **A**ttention
SGD	**S**tochastic **G**radient **D**escent
SN	**S**pectral **N**ormalization
VAE	**V**ariational **A**uto**e**ncoder
VGG	**V**isual **G**eometry **G**roup

Chapter 1

Introduction

1.1 Motivation and Context

Computed tomography (CT) is a cornerstone of modern medical imaging, offering non-invasive, high-resolution visualization of both hard and soft tissues. Its applications are critical for accurate disease diagnosis, treatment planning, and surgical navigation. However, the presence of metallic implants—such as orthopedic screws and prosthetics—can significantly degrade CT image quality. Metal-induced artifacts, including streaks, dark bands, and shadowing effects, obscure vital anatomical details, thereby compromising diagnostic accuracy and increasing the risk of medical errors in treatment.

Traditional metal artifact reduction (MAR) techniques, such as dual-energy CT and iterative reconstruction algorithms, have provided some mitigation of these artifacts. Despite their utility, these methods often fall short in fully eliminating artifacts without introducing secondary issues, such as loss of detail or new distortions (Selles et al., 2023). The complex, non-linear nature of metal-induced artifacts poses significant challenges that traditional algorithms are incapable of handling.

Recent advancements in artificial intelligence, particularly deep learning, offer promising new avenues for MAR. Deep learning models can learn and replicate the intricate patterns of metal artifacts, enabling more effective artifact reduction while preserving anatomical fidelity (Yu, Zhang, Li, Ren, et al., 2021). Various approaches, including supervised, unsupervised, and dual-domain models, have demonstrated potential in both laboratory and clinical settings. Despite these advancements, there remains a lack of systematic evaluations comparing the efficacy of different deep learning-based MAR methods across diverse clinical scenarios.

1.2 Problem Statement

Metal artifacts in CT scans, typically arising from metallic implants like surgical screws or prosthetics, introduce significant visual distortions such as streaks, dark shadows, and noise. These artifacts impede the accurate visualization of anatomical structures, leading to potential diagnostic errors and complications in medical procedures. Traditional MAR methods—including filtered back projection, iterative reconstruction, and sinogram inpainting—often fail to fully address the complexity of metal-induced artifacts. They are limited

by incomplete artifact removal, introduction of secondary artifacts, and inadequate handling of the non-linear properties of metal attenuation in CT imaging (Abdoli et al., 2012).

While deep learning-based MAR models have shown considerable promise in mitigating these challenges by effectively addressing the non-linear and non-local characteristics of metal artifacts, the field lacks comprehensive, systematic evaluations of these models' performance across various clinical contexts. Different deep learning approaches—ranging from supervised to unsupervised methods and single-domain to dual-domain models—exhibit varying levels of effectiveness. This creates uncertainty for clinicians and researchers in selecting the most appropriate MAR technique for specific clinical applications.

This study addresses the question: How do different deep learning models compare in their effectiveness for metal artifact reduction (MAR) in CT imaging? By systematically evaluating a selection of state-of-the-art deep learning models, this research seeks to identify strengths and limitations across methods, and ultimately contribute to the development of improved MAR solutions in medical imaging.

1.3 Objectives

The main objective of this research is to evaluate and compare the effectiveness of various deep learning models for metal artifact reduction (MAR) in CT imaging, aiming to identify the strengths and limitations of each model for clinical application. By systematically analyzing different MAR approaches, this study seeks to provide clear guidance for selecting appropriate MAR models to improve image quality in cases where metallic implants are present.

1.3.1 General Objectives

To evaluate and compare deep learning based MAR techniques for reducing metal artifacts in CT imaging so that CT scans with metallic implants are more diagnostic and clinical useful.

1.3.2 Specific Objectives

- Objective 1: Identify and review the types of deep learning models used for metal artifact reduction, supervised, unsupervised, image domain and dual domain.
- Objective 2: Define the evaluation metrics for comparing MAR model performance, artifact reduction effectiveness, image quality preservation.
- Objective 3: Compare selected MAR models using clinical and synthetic datasets that simulate different types and intensities of metal artifacts.
- Objective 4: Analyze the pros and cons of each model, so that we know the clinical context and artifact type where each model works best.
- Objective 5: Provide guidance to clinicians and researchers on which MAR technique to use based on clinical needs and artifact characteristics, so that we get better CT imaging and diagnosis.

Chapter 2

Background

2.1 Computed Tomography

Computed Tomography (CT) is a common imaging modality that creates cross sectional images of the body by processing X-ray projections taken from many angles around the patient. This is useful for diagnosing everything from bone fractures to complex vascular diseases because it provides high resolution anatomical detail. The basic principle of CT imaging is the measurement of X-ray attenuation as the beams pass through differently reconstructed into 2D slices or 3D images of the scanned area.

An important part of the CT scanning process is an X-ray emitter and detector array. The emitter rotates around the patient, emitting X-rays at different angles continuously while the detectors on the other side capture the attenuated X-rays that pass through the body. As the X-rays pass through different tissues their intensity decreases depending on the tissue density and composition. This attenuation data forms a series of 1D projections, which are combined to create a sinogram—a raw data representation that holds all the information required to reconstruct an image of the scanned area.

2.1.1 Hounsfield Units (HU) and Tissue Radiodensity

A unique feature of CT imaging is its use of Hounsfield Units (HU), a quantitative scale that measures the radiodensity of different tissues. The HU scale is based on the linear attenuation coefficients of materials, providing a standardized way to interpret CT images. Key reference points on the HU scale include:

- Water: Assigned a baseline value of 0 HU.
- Air: Approximately -1000 HU, indicating very low attenuation of X-rays.
- Cortical Bone: Around +1000 HU or higher, due to its high density and substantial X-ray attenuation.
- Muscle Tissue: Typically between +30 and +60 HU, while other soft tissues, such as the liver, kidneys, and blood vessels, have characteristic HU ranges reflecting their densities.

This range of HU values enables clinicians to distinguish between tissue types and detect abnormalities. For instance, fat has relatively low HU values (-50 to -100 HU), while calcified structures or bones have high HU values, enhancing diagnostic clarity and precision.

2.1.2 CT Image Creation Process

The CT image reconstruction process involves several key steps:

1. Data Acquisition: As the emitter and detector array rotate around the patient, X-ray projections are collected from multiple angles. This multi-angle data acquisition is essential to capture the density variations across different tissues and build a complete dataset of the scanned region.

2. Signal Processing: The detected signals are processed to calculate the degree of X-ray attenuation along each projection path, yielding critical information on tissue density and composition. This step is central to generating accurate HU values for each voxel in the final image.

3. Sinogram Formation: The processed data from each projection is organized into a sinogram. The sinogram is a representation of how the X-rays have been attenuated across all angles, containing the raw information needed to reconstruct the final cross-sectional image.

4. Image Reconstruction Using Filtered Back Projection (FBP): The sinogram data is then used to reconstruct a cross-sectional image through a technique called back projection. In this step, each projection is "smeared" or "projected back" across the image plane, allowing the information from all angles to merge into a cohesive image. To reduce blurring and improve image clarity, the data undergoes a filtering step before back projection, hence the term Filtered Back Projection (FBP). FBP is the most widely used reconstruction technique in clinical CT imaging because it is fast and straightforward, allowing for efficient processing while preserving high image quality (Schofield et al., 2020).

2.1.3 Challenges with Artifact in CT Imaging

Despite CT's advantages, the technology is susceptible to artifacts—distortions that can degrade image quality. Artifacts in CT scans arise due to several factors:

- Patient Movement: Movement during scanning can cause blurring or streaks in the reconstructed image, as the alignment of projections may be compromised.

- Metallic Implants: Objects such as screws, or prosthetics produce intense artifacts due to their high density and atomic number, causing streaks, shadows, and distortions in the image.

- Limitations in Reconstruction Algorithms: Some algorithms may struggle with highly attenuated or incomplete data, leading to inaccuracies in the reconstructed image.

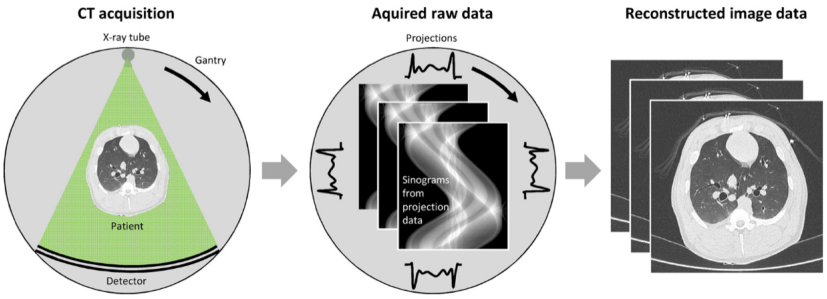


FIGURE 2.1: Simplified schematic of a typical CT scanning process, detailing the steps from data acquisition to image reconstruction using filtered back projection (FBP) in spiral computed tomography (CT). Image from (Stiller, 2018).

Artifacts from metallic objects are particularly problematic because they introduce streaks, shadows, and distortions that obscure surrounding tissues. These metal-induced artifacts arise from multiple physical phenomena, including beam hardening, photon starvation, and scatter. These issues impact the accuracy of Hounsfield Unit (HU) values near metallic implants, complicating diagnosis and treatment planning. Addressing these artifacts is crucial for accurate imaging, especially in clinically relevant areas where detailed anatomical information is essential for accurate diagnostic.

2.2 Metal Artifacts

Metal objects have significantly higher attenuation properties compared to human tissues, particularly soft tissues. Consequently, they can produce artifacts in CT images through phenomena such as beam hardening, scattering, and photon starvation (Boas & Fleischmann, 2012).

2.2.1 Beam Hardening

Beam hardening happens in medical computed tomography (CT) when polychromatic X-ray beams are utilized. In this context, lower-energy photons are attenuated more than higher-energy photons, particularly in the presence of dense objects within the field of view (FOV). The attenuation of low-energy photons is predominantly due to the photoelectric effect.

The photoelectric effect is especially pronounced in materials with high atomic numbers (Z), such as metals, and with low-energy photons. The probability of the photoelectric effect occurring is proportional to $\frac{Z^3}{E^3}$, where E represents the photon energy (Vellarackal & Kaim, 2021). This effect acts similarly to a high-pass filter, resulting in an increase in the mean beam energy—a process referred to as "beam hardening." Locally, this phenomenon causes the beam to behave as though it has traversed a low-attenuating object, which manifests as

dark streaks in the reconstructed CT image. These dark streaks predominantly appear along the axis of greatest attenuation, typically where the object is longest within the 2D slice or situated between two metal objects (see Figure 2.2).

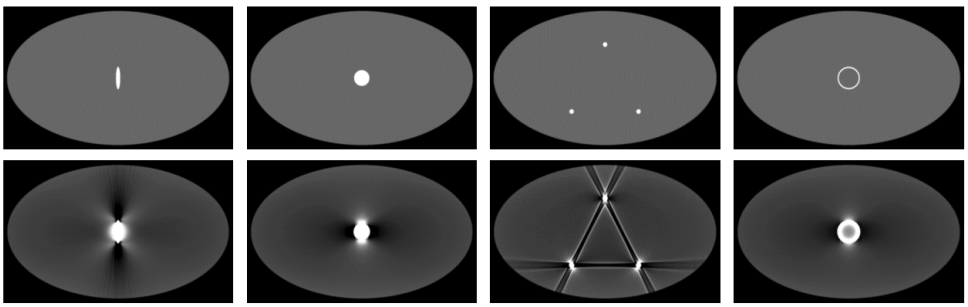


FIGURE 2.2: Simulated scans without (top row) and with (bottom row) beam hardening demonstrate that dark streaks appear along areas of highest attenuation, while bright streaks emerge in other directions. Similar artifacts are also caused by Compton scatter. Image from (Boas & Fleischmann, 2012).

2.2.2 Compton Scattering

At higher energies, Compton scattering becomes the dominant interaction mechanism in CT imaging because it varies inversely $\frac{1}{E}$, rather then $\frac{1}{E^3}$. During Compton scattering, photons are deflected from their original trajectories within the patient's body. These scattered photons may then strike detectors that are not aligned with the central axis of the incident beam, leading to inaccuracies in image reconstruction. When combined with the strong photoelectric effect, Compton scattering contributes to the formation of dark streaks in the reconstructed image, as both phenomena result in an increased number of detected photons that are erroneously attributed to different paths.

2.2.3 Phantom Starvation and Noise

Phantom starvation and noise arise when a significant portion of photons are absorbed by highly attenuating objects within the FOV, resulting in an insufficient number of photons reaching the detectors. This insufficiency leads to elevated statistical errors in the projection data, a phenomenon known as phantom starvation. Consequently, phantom starvation is responsible for the appearance of thin dark and bright streaks in the reconstructed CT images.

2.3 Deep Learning

In recent years, deep learning has revolutionized artificial intelligence, enabling significant advancements across domains such as computer vision, natural language processing, and,

notably, medical imaging. Deep learning models are based on neural networks with many layers, allowing them to learn complex, hierarchical patterns in data. Unlike traditional machine learning, which often requires hand-engineered features, deep learning networks learn features directly from raw data through a process known as feature extraction. This ability to learn from unstructured data, such as images and text, makes deep learning especially powerful for tasks in medical imaging, including classification, segmentation, and artifact reduction.

