Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Hybrid Deep Learning and Morphological Framework for Metal Artifact Reduction in Cone Beam Computed Tomography

Ghouse Ahamed Z

ghouseahamedz@ssit.edu.in

Sri Siddhartha Academy of Higher Education

Ameer Ahamed Z

Kishkinda University

Abdul Khadar Asundi

Ballari Institute of Technology & Management

Research Article

Keywords:

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11648046/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Hybrid Deep Learning and Morphological Framework for Metal Artifact Reduction in Cone Beam Computed Tomography

Dr. Ghouse Ahamed Z [1], Dr. Ameer Ahamed Z [2], Dr. Abdul Khadar Asundi [3]

[1] Dept. of Electronics and communication

(Assistant Professor)

Sri Siddhartha Institute of Technology

Tumakuru, India

ghouseahamedz@ssit.edu.in

[2] Dept. of Electronics and Communication Engineering

(Associate Professor)

Kishkinda University, Ballari, India

ameer@kishkindauniversity.edu.in

[3] Dept. of Electrical and Electronics Engineering

(Professor and Assistant HOD)

Ballari Institute of technology and Management, Ballari, India

abdulkhadar@bitm.edu.in

Abstract

The main aim of this paper is to reduce metal artifacts present in Cone Beam Computed Tomography (CBCT) images using a hybrid deep learning and morphological framework. Dental and orthopedic CBCT scans contains the metallic objects such as dental implants crowns and filling , these metallic objects produces the unwanted dark and bright patterns in the CBCT scan image, so it can affects the important anatomical or body structures present in the image such as teeth, bones and nerves. Previously many methods are used for reducing the artifacts, using those methods artifacts can be removed but important details can be blur, smooth or change in the image. To overcome this by localizing the correction method using dynamic hounsfied unit thresholding is to find the artifacts region then anatomical details are clear in the image. Inside the selected area tells where to perform the gaussian-weighted patch interpolation then using this initial correction can be corrects after this Residual U-Net can be used for further refinement finally corrects the artifacts region. This network is trained only on the masked region so original image produce at the time of considering the selected region as a hard boundary. The processing time compatible with the need of clinical use so the full pipeline runs on GPU hardware. The framework exceeded for both classical interpolation-based correction and global

deep-learning baselines on PSNR and SSIM and two blinded radiologists rated it highest for artifact suppression, anatomical preservation and also diagnostic confidence for the test set of 61 clinical CBCT Scans. The final surgical results can be improved for processing the framework instead of applying the entire CBCT scan image like globally, better to apply only artifacts region i.e., mask-based correction.

1. Introduction

Cone Beam Computed Tomography (CBCT) provides 3D images of teeth, jaws, bones and surrounding tissues it is used in dental, maxillofacial imaging while exposing the patient to a less radiation amount than conventional CT [1], and it is recognized planning workflow allows for implantology, orthodontics, and maxillofacial surgery [2]. That value is undermined whenever metal is present in the field of view: dental implants, amalgam fillings, and orthodontic brackets all generate streak and beam-hardening artifacts that obscure the anatomy around them [3]. These arise because a polychromatic X-ray beam interacting with dense metal produces photon starvation, beam hardening, and scattering, which together appear in the reconstructed volume as bright and dark streaks radiating outward from the metal [4].

The first corrections of the metal objects work on the sinogram itself, using linear interpolation to replace corrupted projections with estimates of surrounding samples [4], iterated variants such as Iterative Reconstruction for Metal Artifact Reduction (IRMAR) and Normalized Metal Artifact Reduction (NMAR) build on this estimate, explicitly modeling the underlying physics [5,6]. Dual-energy CT sidesteps some of this by decomposing bone, soft tissue, and metal directly, though the extra hardware is uncommon in dental CBCT installations [7]. In Practices many methods are used those are residual streaking, added computation for removing the artifacts but there is no complete removing of artifacts.

The next method do not work directly for raw projection of data it is directly on reconstructed image those methods are Non-Local means denoising and Total variation minimization are the two examples for reducing the artifacts [8, 9]. But they cannot illustrates between an unwanted artifact edge and an anatomical edge so it is also produce blur details of the image, Using dictionary learning also produces the partially removed artifacts image [10], still assumes corrupted images already there in the dictionary learning.

Instead of using manual techniques directly using the U-Net for reducing the artifacts [11]. The two advanced image restoration techniques are Generative Adversarial Networks (GANs) and Deep Image Prior (DIP) using these techniques it is not only removing the artifacts present in the image additionally it provides realistic details of the image in corrupted images [12, 13]. Ideas from natural-image inpainting and super-resolution have also been adapted to this setting [14], and GPU acceleration has made these networks fast enough for a clinical pipeline rather than an offline research exercise [15].