Deep learning's layered architecture is inspired by the structure and function of the human brain. At its core, a deep neural network is composed of layers of neurons, each of which processes information and identifies patterns of increasing complexity. For metal artifact reduction (MAR) in CT imaging, deep learning enables models to recognize and correct the non-linear and non-local artifact patterns caused by metal implants. Key architectures in deep learning for MAR include Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), each providing distinct advantages for artifact correction.

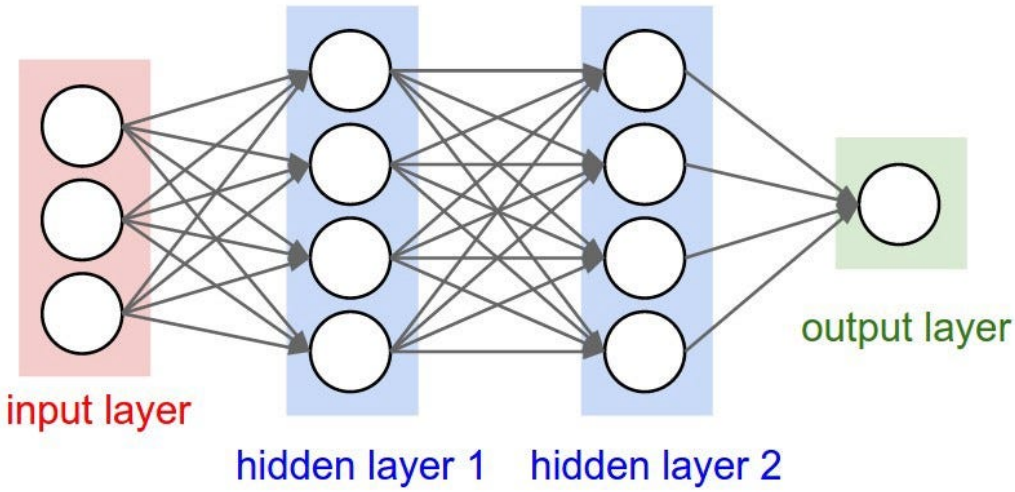


FIGURE 2.3: Example of a neural network with an input layer, 2 hidden layers of 4 neurons each and an output layer. Data moves through the network from left to right.

2.3.1 Core Components of Deep Learning Models

Deep learning models consist of several key components that work together to process data, learn patterns, and make predictions. Understanding these components is important for grasping how different architectures, like Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), perform their specific tasks. The input layer is the starting point where raw data, such as images or text, is received and prepared for processing. The hidden layers, composed of various types such as dense (fully connected), convolutional, recurrent, or transformer layers, extract hierarchical features through successive transformations. These layers rely on weights and biases that are adjusted during training to minimize errors. The weights and biases are fundamental to the network and is what allows the networks to "learn". During a forward pass, the process by which a neuron's output is

calculated is as follows:

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right)$$

where x_i represents the input features, w_i are the weights, b is the bias, and f is the activation function. This equation defines how each neuron transforms its inputs, which is fundamental to understanding how deep learning models operate.

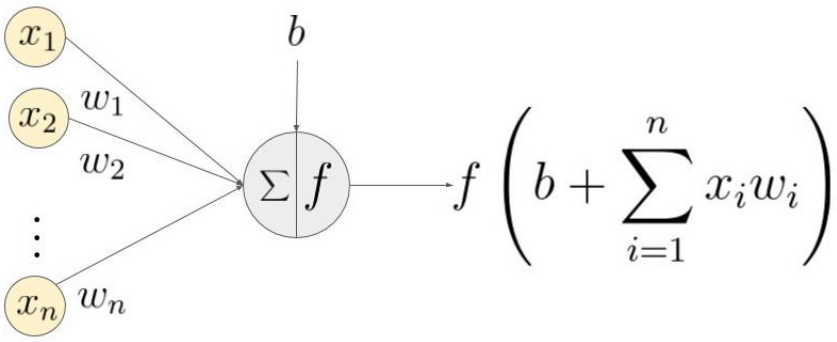


FIGURE 2.4: Python visualization of the elements that go into the computation for a single neuron (Gupta, 2017).

Activation functions play a critical role in deep learning models by introducing non-linearity, which allows the network to learn and represent complex patterns. Common activation functions include:

- ReLU (Rectified Linear Unit): Is one of the most widely used activation functions in deep learning due to its simplicity and efficiency. It outputs the input directly if it is positive and zero otherwise, effectively introducing sparsity into the network. It's formula $g(z) = \max(0, z)$ means that for inputs $z > 0$, the function returns z, and for $z <= 0$, it returns 0. This simplicity reduces computational overhead and reduces the vanishing gradient problem.
- Sigmoid: The sigmoid activation function maps input values to a range between 0 and 1, making it suitable for probabilistic interpretations in binary classification tasks. It's formula $g(z) = \frac{1}{1+e^{-z}}$ means that as $z \to +\infty$, $g(z) \to 1$; as $z \to -\infty$, $g(z) \to 0$. Sigmoid can suffer from the vanishing gradient problem for large or small inputs.
- Tanh (Hyperbolic Tangent): Similar to sigmoid, the tanh function maps input values to a range between -1 and 1, which centers the output around zero, making it often preferable to sigmoid in practice. It's formula

  $$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

  means that for large positive x, $f(x) \to 1$, and for large negative x, $f(x) \to -1$. Tanh addresses some of the issues with sigmoid but still suffers from the vanishing gradient problem.

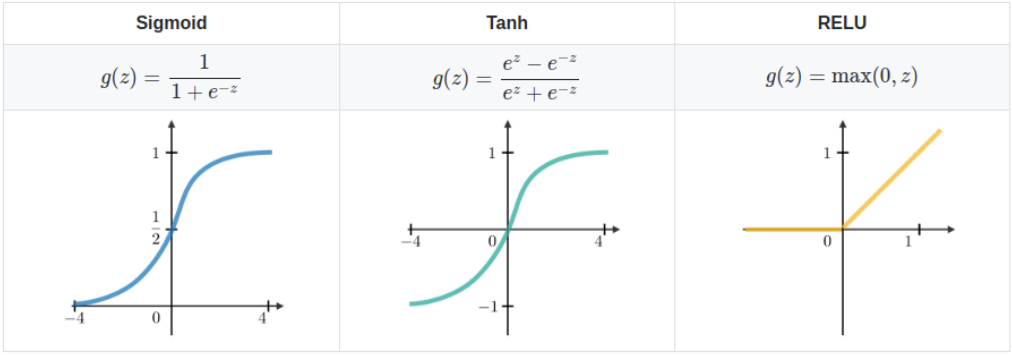


FIGURE 2.5: Visualisation of the Sigmoid, Tanh and ReLU activation functions used in neural networks.

Each activation function has strengths and weaknesses, and the choice depends on the specific task and network architecture. For instance, ReLU is popular in hidden layers, while sigmoid is commonly used in output layers for binary and multi-class classification tasks.

2.4 Optimization Techniques

Optimization plays a crucial role in training deep learning models, as it determines how effectively a model can minimize its error and improve performance. The process involves adjusting the model's parameters (weights and biases) to minimize a loss function, which measures the difference between the predicted outputs and the ground truth.

2.4.1 Gradient Descent

The gradient descent formula is a widely used optimization algorithm in deep learning that minimizes a loss function $L(\mathbf{w})$ by iteratively updating the model's parameters $\mathbf{w}$ in the direction of the negative gradient which reduces the loss. The general formula is given by:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla_{\mathbf{w}} L(\mathbf{w}_t),$$

where $\mathbf{w}_t$ represents the parameter vector at iteration t, $\mathbf{w}_{t+1}$ is the updated parameter vector for the next iteration, η is the learning rate controlling the step size, and $\nabla_{\mathbf{w}} L(\mathbf{w}_t)$ denotes the gradient of the loss function with respect to $\mathbf{w}_t$. The gradient is a vector of partial derivatives:

$$\nabla_{\mathbf{w}} L(\mathbf{w}) = \left[\frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, \ldots, \frac{\partial L}{\partial w_n} \right]^T.$$

The algorithm aims to minimize $L(\mathbf{w})$, typically representing the error or cost of the model. The learning rate η plays a critical role; if it is too large, updates may overshoot the optimal solution, while a small learning rate can result in slow convergence.

Variants of gradient descent include batch gradient descent, which computes the gradient over the entire dataset, stochastic gradient descent (SGD), which uses a single data point for updates, and mini-batch gradient descent, which updates parameters using a small subset of the data, balancing efficiency and stability. For instance, for a simple quadratic

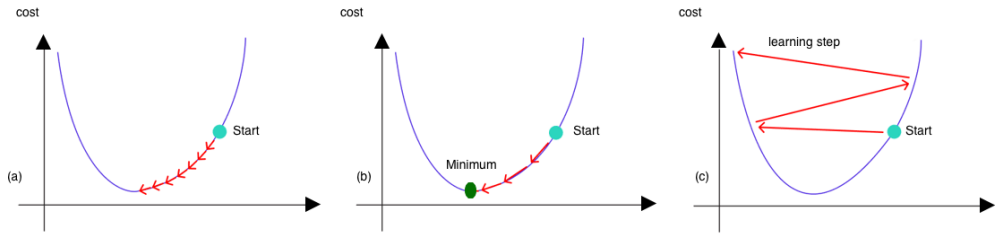


FIGURE 2.6: The image above shows 3 different learning rate settings. (a) is set too low and the model never reaches the optimal loss. (b) is set just right and loss is minimized. (c) is set too high causing the updates to overshoot the optimal solution.

loss function $L(w) = w^2$, the gradient is $\frac{\partial L}{\partial w} = 2w$, and the update rule becomes $w_{t+1} = w_t - \eta(2w_t)$. This iterative process reduces w toward 0, minimizing $L(w)$.

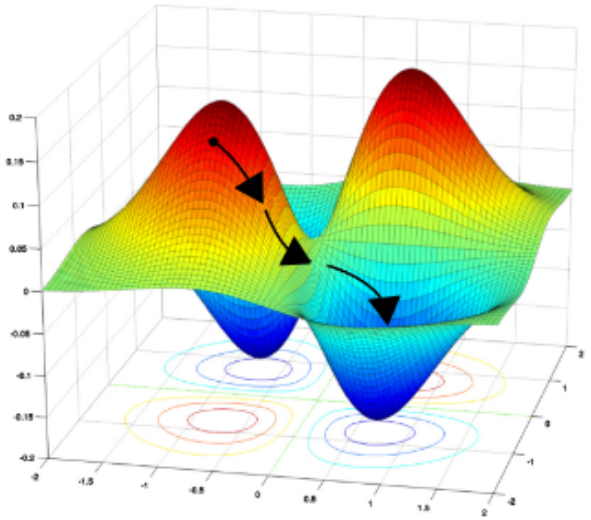


FIGURE 2.7: Visualization of Gradient Descent: Iterative optimization process showing the parameter updates (arrow) moving toward the minimum of the loss function.

2.5 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are foundational to deep learning in image processing. CNNs are particularly useful in medical imaging tasks, such as MAR in CT scans, as they can learn to distinguish metal-induced artifacts from anatomical structures. CNNs were at first created to process images, leveraging their unique architecture to account for the spatial structure of images. This makes them particularly well-suited for tasks like image classification (Nielsen, 2015). When identifying objects in an image, it's essential to recognize that the features describing an object are often located near each other, and their exact position within the image is less important than their relative arrangement. CNNs exploit these principles, known as locality and translational invariance, to achieve superior performance in image analysis.

Similar to standard neural networks, CNNs consist of neurons with learnable weights and biases. However, the primary distinction lies in how data is organized and the mathematical operations used in the hidden layers, which include convolutional and pooling layers. Instead of treating input as a one-dimensional vector, CNNs maintain the two-dimensional structure of images, preserving the grid-like arrangement of pixels. For example, consider a CNN processing a grayscale image with a size of 7×7 pixels. Each pixel represents a grayscale intensity value, and each hidden neuron in the convolutional layer is connected to a small region of the input, called the local receptive field. This receptive field is typically a small window, such as 3×3 pixels, which scans over the entire input image (see Fig. 2.8).

The size of the receptive field and its movement step, called the stride, determine the number of neurons in the convolutional layer. For instance, in the example shown in Fig. 2.8, a 3×3 receptive field with a stride of 1 slides across a 7×7 input, resulting in a 5×5 grid of neurons in the convolutional layer. Both the receptive field size and stride length are key hyperparameters that define the network's architecture. The relationship between the input size, filter size, stride and output size can be shown in the formula:

$$\text{Output size} = \frac{\text{Input size} - \text{Filter size}}{\text{Stride}} + 1$$

A unique feature of CNNs is that all neurons in a convolutional layer share the same set of weights and biases, often referred to as the filter K. This weight sharing ensures that the same feature is detected across various locations in the image, providing the translational invariance characteristic of CNNs. The output of this operation is known as a feature map, and each convolutional layer typically includes multiple feature maps to detect various localized features.

Following the convolutional layer, a pooling layer is applied to condense the feature maps, reducing their dimensionality while retaining essential information (see Fig. 2.9). A common pooling method is max-pooling, where each pooling unit takes the maximum value from a small region, such as a 2×2 area, in the feature map. Pooling can be thought of as a query that determines whether a feature is present, without regard to its precise position within the image.

At the end of the network, a fully connected layer integrates all the learned features to perform the final classification task. By stacking multiple convolutional and pooling layers, CNNs can construct deep, highly expressive architectures capable of learning complex patterns. The use of shared weights and biases, along with pooling operations, reduces the number of trainable parameters significantly, allowing CNNs to learn faster and more efficiently than fully connected neural networks with comparable capacity(Nielsen, 2015). Additionally, the backpropagation algorithm used to train CNNs requires only minor modifications to accommodate the convolutional and pooling operations.

2.5.1 Encoder-Decoder Architectures

The encoder-decoder architecture is a specialized CNN-based model designed to compress input data into a lower-dimensional representation and then reconstruct it with high fidelity.

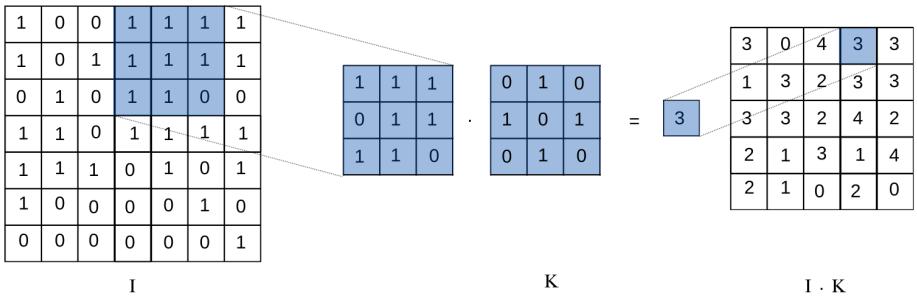


FIGURE 2.8: An example of a 3×3 receptive field scanning a 7×7 image with a stride of 1, resulting in a 5×5 feature map.

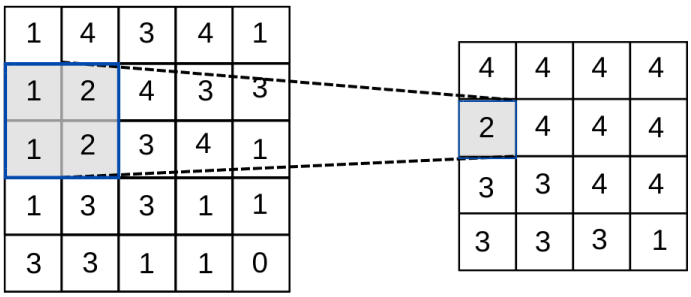


FIGURE 2.9: The max-pooling operation using a 2×2 filter and stride of 1 $(max\{1,2,1,2\} = 2)$

This approach is particularly effective for tasks requiring pixel-level accuracy, such as metal artifact reduction (MAR), where a standard CNN might fall short. A notable example of this architecture is the U-Net model, widely used in medical image segmentation. Its ability to preserve spatial details during transformations makes it well-suited for MAR tasks, where maintaining clear boundaries between artifacts and anatomical structures is critical.

The encoder-decoder architecture consists of two primary components: the encoder and the decoder. The encoder compresses the input image through a series of convolutional and pooling layers, producing a lower-dimensional bottleneck representation. This compression retains the most essential features while eliminating redundant details, enabling the network to effectively distinguish between artifacts and anatomical structures. The decoder then reconstructs the image to its original resolution using up-sampling layers, typically implemented with transposed convolutions. By reversing the encoding process, the decoder produces an artifact-reduced output image that remains faithful to the original anatomical structure. This combination of compression and reconstruction allows the encoder-decoder architecture to deliver precise and high-quality results for MAR and similar pixel-level tasks.

2.6 Generative Adversarial Networks (GANs)

Generative Adversarial Networks (GANs) are a groundbreaking framework in deep learning, particularly known for their capabilities in generative modeling. GANs were introduced by Goodfellow et al. (2014) to generate new data samples that mimic a given dataset. Unlike

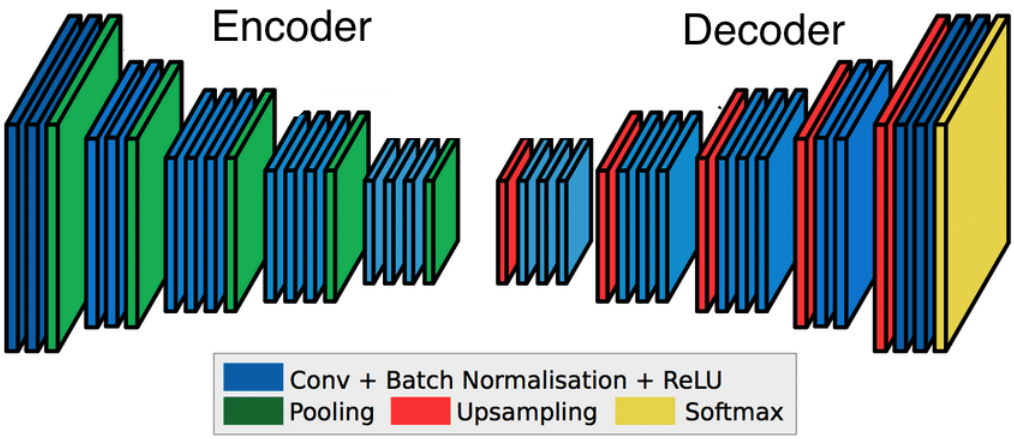


FIGURE 2.10: The encoder compresses input data into a latent representation, capturing essential features, while the decoder reconstructs the artifact-free image, integrating skip connections to preserve spatial resolution. Image from (Hussein & Ali, 2022).

Convolutional Neural Networks (CNNs), which are primarily used for feature extraction and classification tasks, GANs have found applications in fields such as medical imaging, where they can generate synthetic data or reconstruct images with reduced artifacts (Nie et al., 2018).

GANs consist of two neural networks—a generator G and a discriminator D—that compete in a zero-sum game. The generator aims to produce data samples resembling the real dataset, while the discriminator attempts to distinguish between real and synthetic samples. This adversarial training process can be likened to a forger attempting to create counterfeit artwork indistinguishable from authentic pieces, with the discriminator acting as an art critic (see Fig. 2.11).

The generator maps a random noise vector $z \sim \mathcal{N}(0,1)$ to the data space $x \in \mathcal{X}$ through a series of learned transformations, while the discriminator outputs the probability $P(x)$ that a given sample is real. The optimization objective for this setup is given by:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))].$$

Here, $p_{\text{data}}(x)$ is the distribution of real data, and $p_z(z)$ is the prior noise distribution.

GANs have seen tremendous success in medical imaging, particularly in reducing noise and artifacts. For example, in metal artifact reduction (MAR) in CT images, GANs have been used to reconstruct artifact-free images by learning the mapping between corrupted images and their ground-truth counterparts (Nie et al., 2018).

Despite their effectiveness, training GANs is notoriously challenging due to issues like mode collapse, non-convergence, and gradient instability. Techniques such as Wasserstein loss and spectral normalization have been introduced to stabilize training and improve performance (Arjovsky et al., 2017).

GANs are a versatile and powerful tool in deep learning. Their application in medical

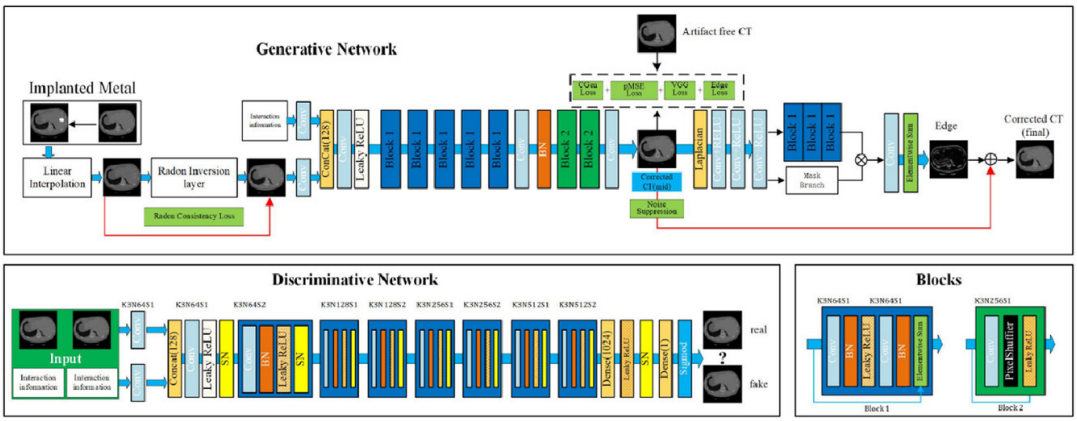


FIGURE 2.11: A schematic representation of a specialized GAN model. The MFE-GAN proposed by (Huang et al., 2022) . The generator learns to produce realistic images, while the discriminator evaluates their authenticity.

imaging, particularly in challenging tasks such as artifact reduction, highlights their potential to transform diagnostic workflows by generating high-quality, artifact-free images while maintaining anatomical fidelity.

Chapter 3

Related Work

3.1 Hand-Crafted Approaches

Research on metal artifact reduction (MAR) in computed tomography (CT) has evolved significantly, transitioning from traditional hand-crafted methods to advanced deep learning-based approaches. This chapter reviews key methods in both categories, setting the stage for the methodological details in Chapter 4.

3.1.1 Hand-Crafted Approaches

Hand-crafted approaches to MAR focus on leveraging domain-specific knowledge of CT imaging physics to mitigate artifacts. A prominent category includes projection-based techniques, such as linear and polynomial interpolation, which aim to fill corrupted regions in the sinogram with estimated values derived from adjacent unaffected data (Abdoli et al., 2012). While effective at reducing streak artifacts caused by beam hardening and photon starvation, these methods often lead to secondary artifacts due to inconsistencies between interpolated and non-interpolated regions.

Dual-energy CT (DECT) has been another transformative development. By exploiting energy-dependent attenuation properties of materials, DECT creates virtual monochromatic images that minimize beam-hardening artifacts near metallic implants. Although effective, the need for specialized hardware and higher radiation doses limits its widespread adoption in routine clinical settings (Kawahara et al., 2019). Another significant advancement in this category is model-based iterative reconstruction (MBIR), which uses physics-driven priors to iteratively refine the reconstructed image. MBIR has been shown to outperform direct sinogram correction methods by addressing multiple sources of artifacts, including scatter and noise, in a unified framework (Thibault et al., 2007).

Despite their utility, hand-crafted methods face limitations in handling complex, non-local artifact structures. Their reliance on simplified physical models often fails to account for diverse imaging conditions, making them less robust than data-driven techniques.

3.1.2 Deep Learning-Based Approaches

The emergence of deep learning has revolutionized MAR by enabling models to learn complex artifact patterns directly from data. Supervised learning methods dominate this space,

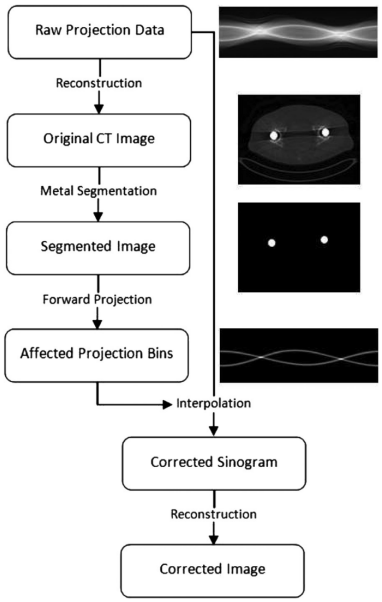


FIGURE 3.1: General flowchart of interpolation-based sinogram correction MAR methods. (Abdoli et al., 2012)

leveraging paired datasets of artifact-corrupted and artifact-free images. For instance, InDuDoNet alternates between sinogram correction and image reconstruction, embedding the physical constraints of CT imaging into its dual-domain learning framework. It has shown significant improvements in reducing artifacts while preserving anatomical details (H. Wang et al., 2021). Building on this foundation, InDuDoNet+ incorporates additional priors to enhance generalization and robustness across diverse datasets (H. Wang, Li, Meng, & Zheng, 2022).

Image-domain methods, such as DICDNet, represent another branch of deep learning MAR. These approaches operate directly on reconstructed images, focusing on enhancing visual quality while reducing streak and shading artifacts. Unlike dual-domain models, which require access to sinogram data, image-domain models are more practical for clinical deployment as they integrate seamlessly into existing workflows (H. Wang, Li, He, et al., 2022).

Unsupervised and semi-supervised methods have recently gained traction for their ability to handle limited or unpaired data. The "Stay in the Middle" framework is a notable example, combining labeled and unlabeled datasets to achieve a balance between supervised accuracy and unsupervised generalization. This model introduces a middle-domain constraint that bridges the gap between labeled and unlabeled data distributions, demonstrating robust performance on challenging datasets. A key aspect of this approach is the explicit subtraction of the clean portion from the metal-corrupted image, isolating the artifact for targeted correction (T. Wang et al., 2023).

Deep learning methods, while highly effective, are not without challenges. Training these models requires large datasets and significant computational resources, and their performance can degrade when faced with out-of-distribution artifacts. However, the ability of these methods to adapt to complex imaging scenarios makes them a compelling alternative to traditional approaches.

Chapter 4

Methodology

This chapter provides a detailed overview of the methodologies used in state-of-the-art deep learning-based metal artifact reduction (MAR) approaches. The discussion is focused on dual-domain methods, including InDuDoNet and its successor InDuDoNet+, and image-domain methods, such as DICDNet (Deep Interpretable Convolutional Dictionary Network) and the semi-supervised model "Stay in the Middle." Each method reflects a unique strategy for addressing the challenges of MAR, leveraging deep learning architectures and domain-specific knowledge.

4.1 Duo Domain

Dual-domain methods address metal artifact reduction (MAR) by simultaneously optimizing in both the sinogram and image domains, leveraging the physical constraints of CT imaging. These approaches iteratively refine corrupted projection data while reconstructing artifact-free images, ensuring consistency across domains. In this study I will focus on 2 architectures of the Duo Domain framework: InDuDoNet and its enhanced version InDuDoNet+.

4.1.1 InDuDoNet: Interpretable Dual Domain Network

InDuDoNet is a pioneering dual-domain network specifically designed to address the metal artifact reduction (MAR) task by combining model-driven and data-driven methodologies. The network is constructed using a deep unfolding framework, which maps the iterative optimization process of a MAR algorithm into a learnable structure. The architecture consists of three main components: Prior-net, N-stage $\tilde{S}$-net, and N-stage X-net, as illustrated in Fig. 4.1.

Prior-net: This module generates an initial estimate $\tilde{Y}$ by leveraging the metal-affected image (X_{ma}) and a linearly interpolated corrected image (X_{LI}). The Prior-net adopts a U-shaped architecture similar to previous models and acts as a precursor for subsequent refinement by providing a normalized input to the next stages.

$\tilde{S}$-net and X-net: The core of the architecture involves N-stages of alternating updates between the $\tilde{S}$-net and X-net.

$\tilde{S}$-net: Processes the normalized sinogram $\tilde{S}_{n-1}$ generated in the previous stage and updates it using a deep network, referred to as $\text{proxNet}_{\theta_{\tilde{s}}^{(n)}}$. This operation corresponds to

the proximal operator in the optimization framework, ensuring that the sinogram remains physically consistent while removing artifacts. **X-net**: Takes the artifact-reduced image X_{n-1} from the prior stage and refines it through another deep network, $\text{proxNet}_{\theta_x^{(n)}}$. This step reconstructs the image by eliminating residual artifacts and ensuring anatomical fidelity.

Each proxNet module in $\tilde{S}$-net and X-net comprises four residual blocks, where each block consists of convolutional layers, batch normalization, ReLU activation, and skip connections. These residual structures enhance feature extraction while maintaining computational efficiency.

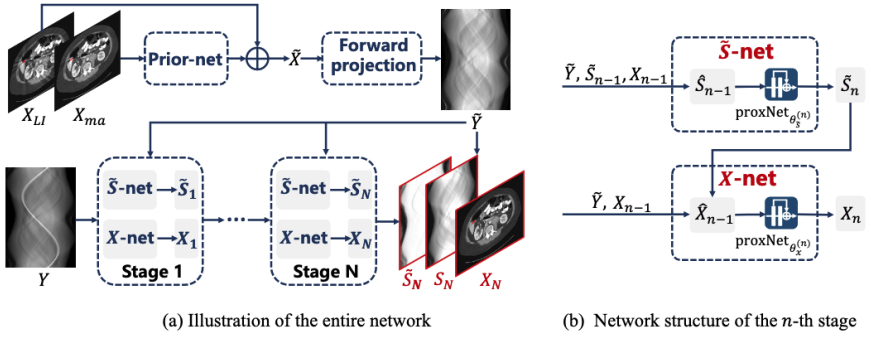
Forward and back projection operations integrate CT imaging constraints into the framework by simulating the reconstruction process. These modules ensure consistency between the sinogram and image domains throughout the network.

Training and Optimization: The training process minimizes a composite loss function that supervises outputs at every stage n for both the sinogram and image domains:

$$\mathcal{L} = \sum_{n=0}^{N} \beta_n \|X_n - X_{\text{gt}}\|_F^2 \odot (1 - M) + \gamma \left(\sum_{n=1}^{N} \beta_n \|\tilde{Y}\tilde{S}_n - Y_{\text{gt}}\|_F^2 \right),$$

where X_{gt} and Y_{gt} are the ground truth artifact-free image and sinogram, respectively. The weights β_n ensure that the final stage plays a dominant role, while intermediate stages are lightly supervised.

Key Features: InDuDoNet combines the strengths of prior-based methods and deep learning. Unlike traditional MAR techniques, it flexibly learns sinogram- and image-related priors directly from training data, while integrating CT imaging constraints and dual-domain priors. This hybrid approach provides physical interpretability and enhances generalization across diverse datasets.



(a) Illustration of the entire network (b) Network structure of the n-th stage

FIGURE 4.1: Illustration of the InDuDoNet architecture, including Prior-net, N-stage $\tilde{S}$-net, and N-stage X-net. The framework alternates between sinogram and image refinement.(H. Wang et al., 2021)

4.1.2 InDuDoNet+: Deep Unfolding Dual Domain Network

InDuDoNet+ builds upon the foundational InDuDoNet framework by addressing two key limitations of traditional deep learning-based metal artifact reduction (MAR) methods: the

insufficient incorporation of CT imaging geometry constraints and the lack of interpretability in network modules. By embedding the CT imaging process more comprehensively and introducing new prior-based enhancements, InDuDoNet+ achieves superior performance and generalization.

Key Enhancements over InDuDoNet: InDuDoNet+ refines the dual-domain approach by embedding the physical properties of CT imaging more comprehensively into its network design. This refinement is achieved through the derivation of a joint spatial and Radon domain reconstruction model, which tightly couples the image and sinogram domains. Unlike InDuDoNet, which alternates between domain corrections, InDuDoNet+ employs a simplified optimization algorithm to ensure consistency throughout the MAR process. Additionally, InDuDoNet+ addresses the lack of interpretability in traditional deep learning methods by unfolding the iterative steps of its optimization algorithm into corresponding network modules. Each module represents a specific step, making the role of individual components easier to evaluate in the context of artifact reduction. Another key improvement is the incorporation of a prior network that leverages observations of CT value distributions across different tissues. By guiding the network's learning process with this prior knowledge, InDuDoNet+ significantly enhances its ability to generalize to unseen data, particularly in scenarios involving complex artifact patterns.

Optimization Framework and Training: InDuDoNet+ adopts a deep unfolding methodology, where the iterative steps of the optimization process are translated into learnable modules. Each stage alternates between sinogram correction and image refinement, similar to InDuDoNet, but is guided by the prior network to further enhance corrections. The training process minimizes the following composite loss function:

$$\mathcal{L} = \sum_{n=0}^{N} \beta_n \| X_n - X_{\text{gt}} \|_F^2 \odot (1 - M) + \gamma \left(\sum_{n=1}^{N} \beta_n \| \tilde{Y} \odot \tilde{S}_n - Y_{\text{gt}} \|_F^2 \right),$$

where X_{gt} and Y_{gt} are the ground truth artifact-free image and sinogram, respectively. The weights β_n prioritize the final stage while lightly supervising intermediate outputs.

InDuDoNet+ builds on the foundation of InDuDoNet by incorporating advanced priors and constraints to enhance generalization and robustness. (H. Wang, Li, Zhang, et al., 2023).

4.2 Image Domain

Image-domain methods focus on improving reconstructed CT images directly, without relying on the intermediate sinogram domain. These approaches aim to enhance image quality by suppressing artifacts in post-reconstruction images, making them computationally efficient and easier to integrate into existing clinical workflows. In this study, I will focus on two architectures within the image-domain framework: the fully supervised DICDNet and the semi-supervised "Stay in the Middle" model, which addresses the challenges of limited labeled datasets through innovative use of both supervised and unsupervised learning strategies.

DICDNet: Deep Interpretable Convolutional Dictionary Network

DICDNet is an image-domain model specifically designed for metal artifact reduction (MAR) by directly operating on reconstructed CT images. Unlike sinogram-based approaches, DICDNet eliminates the need for sinogram data, making it more practical for real-world applications where sinogram information may be inaccessible. This approach frames MAR as an image restoration problem, utilizing tailored network modules and prior-based methodologies to suppress artifacts.

The DICDNet framework leverages the intrinsic prior structure of metal artifacts, which often exhibit repetitive non-local streaking and star-shaped patterns. To explicitly model these observations, the network employs a convolutional dictionary model in which convolutional kernels represent repetitive local patterns, and feature maps encode the spatial locations of artifacts. This formulation ensures that the network learns to identify and suppress artifact patterns effectively.

The network adopts a stage-wise architecture, where each stage comprises two modules: M-net and X-net. M-net extracts artifact-related feature maps ($M^{(s)}$), which represent the locations and intensity patterns of metal-induced artifacts. These maps are refined to isolate non-local streaking and star-shaped patterns commonly observed in metal-corrupted images. X-net uses these refined maps to update the corrected CT image ($X^{(s)}$), progressively suppressing residual artifacts while preserving structural details. This iterative design allows DICDNet to leverage convolutional dictionary models to encode repetitive artifact patterns and spatial locations effectively.

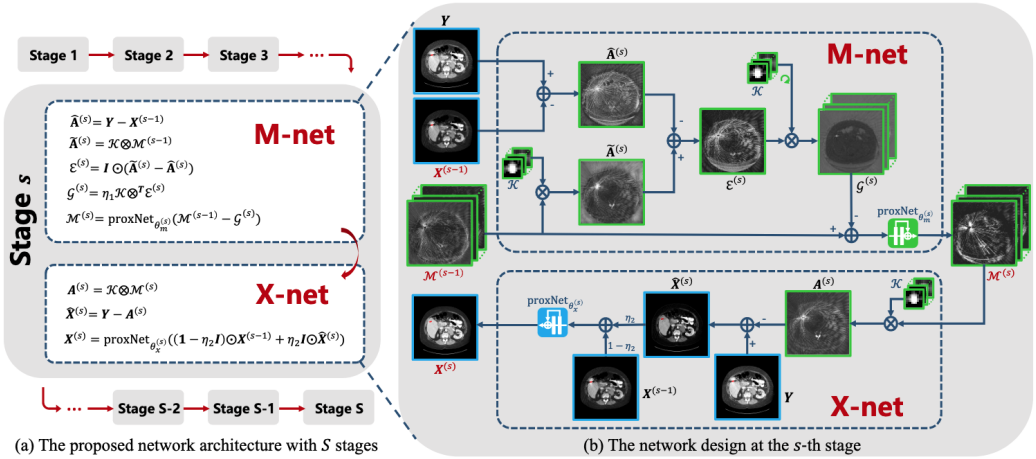


FIGURE 4.2: The DICDNet architecture operates in S iterative stages, where each stage updates the artifact-related feature map $M^{(s)}$ and the corrected CT image $X^{(s)}$ using two dedicated modules: M-net and X-net. M-net identifies and refines metal artifact patterns, while X-net progressively reconstructs the artifact-reduced image. These modules are based on computational operators derived from the underlying optimization framework, ensuring interpretability and effectiveness. The network is trained on image patches and tested on full metal-corrupted CT images, demonstrating robust performance in removing artifacts while preserving anatomical details. (T. Wang et al., 2023)

Compared to sinogram-involved approaches, DICDNet is easier to integrate into existing clinical workflows as a plug-in module. Furthermore, the explicit incorporation of artifact

priors provides the network with superior generalization capabilities, as demonstrated by its performance on both synthesized and clinical datasets. The network has achieved state-of-the-art (SOTA) results for MAR tasks and has been shown to improve the accuracy of downstream applications such as multi-bone segmentation in pelvic fracture cases (T. Wang et al., 2023).

ACDNet: Adaptive Convolutional Dictionary Network

The Adaptive Convolutional Dictionary Network (ACDNet) is a deep learning method in the image domain specifically developed for metal artifact reduction (MAR) in computed tomography (CT) images. ACDNet merges the interpretability of model-based techniques with the powerful representation capabilities of deep neural networks, effectively overcoming the limitations of conventional sinogram-based approaches.

ACDNet explicitly models metal artifacts through a weighted convolutional dictionary model, capturing both common artifact patterns and their specific variations across individual CT images. This adaptive mechanism encodes non-local repetitive streaking patterns typically present in metal artifacts, facilitating a tailored and effective artifact reduction strategy for each image.

The architecture of ACDNet employs a clear iterative unfolding structure based on optimization algorithms. It consists of distinct modules—K-net, M-net, and X-net—which sequentially update the artifact weighting coefficients (K), the artifact feature maps (M), and the artifact-reduced CT image (X). This design provides clear interpretability by embedding explicit prior knowledge about artifact characteristics into the network, thus ensuring transparent learning and superior generalization capabilities.

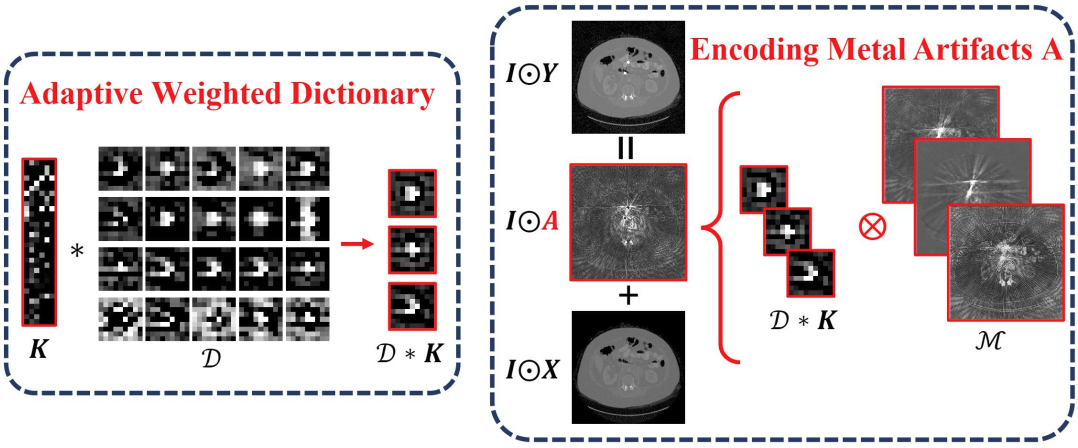


FIGURE 4.3: Illustration of the proposed weighted convolutional dictionary model for metal artifacts A, represented as $(D * K) \otimes M$. The dictionary D remains fixed across all samples, whereas the weighting coefficient K varies per sample. By adjusting K, the artifact representation kernel $(D * K)$ can adapt dynamically to each individual input image Y. (H. Wang, Li, Meng, & Zheng, 2022)

Compared to the earlier DICDNet, ACDNet builds upon similar conceptual foundations but introduces an adaptive weighting scheme that dynamically adjusts convolutional kernels based on input specifics. While DICDNet primarily relies on fixed convolutional kernels

within its dictionary model, ACDNet enhances flexibility by adapting kernels to individual images, thereby further improving artifact suppression and generalization across diverse scenarios.

Due to its purely image-domain operation, ACDNet easily integrates into clinical workflows as a plug-in module. Comprehensive experimental validation using synthetic and clinical datasets demonstrates its superior performance over conventional and contemporary MAR methods, with proven effectiveness in critical downstream medical applications such as anatomical segmentation tasks affected by metal artifacts.

OSCNet: Orientation-Shared Convolutional Network

The Orientation-Shared Convolutional Network (OSCNet) is an advanced deep-learning model specifically designed for metal artifact reduction (MAR) in computed tomography (CT) images. OSCNet distinguishes itself by explicitly incorporating a critical prior knowledge about metal artifacts: their characteristic rotationally symmetrical streaking patterns, a result of rotational scanning in CT imaging.

OSCNet introduces an orientation-shared convolution representation mechanism to model these rotationally symmetrical artifacts effectively. By employing Fourier-series-expansion-based filter parameterization, the model adeptly captures the inherent repetitive streaking structures across multiple rotation angles. This approach significantly reduces the number of learnable parameters while ensuring accurate representation and extraction of metal artifacts.

The architecture of OSCNet leverages deep unfolding techniques, building upon iterative optimization methods to achieve clear interpretability and robust performance. It comprises sequential modules, M-net and X-net, that iteratively refine artifact-related feature maps and corrected CT images, respectively. Furthermore, OSCNet includes a dynamic sub-network, termed OSCNet+. OSCNet+ differs from OSCNet as it adds dynamic inference capabilities by allowing the network to adjust the convolutional filters for each input image (sample-variant) creating a more flexible artifact modeling. These, along with a sub-network that dynamically computes the filter parameters during each iteration, increase flexibility and generalization performance.

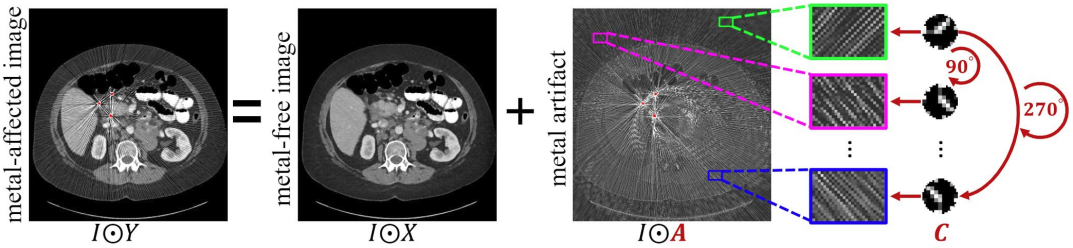


FIGURE 4.4: Illustration of the workings of OSCNet. The metal artifact A exhibits distinct rotational streak patterns, which can be captured by rotating a convolution filter C across multiple orientations (i.e., orientation-shared). In this context, Y, X, and I denote the metal-affected CT image, the ground truth image, and the non-metal region mask, respectively. The red pixels in both Y and A indicate the presence of metallic implants. (H. Wang, Xie, et al., 2023)

OSCNet operates entirely within the image domain, avoiding dependence on sinogram data, which simplifies its integration into clinical workflows. Extensive evaluations on both synthetic and clinical datasets demonstrate OSCNet's superior capability in artifact suppression, outperforming conventional and other contemporary deep-learning MAR methods. Its effectiveness has also been validated through enhanced accuracy in clinical downstream applications, indicating substantial potential for real-world clinical adoption.

Stay in the Middle: A Semi-Supervised Approach

"Stay in the Middle" introduces a semi-supervised learning framework for metal artifact reduction (MAR) that combines the strengths of supervised and unsupervised methods. Its core innovation is the "middle domain," a shared latent space that aligns representations from labeled and unlabeled data. This enables the model to leverage the detailed supervision from paired data and the generalization capabilities of unpaired data, achieving robust artifact correction across diverse datasets.

The architecture, shown in Fig. 4.5, comprises a paired encoder-decoder network, denoted as NetE and NetD, respectively. Given a metal-corrupted CT image (X_{low}), a pre-trained prior image (X_{pre}), and an unpaired clean CT image (X_{high}), the encoder extracts invariant content features. These features are mapped back to their corresponding image domains by the decoder, enabling the generation of a corrected CT image (X_{corr}) and other reconstructions. The encoder functions as a degeneration-invariant feature extractor, ensuring that all inputs can be mapped into a shared latent space.

The process explicitly models artifacts as additive components, represented as the difference between corrupted and corrected images ($X_{\text{art}} = X_{\text{low}} - X_{\text{corr}}$). These artifact maps are then used to synthesize new metal-corrupted images ($X_{\text{hl}} = X_{\text{high}} + X_{\text{art}}$) and further refine reconstructions through cyclic consistency. This cyclic reconstruction ensures minimal information loss during artifact disentanglement, enforcing robust learning of the inverse mapping between corrupted and clean images.

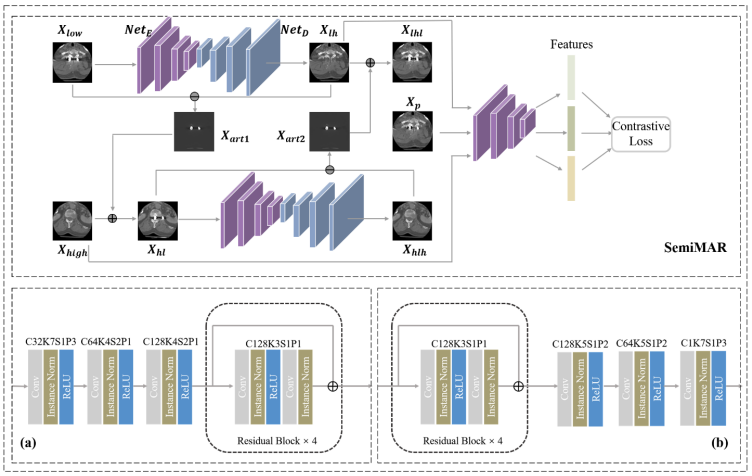


FIGURE 4.5: Overview of the SemiMAR architecture. NetE and NetD represent the encoder and decoder, respectively. (T. Wang et al., 2023).

The symmetrical architecture of NetE and NetD uses four residual blocks equipped with downsampling and upsampling layers, as shown in Fig. 4.5 (a) and (b). The adversarial training framework employs PatchGAN-based discriminators to distinguish between real and generated metal-corrupted and artifact-free images, improving the realism and quality of the outputs. The network effectively balances artifact suppression with structural preservation, demonstrating superior performance on both synthetic and clinical datasets.(T. Wang et al., 2023).

Chapter 5

Datasets and Evaluation Metrics

5.1 Datasets

For this study, I plan to use three main datasets: Deep Lesion, SynDeepLesion and the CT-Spine1K. Each of these was used with the different Deep Learning architectures I will be comparing in this research.

5.1.1 Deep Lesion

The Deep Lesion dataset, provided by the National Institutes of Health (NIH), is a large-scale radiological image dataset designed to support advanced medical imaging research. It contains 32,735 annotated lesions in 32,120 CT slices from 10,594 studies and 4,427 patients, capturing a diverse range of lesion types such as lung nodules, liver tumors, and enlarged lymph nodes (Yan et al., 2017). Each lesion is annotated with bounding boxes and clinical measurements added by radiologists during diagnostic interpretation, ensuring a high degree of clinical realism. These annotations provide positional and contextual information that mimic real-world diagnostic scenarios.

The dataset spans multiple anatomical regions, including thoracic, abdominal, and pelvic CT images, and features a variety of lesion sizes, shapes, and contexts. This diversity makes it an ideal resource for evaluating imaging algorithms that require generalizability across heterogeneous conditions.

FIGURE 5.1: Example slices from the Deep Lesion dataset showing bounding box annotations.

In this research, the Deep Lesion dataset serves as the foundation for creating synthetic metal artifacts, enabling precise comparisons between artifact-affected images and their artifact-free ground truth counterparts. Inspired by methodologies in studies such as *InDuDoNet*,

InDuDoNet+, *DICDNet*, and *STAY IN THE MIDDLE*, synthetic artifacts are applied to simulate realistic clinical challenges caused by metallic implants. The dataset's diversity ensures robust testing of artifact removal methods across varying levels of structural complexity and noise, supporting detailed evaluations of the selected MAR algorithms.

5.1.2 SynDeepLesion

The SynDeepLesion dataset is a synthetically generated collection of CT images designed to simulate realistic metal artifacts, facilitating a systematic evaluation of metal artifact reduction (MAR) techniques. Using clean CT images from the Deep Lesion dataset and metal masks, controlled artifacts are introduced to enable precise comparisons with artifact-free ground truth images.

Construction and Simulation Process

The SynDeepLesion dataset is constructed following the simulation settings in (Yu, Zhang, Li, & Xing, 2021), combining fan-beam geometry with 100 metallic implant masks from (Yan et al., 2017):

- **Training Dataset:** 1,000 clean CT images paired with 90 metal masks to create synthetic training samples.
- **Testing Dataset:** 200 CT images from 12 patients paired with the 10 remaining metal masks, generating 2,000 synthetic test images. Metal masks vary in size, ranging from 2,061 to 35 pixels, grouped by size for average MAR performance evaluation.

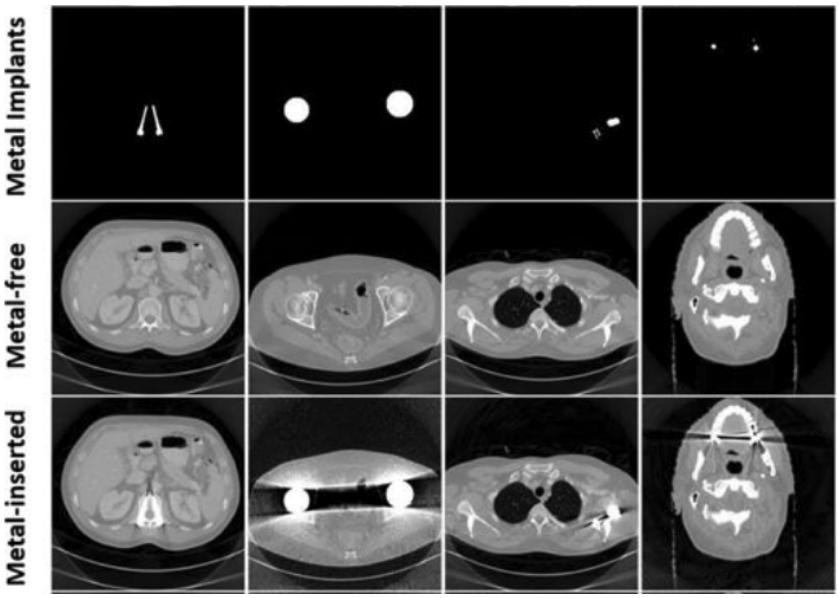


FIGURE 5.2: Example slices from the Synthetic Deep Lesion dataset showing metal artifacts, metal free and metal inserted images. Images from (Zhang & Yu, 2018)

Artifacts are simulated using widely adopted procedures (Zhang & Yu, 2018), incorporating polychromatic X-ray effects, beam hardening, partial volume effects, and Poisson noise.

All images are resized to 416 × 416 pixels, with sinograms set to 641 × 640 dimensions for uniform processing.

5.1.3 CTSpine1K

The CTPelvic1K dataset is a large-scale collection of pelvic CT images, featuring 1,184 3D volumes and over 320,000 CT slices, curated to replicate real-world clinical variability. Among its seven sub-datasets, the CLINIC and CLINIC-metal subsets are particularly relevant for pelvic fracture analysis and metal artifact reduction research.

- **CLINIC:** Includes 103 preoperative CT volumes free from metal artifacts, offering clean imaging data for training and validation of segmentation models.
- **CLINIC-metal:** Consists of 75 postoperative CT volumes containing metal artifacts from implants such as screws and prosthetics. Of these, 14 cases are annotated for detailed segmentation and artifact studies, while the remaining 61 cases are suitable for unsupervised learning tasks.

Both datasets are stored in DICOM format, preserving scanner-specific metadata, and have been reformatted to NIfTI for compatibility with preprocessing pipelines. These subsets provide a valuable resource for studying metal artifact challenges, enabling controlled experiments with clean ground truth images and artifact-affected counterparts. Notably, the DICDNet paper utilizes the CLINIC dataset to train a pelvic segmentation model and tests its performance on artifact-reduced images from the CLINIC-metal subset.

CTPelvic1K, including its CLINIC and CLINIC-metal subsets, is publicly available under the Creative Commons CC-BY-NC-SA 4.0 license, supporting advancements in segmentation and artifact reduction research.

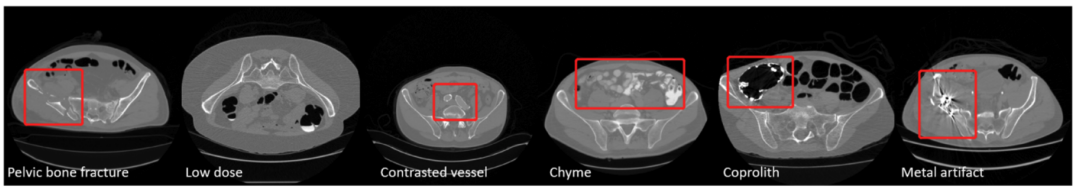


FIGURE 5.3: CTSpine1K CT image examples with various conditions

5.2 Evaluation Metrics

The performance of the proposed metal artifact reduction methods will be quantitatively calculated before and after application of MAR using the following evaluation metrics:

5.2.1 Structural Similarity Index (SSIM):

This metric evaluates the perceptual quality of images by quantifying the similarity between the artifact-corrected image and the ground truth. SSIM measures structural information, contrast, and luminance consistency, providing a holistic assessment of image fidelity.

Higher SSIM values indicate more effective artifact suppression and better preservation of fine structural details in the corrected images.

In the context of metal artifact reduction, the *ground truth image* (gt) represents the artifact-free reference image, which serves as the baseline for comparison. The *metal artifact reduced image* (MAR) refers to the image obtained after applying the artifact correction technique being evaluated. SSIM is used to determine how closely the corrected image (MAR) aligns with the reference image (gt) in terms of structural similarity. The calculation of SSIM is defined by the following formula:

$$\text{SSIM}(\text{gt}, \text{MAR}) = \frac{(2\mu_{\text{gt}}\mu_{\text{MAR}} + C_1)(2\sigma_{\text{gt,MAR}} + C_2)}{(\mu_{\text{gt}}^2 + \mu_{\text{MAR}}^2 + C_1)(\sigma_{\text{gt}}^2 + \sigma_{\text{MAR}}^2 + C_2)} \tag{5.1}$$

Here:

- μ_{gt} and μ_{MAR} are the mean intensity values of the ground truth and metal artifact reduced images, respectively.
- σ_{gt}^2 and σ_{MAR}^2 are the variances of the ground truth and metal artifact reduced images.
- $\sigma_{\text{gt,MAR}}$ is the covariance between the ground truth and the corrected image.
- C_1 and C_2 are small constants added for numerical stability, typically defined as $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$, where L is the dynamic range of the pixel values, and k_1, k_2 are constants (commonly $k_1 = 0.01$ and $k_2 = 0.03$).

The SSIM metric, with its emphasis on perceptual quality, is particularly suited for evaluating metal artifact reduction techniques as it ensures the reconstructed image retains clinically relevant structural information.

5.2.2 Peak Signal-to-Noise Ratio (PSNR):

This metric evaluates the fidelity of reconstructed images by comparing the pixel-wise differences between the corrected image and the ground truth. PSNR quantifies the ratio of the maximum possible signal power to the power of corrupting noise, expressed in decibels (dB). Higher PSNR values indicate better image quality, characterized by reduced noise and minimal distortions.

In the context of metal artifact reduction, the *ground truth image* (gt) serves as the reference artifact-free image, while the *metal artifact reduced image* (MAR) represents the reconstructed image after applying the correction technique. PSNR provides a measure of how closely the MAR image approximates the gt image in terms of pixel intensity accuracy. The calculation of PSNR is defined as follows:

$$\text{PSNR}(\text{gt}, \text{MAR}) = 10 \cdot \log_{10}\left(\frac{L^2}{\text{MSE}(\text{gt}, \text{MAR})}\right) \tag{5.2}$$

Here:

- L is the maximum possible pixel intensity value in the image. For example, for 8-bit grayscale images, $L = 255$.

- $\mathrm{MSE}(\mathrm{gt}, \mathrm{MAR})$ is the Mean Squared Error between the ground truth image (gt) and the corrected image (MAR), defined in Equation 5.3.

PSNR is particularly useful for assessing the global similarity between two images. In the context of artifact reduction, higher PSNR values indicate that the corrected image (MAR) is closer to ground truth (gt) in terms of pixel-wise intensity, signifying effective artifact suppression and preservation of fidelity.

5.2.3 Mean Squared Error (MSE):

This metric computes the average squared differences between the pixel values of the corrected image and the ground truth image. It is a fundamental measure of error that quantifies the degree to which the reconstructed image (MAR) approximates the reference image (gt). Lower MSE values indicate a higher degree of similarity to the ground truth, reflecting better reconstruction fidelity and more effective artifact suppression. The MSE is calculated as follows:

$$\mathrm{MSE}(\mathrm{gt}, \mathrm{MAR}) = \frac{1}{N} \sum_{i=1}^{N} (I_{\mathrm{gt},i} - I_{\mathrm{MAR},i})^2 \tag{5.3}$$

Here:

- $I_{\mathrm{gt},i}$ and $I_{\mathrm{MAR},i}$ represent the pixel intensity values at the i-th position in the ground truth and corrected images, respectively.
- N denotes the total number of pixels in the image.

MSE has several important characteristics:

- **Sensitivity to large errors:** The squaring operation amplifies the larger differences between the pixel values, making MSE particularly sensitive to significant artifacts or distortions.
- **Interpretability:** MSE provides an intuitive measure of error, where a value of 0 indicates a perfect match between the corrected and ground truth images.
- **Scale dependence:** Since the MSE values depend on the dynamic range of the pixel intensities, it is often paired with other metrics such as PSNR to provide a normalized interpretation of image quality.

In the context of metal artifact reduction, MSE evaluates the pixel-wise accuracy of the corrected image and complements perceptual metrics like SSIM and normalized metrics like PSNR. Together, these metrics offer a comprehensive assessment of the proposed method's ability to suppress artifacts while preserving clinically relevant image details.

Chapter 6

Experimental Results

6.1 Implementation Details

The experiments were conducted using a custom built desktop computer running Linux Ubuntu equipped with an NVIDIA RTX 3090 Ti GPU featuring 24 GB of memory, supported by an Intel CPU i7 12th gen and 128 GB RAM. The deep learning models were implemented using Python 3.6, the PyTorch 1.4.0 framework, optimized to run with CUDA 10.1. Python scripts for data preprocessing, model training, and evaluation were executed within a Conda-managed environment, ensuring reproducibility and dependency management. While the original papers utilized an NVIDIA Tesla V100-SMX2 GPU, the NVIDIA RTX 3090 Ti is used in this study. It provided improved computational performance and greater memory capacity allowing to run the models significantly faster.

All of the models were optimized using the Adam optimizer with $(\beta_1, \beta_2) = (0.5, 0.999)$. An initial learning rate of 2×10^{-4} is used for all models however, it is halved at different epochs for each model as follows:

- InDuDoNet: every 40 epochs
- InDuDoNet+: every 40 epochs
- DICDNet: every 30 epochs
- ACDNet: every 50 epochs
- OSCNet: every 30 epochs
- OSCNet+: every 30 epochs

Batch sizes of 16 were used for DICDNet, OSCNet, OSCNet+ and ACDNet with input patches of size 64×64 pixels randomly flipped horizontally and vertically for data augmentation.

6.2 Quantitative Analysis

All six models were successfully implemented with the DeepLesion dataset, and their performance metrics closely matched those reported in previous studies, confirming reproducibility and accuracy. The quantitative results obtained using PSNR and SSIM metrics for each model are summarized and presented clearly in the table below:

Model	PSNR (Orig.)	PSNR (Repro.)	PSNR (STD)	SSIM (Orig.)	SSIM (Repro.)	SSIM (STD)
InDuDoNet	41.48	42.61	± 5.67	0.9904	0.9937	± 0.0042
InDuDoNet+	41.50	40.69	± 5.42	0.9891	0.9917	± 0.0058
DICDNet	41.83	41.25	± 4.88	0.9923	0.9864	± 0.0069
ACDNet	40.68	41.43	± 4.49	0.9933	0.9874	± 0.0065
OSCNet	42.19	41.13	± 4.71	0.9931	0.9875	± 0.0061
OSCNet+	42.93	39.55	± 4.40	0.9943	0.9869	± 0.0068

TABLE 6.1: Quantitative comparison of reproduced results and original authors' metrics (PSNR, SSIM) including standard deviations for both metrics.

6.3 Qualitative analysis

The qualitative analysis visually compares different MAR methods applied to the synthesized DeepLesion dataset. Three representative cases were chosen to demonstrate the effectiveness of MAR techniques across varying sizes of metallic implants. Small, medium and large. Each case shows the input images alongside their ground truth, highlighting specific artifacts caused by the metallic implants. Subsequently, images reconstructed by different MAR techniques—including InDuDoNet, InDuDoNet+, DICDNet, ACDNet, OSCNet, and OSCNet+, are presented for direct visual comparison. This analysis focuses on observable improvements, such as artifact reduction, preservation of anatomical structures, and clarity enhancement. These visual comparisons support the evaluation of each method's effectiveness and highlight strengths or limitations that numeric metrics alone may not capture.

To further understand the strengths and weaknesses of each metal artifact reduction (MAR) method, I analyzed representative cases across small, medium, and large metallic implants. Each row in the figure illustrates how well the models recover anatomical structures while suppressing artifacts, offering insight into their robustness under varying levels of metal-induced distortion.

In the case of small metallic implants (row a), the input image contains moderate streaking that radiates from the implant site and disrupts soft tissue continuity. InDuDoNet and its improved variant, InDuDoNet+, reduce artifacts effectively, though some residual streaks remain, especially in the periphery (this is particularly clear in the full CT images). DICDNet also significantly reduces the artifacts, though some smoothing in the soft tissues is visible. ACDNet appears to better balance artifact removal and texture retention, with sharper details in the surrounding bowel region. OSCNet and OSCNet+ both offer effective artifact suppression, with OSCNet+ performing particularly well in preserving tissue contrast and minimizing distortion near the implant site.

With medium metallic implants (row b), artifact severity increases, especially around dense bone structures such as the lumbar vertebra. The input image shows substantial streaking obscuring the lumbar spine. InDuDoNet exhibits some edge degradation and fails to fully restore the underlying structures. InDuDoNet+ corrects this partially but still underperforms compared to the OSC-based approaches. This is visible with the remaining light and dark artifacts below and around the spinous process of the lumbar spine. In contrast, DICDNet again suppresses artifacts but blurs the cortical bone boundary slightly. ACDNet provides a good compromise, maintaining clearer anatomical structure with acceptable artifact levels. OSCNet shows noticeable improvement in streak removal, and OSCNet+

enhances this by preserving the bone detail with minimal over-smoothing.

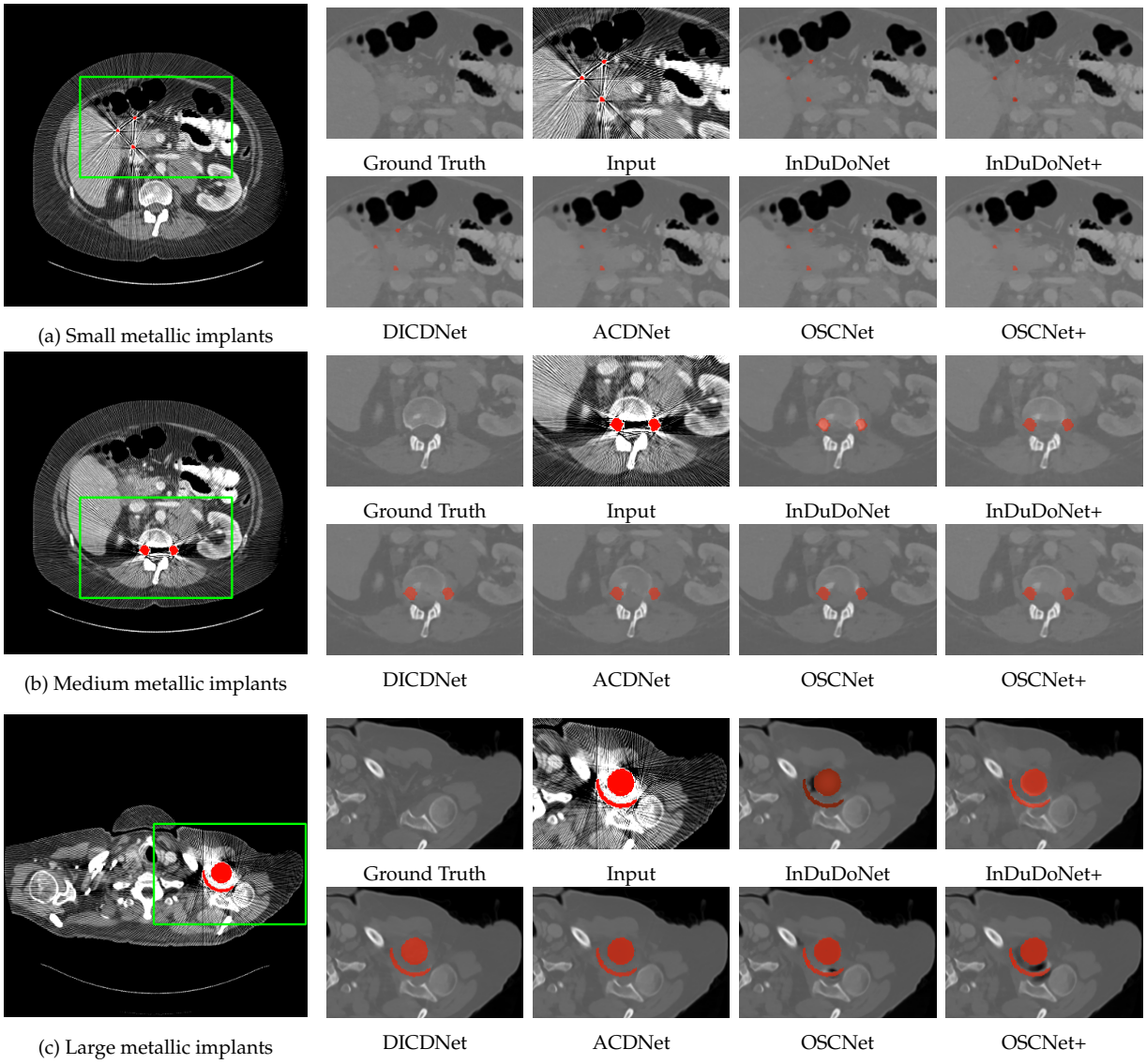


FIGURE 6.1: Here we compare different MAR techniques using the synthesized DeepLesion dataset on three distinct metallic implant sizes.

For large metallic implants (row c), the challenge is more severe. The input image reveals extensive beam hardening and star-shaped artifacts that severely compromise visibility of both soft and hard tissue structures. Though InDuDoNet introduces a dark patch to the left of the metal implant it does reduce artifact presence. InDuDoNet+ similarly reduces strong artifacts but struggles to recover finer anatomical boundaries and shows signs of low fidelity in high-intensity areas. DICDNet removes many of the artifacts but introduces a visibly smooth appearance, causing loss of anatomical detail (visible in the lower left of the femoral bone). ACDNet performs relatively better, with more defined edges around the femoral bone. OSCNet seems to struggle with the high intensity areas and introduces dark spots in places(possibly due to beam hardening); however, it retains good tissue definition

and preserves anatomical areas quite well around the metal inplant. OSCNet+ also struggles with the high intensity areas and introduces even more dark spots to the image. Irrespective of this it performs well by restoring continuous bone structure (visible on the lower left of the femoral bone) and soft tissue clarity even near the metal regions.

Overall, OSCNet+ consistently demonstrates strong generalization and performance across all implant sizes, striking the best balance between artifact suppression and structure preservation. ACDNet also shows strong performance, particularly in maintaining soft tissue integrity. DICDNet remains effective in basic artifact removal but may oversmooth anatomical detail, while InDuDoNet and InDuDoNet+ offer reasonable improvements but are less effective under higher artifact conditions.

6.3.1 Detailed Examination of Model Outputs

To further evaluate and compare the qualitative performance of the MAR models, I closely examined specific regions of interest (ROIs) in the reconstructed CT images that included large metal implants. My analysis primarily aimed to understand each model's capability to reconstruct bone, soft tissue, and anatomical details while effectively suppressing metal artifacts.

The ground truth image serves as a clean reference and clearly depicts well-defined anatomical structures without any visible artifacts. In contrast, the input image, which is fed to the models, shows significant streaking artifacts caused by metallic implants. These artifacts severely obscure anatomical details, especially near the metal sites, affecting both bone and soft tissue clarity.

Beginning with InDuDoNet, the model reduces streaking artifacts to a moderate extent. However, it tends to oversmooth soft tissue regions, leading to noticeable loss of detail. Metal artifacts are still clearly visible emanating horizontally from the metal implants on the center of the image. This shows while the model can reduce most artifacts it struggles to remove the artifacts fully under the high intensity areas found in large implants. InDuDoNet also tends to produce slightly brighter reconstruction of the image structure around metal implants compared to other models, as seen visibly in the remaining outline of the implant, the anterior surface of the vertebral cortical bone, and the surrounding areas. InDuDoNet+ improves slightly on this result, offering somewhat sharper anatomical boundaries and better quality reconstruction of organs. This is visible in the complete reconstruction of the liver where InDuDoNet struggled. Though this model shows an improved performance residual linear artifacts are still visible, especially around the periphery of the implants. These remaining artifacts are slightly darker than the ones left behind by InDuDoNet resulting in a higher quality image recosntruction.

DICDNet achieves cleaner artifact removal compared to the previous two models, significantly reducing streaks to the point where no visible streaks remain. However, it exhibits excessive smoothing of structures. This smoothing particularly affects finer textures and reduces clarity at boundaries adjacent to the metal implants. This effect is particularly strong on the (image) right metal implant where the entire radius around the artifact has been blurred to form one uniform structure. To a lesser degree, the smoothing effect on the (image) left metal implant has slightly merged the kidney with the patient's liver. Finally on

possible negative impact of this model is that the metal implants merge with the anatomical structures, making it hard to distinguish between the two which may complicate diagnostics.

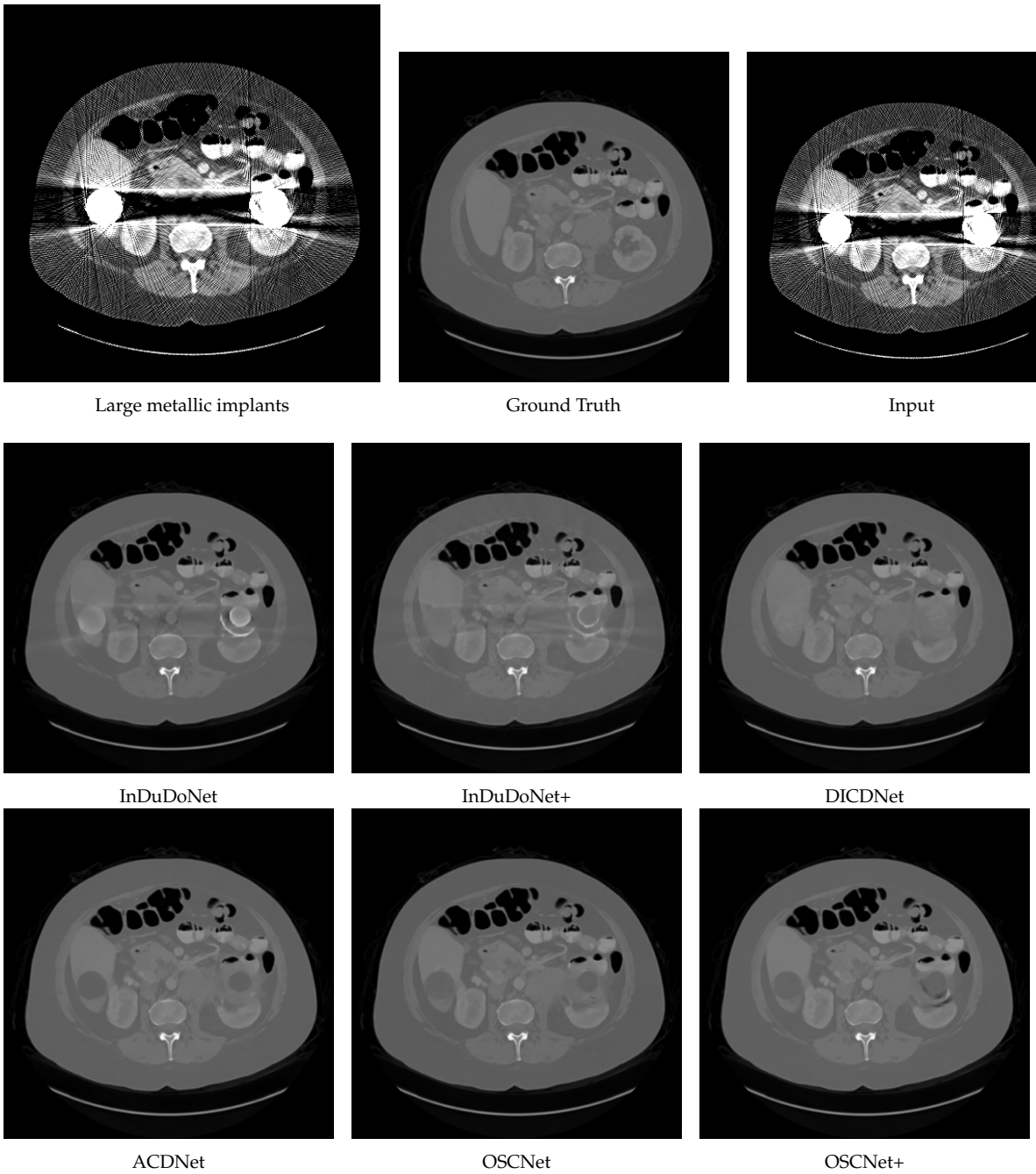


FIGURE 6.2: Closer investigation of reconstructed images of large metallic implants. The metal-corrupted CT images is from the synthesized DeepLesion dataset.

ACDNet demonstrates robust artifact suppression and maintains a balanced reconstruction of bone and soft tissue details. It preserves tissue contrast very well, providing slightly better soft tissue definition compared to OSCNet+. This is very clear in the contrast definition between the kidney and liver structures as well as the kidney reconstruction itself. The

metal implants are clearly delineated(which may aid in diagnostics) and bright metal streaks are mostly absent although faint artifact residues can occasionally still be observed. A notable observation is the small bright spot just below and to the right of the metal implant on the image's right side as well as the star-shaped pattern visible to the left of the same metal implant.

Similar to ACDNet, OSCNet effectively suppresses most artifacts, notably reducing streaks and preserving tissue contrast around the lumbar spine. Despite these strengths, slight blurring is detectable in finer soft tissue textures. Small dark patches are also visible around the metal implants. One patch is visible above the metal implant on the images's left as well as two elongated dark patches below the metal implant on the image's right side. Though the metal implants are outlined clearly these dark spots may mislead in diagnostics.

Finally, OSCNet+ produces refined reconstruction results, similar to its parent model, OSCNet. It improves edge sharpness, clearly delineates bone structures, and maintains a good balance between artifact removal and anatomical accuracy. However, a key difference is the stronger presence and introduction of dark metal artifacts around high-intensity areas near the implants. Although the model successfully reconstructs and preserves the overall anatomical structures, the regions immediately adjacent to the metal implants are notably affected. This issue is especially pronounced around the implant on the right side of the image, where two separate metal structures interact. Here, the model struggles more significantly compared to its performance around the implant on the left side. Additionally, while the metal implants appear as dark grey regions, a slight brighter halo is visible around them.

Overall, while it is difficult to identify a single absolute best performer, OSCNet, OSCNet+, and ACDNet stand out for their strong balance between artifact suppression and structural preservation. They capture the complexity and fidelity of soft tissues effectively while minimizing residual artifacts, outperforming earlier methods such as InDuDoNet, InDuDoNet+ and DICDNet in clarity and anatomical detail preservation.

Chapter 7

Conclusions and Future Work

7.1 Discussion and Conclusion

This thesis undertook a systemic evaluation of the effectiveness of various deep learning architectures for metal artifact reduction (MAR) in CT imaging. The primary motivation was to address the significant challenges posed by metal-induced artifacts, which degrade image quality, compromise diagnostic accuracy, and complicate treatment planning. The research aimed to compare different deep learning approaches specifically dual-domain (InDuDoNet, InDuDoNet+) and image-domain (DICDNet, ASCNet, OSCNEt and OSCNet+) models to identify their respectve strenghts and limitations in clinical contexts.

The experimental results presented in Chapter 6 demonstrated that deep learning models offer substantial improvements over traditional MAR techniques. Quantitative analysis using PSNR and SSIM metrics (Table 6.1) confirmed the reproducibility of results reported in prior studies and highlighted the high fidelity achieved by these models. Qualitative analysis across small, medium, and large metallic implants (Figures 6.1 and 6.2) provided further insights into the model's performance nuances.

Key findings indicate that while all evaluated models achieved significant artifact reduction, their effectiveness varied depending on the artifact severity and the specific model architecture. Dual-domain methods like InDuDoNet and InDuDoNet+ effectively incorporated CT imaging physics but sometimes struggled with residual artifacts or slight oversmoothing, particularly with larger implants. Image-domain models showed considerable promise. DICDNet offered clean artifact removal but occasionally exhibited excessive smoothing, potentially losing fine anatomical details. ACDNet, OSCNet, and particularly OSCNet+ stood out for their ability to robustly suppress complex artifacts while preserving anatomical structures and tissue contrast effectively across different implant sizes. The adaptive nature of ACDNet and the orientation-shared mechanism in OSCNet appear particularly beneficial for handling the complex, non-local patterns characteristic of metal artifacts.

In conclusion, this study confirms the potential of deep learning, especially advanced image-domain networks like ACDNet, OSCNet, and OSCNet+, to significantly enhance CT image quality in the presence of metallic implants. These methods demonstrate a superior balance between artifact suppression and structural preservation compared to earlier models, fulfilling the objective of identifying more effective MAR solutions.

7.1.1 Limitations

Despite the promising results, this research has several limitations that should be acknowledged. Firstly, the evaluation relied heavily on a synthetically generated dataset (SynDeepLesion), created using simulations based on the Deep Lesion dataset. While this approach allows for controlled comparisons with ground truth data, synthetic artifacts may not fully capture the complexity and variability of artifacts encountered in diverse clinical scenarios and real patient data.

Secondly, the selection of deep learning models, while representative of state-of-the-art approaches, is not exhaustive. Other architectures or emerging techniques might offer different performance characteristics. The semi-supervised "Stay in the Middle" approach, mentioned in the methodology, was not included in the final experimental comparison presented (***DECIDE WHAT TO DO HERE. EITHER MENTION THIS OR REMOVE THE SEMI-SUPERVISED SECTION OF THE PAPER), representing a gap in evaluating methods that handle limited labeled data.

Thirdly, the evaluation metrics used (PSNR, SSIM), while standard in image processing, primarily assess pixel-level fidelity and structural similarity. They may not fully correlate with clinical diagnostic utility or perceptual quality evaluated by radiologists. More detailed evaluation involving clinical experts would be necessary to determine the real-world value of MAR improvements.

Finally, the computational resources and the time required for training these deep learning models are substantial, which could pose a barrier to widespread clinical implementation, particularly in settings with limited computational infrastructure. Furthermore, the practical deployment, ongoing management, and maintenance of these complex models require expertise in areas such as data science and AI engineering, which typically falls outside the scope of clinicians and radiologists. This requires careful integration into existing clinical workflows and potentially requires dedicated technical support teams to ensure reliable and effective operation.

7.1.2 Future Work

Building upon the findings and limitations of this thesis, several avenues for future research emerge. A crucial next step is the validation of the top-performing models (ACDNet, OSCNet, OSCNet+) on larger and more diverse clinical datasets containing real metal artifacts from various types of implants and anatomical locations. This would provide a more robust assessment of their generalization capabilities.

Further research could explore hybrid MAR approaches that combine the strengths of dual-domain and image-domain methods, or integrate physics-based priors more deeply into network architectures to enhance interpretability and robustness. Investigating unsupervised or self-supervised learning techniques for MAR remains a promising direction, potentially reducing the reliance on large paired datasets, which are often difficult and expensive to obtain.

Exploring the impact of improved MAR on downstream clinical tasks beyond image quality assessment, such as automated disease detection, or quantitative analysis of tissues near implants, would be valuable. Additionally, optimizing model efficiency for potential

real-time applications in interventional radiology or faster integration into routine diagnostic workflows warrants investigation.

7.1.3 Clinical Implications

The successful reduction of metal artifacts demonstrated by the deep learning models evaluated holds significant potential for clinical practice. Improved MAR methods, particularly those like the ACDNet and OSCNet models, which excel in preserving anatomical detail while suppressing artifacts, could substantially enhance the diagnostic utility of CT scans in patients with metallic implants. According to clinical literature, metal artifacts frequently obscure critical structures, hindering accurate diagnosis and treatment decisions.

By providing a clearer visualization of the tissues and bones surrounding the implants, these advanced MAR techniques could lead to improved diagnostic confidence, more accurate delineation of tumors or other pathologies, and potentially improved efficacy and safety in radiation therapy planning. Although further clinical validation is essential before widespread adoption, the technology is progressing at a significant pace, and it may be a matter of time before these methods are implemented in clinical settings.

7.1.4 Ethical Aspects

The application of deep learning in medical imaging requires careful consideration of ethical aspects. Data privacy is paramount; this study utilized large-scale, anonymized datasets like Deep Lesion, adhering to standards for patient data protection. However, future work involving clinical data must ensure rigorous anonymization and compliance with data protection regulations (e.g., GDPR, HIPAA).

Potential biases in algorithms represent another ethical challenge. Models trained predominantly on specific populations or scanner types might perform differently on underrepresented groups, potentially exacerbating health disparities. Ensuring fairness and robustness across diverse patient demographics and imaging equipment is crucial. Transparency and interpretability of MAR models, while improving, remain areas for development, particularly for safety-critical applications in healthcare. Responsible development and deployment require ongoing validation, monitoring for unintended consequences, and clear communication about the capabilities and limitations of these AI-driven tools to clinicians and patients.

Bibliography

Abdoli, M., Dierckx, R. A. J. O., & Zaidi, H. (2012). Metal artifact reduction strategies for improved attenuation correction in hybrid PET/CT imaging. *Medical Physics*, *39*(6), 3343–3360. https://doi.org/10.1118/1.4709599

Arjovsky, M., Chintala, S., & Bottou, L. (2017, December 6). Wasserstein GAN. https://doi.org/10.48550/arXiv.1701.07875

Boas, F. E., & Fleischmann, D. (2012). CT artifacts: Causes and reduction techniques. *Imaging in Medicine*, *4*(2), 229–240. https://doi.org/10.2217/iim.12.13

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014, June 10). Generative adversarial networks. https://doi.org/10.48550/arXiv.1406.2661

Gupta, V. (2017). *Understanding feedforward neural networks*. https://learnopencv.com/understanding-feedforward-neural-networks/ (accessed: 20.12.2024).

Huang, Z., Zhang, G., Lin, J., Pang, Y., Wang, H., Bai, T., & Zhong, L. (2022). Multi-modal feature-fusion for CT metal artifact reduction using edge-enhanced generative adversarial networks. *Computer Methods and Programs in Biomedicine*, *217*, 106700. https://doi.org/10.1016/j.cmpb.2022.106700

Hussein, S., & Ali, K. (2022). Semantic segmentation of aerial images using u-net architecture. *Iraqi Journal for Electrical and Electronic Engineering*, *18*(1), 58–63. https://doi.org/10.37917/ijeee.18.1.7

Kawahara, D., Ozawa, S., Yokomachi, K., Higaki, T., Shiinoki, T., Saito, A., Kimura, T., Nishibuchi, I., Takahashi, I., Takeuchi, Y., Imano, N., Kubo, K., Mori, M., Ohno, Y., Murakami, Y., & Nagata, Y. (2019). Metal artifact reduction techniques for single energy CT and dual-energy CT with various metal materials. *BJR | Open*, *1*(1), bjro.20180045. https://doi.org/10.1259/bjro.20180045

Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., & Shen, D. (2018). Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, *65*(12), 2720–2730. https://doi.org/10.1109/TBME.2018.2814538

Nielsen, M. (2015). *Neural networks and deep learning*. Determination Press. http://neuralnetworksanddeeplearning.com

Schofield, R., King, L., Tayal, U., Castellano, I., Stirrup, J., Pontana, F., Earls, J., & Nicol, E. (2020). Image reconstruction: Part 1 – understanding filtered back projection, noise and image acquisition. *Journal of Cardiovascular Computed Tomography*, *14*(3), 219–225. https://doi.org/10.1016/j.jcct.2019.04.008

Selles, M., Slotman, D. J., Van Osch, J. A., Nijholt, I. M., Wellenberg, R., Maas, M., & Boomsma, M. F. (2023). Is AI the way forward for reducing metal artifacts in CT? development of a generic deep learning-based method and initial evaluation in patients with sacroiliac joint implants. *European Journal of Radiology*, *163*, 110844. https://doi.org/10.1016/j.ejrad.2023.110844

Stiller, W. (2018). Basics of iterative reconstruction methods in computed tomography: A vendor-independent overview. *European Journal of Radiology*, *109*, 147–154. https://doi.org/10.1016/j.ejrad.2018.10.025

Thibault, J.-B., Sauer, K. D., Bouman, C. A., & Hsieh, J. (2007). A three-dimensional statistical approach to improved image quality for multislice helical CT. *Medical Physics*, *34*(11), 4526–4544. https://doi.org/10.1118/1.2789499

Vellarackal, A. J., & Kaim, A. H. (2021). Metal artefact reduction of different alloys with dual energy computed tomography (DECT). *Scientific Reports*, *11*(1), 2211. https://doi.org/10.1038/s41598-021-81600-1

Wang, H., Li, Y., He, N., Ma, K., Meng, D., & Zheng, Y. (2022). DICDNet: Deep interpretable convolutional dictionary network for metal artifact reduction in CT images. *IEEE Transactions on Medical Imaging*, *41*(4), 869–880. https://doi.org/10.1109/TMI.2021.3127074

Wang, H., Li, Y., Meng, D., & Zheng, Y. (2022, June 16). Adaptive convolutional dictionary network for CT metal artifact reduction. Retrieved October 1, 2024, from http://arxiv.org/abs/2205.07471

Wang, H., Li, Y., Zhang, H., Chen, J., Ma, K., Meng, D., & Zheng, Y. (2021, September 11). InDuDoNet: An interpretable dual domain network for CT metal artifact reduction. Retrieved November 10, 2024, from http://arxiv.org/abs/2109.05298

Wang, H., Li, Y., Zhang, H., Meng, D., & Zheng, Y. (2023). InDuDoNet+: A deep unfolding dual domain network for metal artifact reduction in CT images. *Medical Image Analysis*, *85*, 102729. https://doi.org/10.1016/j.media.2022.102729

Wang, H., Xie, Q., Zeng, D., Ma, J., Meng, D., & Zheng, Y. (2023). OSCNet: Orientation-shared convolutional network for ct metal artifact learning. *IEEE Transactions on Medical Imaging*. https://doi.org/10.1007/978-3-031-16446-0_63

Wang, T., Yu, H., Lu, Z., Zhang, Z., Zhou, J., & Zhang, Y. (2023). Stay in the middle: A semi-supervised model for CT metal artifact reduction. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. https://doi.org/10.1109/ICASSP49357.2023.10095681

Yan, K., Wang, X., Lu, L., & Summers, R. M. (2017, October 10). DeepLesion: Automated deep mining, categorization and detection of significant radiology image findings using large-scale clinical lesion annotations. Retrieved November 10, 2024, from http://arxiv.org/abs/1710.01766

Yu, L., Zhang, Z., Li, X., Ren, H., Zhao, W., & Xing, L. (2021). Metal artifact reduction in 2d CT images with self-supervised cross-domain learning. *Physics in Medicine & Biology*, *66*(17), 175003. https://doi.org/10.1088/1361-6560/ac195c

Yu, L., Zhang, Z., Li, X., & Xing, L. (2021). Deep sinogram completion with image prior for metal artifact reduction in CT images. *IEEE Transactions on Medical Imaging*, *40*(1), 228–238. https://doi.org/10.1109/TMI.2020.3025064

Zhang, Y., & Yu, H. (2018). Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE Transactions on Medical Imaging*, *37*(6), 1370–1381. https://doi.org/10.1109/TMI.2018.2823083