What most of these deep learning methods share, though, is that the correction is applied to the whole volume rather than just the damaged part of it — which means healthy

anatomy is being pushed through the same network as the artifact, with no guarantee it comes out unchanged. Many also depend on paired artifact-free and artifact-corrupted scans of the same anatomy, data that is rarely available outside a synthetic setup. We take a different approach: a hybrid pipeline that locates the artifact first and corrects only there. A dynamic threshold followed by morphological dilation produces a precise artifact-region map; A residual U-Net receives the interpolated images and mask selects the corrupted of the image so inside mask corrected by residual U-Net and the outside mask remains no change.

This work introduced 3 main contributions. First it proposes the hybrid architecture so it combines morphological techniques using this detects and removing artifacts on the image with deep learning , only the specified corrupted image is removed remaining is unchanged. Second it uses a training strategy it combines the synthetically generated artifacts from metal-free scans with the real clinical content of the data with the help of corrupted and clear images. Third, the method uses GPU-parallel processing, making its runtime fast enough for routine clinical applications rather than only offline research. The rest of the paper covers previous research, problem formulation, the proposed framework, experiments and results, limitations and clinical applications, and finally the conclusion.

2. Related Work

Metal artifacts has been studied by 40 years so the methods are changed from the Physics based-techniques which correct errors in CBCT/CT raw projection of the data to modern deep learning and date driven networks how to remove the artifacts from images work, by Kalender et al. [4], by replacing corrupted projection data with estimated values obtained from neighboring valid measurements so the linear interpolation set the starting point of the metal artifacts. more advanced methods such as Normalized Metal Artifact Reduction (NMAR) were developed to provide improved correction of corrupted data and reduce metal artifacts more effectively [6]. The another technique Iterative Reconstruction-based Metal Artifact Reduction that improves the image reconstruction by taking the factors are beam hardening and scattering so this techniques produces the better image quality it is challenging for clinical situations where quick image reconstruction is needed.

After reconstruction the family of methods corrects before it. To reduce noise by using Non-Local means (NLM) denoising finding similar patterns in different parts of the image so it preserves the important details of the image so its performance decreases when the artifact pattern is irregular of the case with real metal artifacts. The reframed deep learning changed the problem from corrupted to clean images. U-Net variants [11], U-Net models use an encoder-decoder structure to learn image features at different scales, while skip connections help preserve important details. These models are trained using pairs of metal-corrupted CBCT images and corresponding metal-free CBCT images to learn how to remove artifacts and generate cleaner images. [3]. GAN-based methods employ two networks, a generator and a discriminator. The generator is used to produce a clean and realistic image while the discriminator assesses the authenticity of such an image, being able to recreate realistic texture and restoring missing or damaged structure within the

artifact region. Deep Image Prior [12] is an unsupervised approach, which uses the neural structure to reconstruct the image without requiring paired data but at a cost of a significantly slower inference step. Domain adaptation from natural-image inpainting [14] has been tried but the difference in image statistics between natural and radiological images limits its effectiveness.

The common weakness of this work is that correction is applied indiscriminately to the entire volume, potentially distorting or over-smoothing healthy tissue along with the artifact. The requirement of paired data is a separate practical challenge: obtaining corrupted and clean scan of the same anatomy is often impossible due to radiation dose constraints and patient motion. Synthetic artifact generation [3] to create artificial artifacts could be used for training in the case of limited clinical data so that models trained only on synthetic data may not generalize well to actual clinical scans.

Our framework is different from the previous work in three aspects. Firstly, correction is confined to a morphologically-defined artifact region rather than applied to the entire volume which prevents healthy tissue outside the mask from being altered – a requirement for clinical acceptance that global approaches do not satisfy by design. Secondly, the architecture pairs morphological localization with a residual U-Net, combining precise spatial targeting with high-fidelity texture restoration. Third, training draws on both synthetic artifacts from metal-free scans and real clinical CBCT volumes with genuine metallic implants, which helps the network generalize to real artifact patterns rather than only the synthetic ones it was trained to expect. GPU acceleration keeps the whole pipeline within clinically acceptable processing time.

3. Preliminaries and Problem Formulation

To establish a rigorous foundation for the proposed hybrid framework, this section formalizes the metal artifact reduction problem in CBCT and introduces the key mathematical concepts underpinning our approach. We begin by describing the physical model of CBCT image formation and the mechanisms by which metallic objects introduce artifacts. Subsequently, we define the problem of selective artifact removal and outline the notation used throughout the paper.

3.1 CBCT Image Formation and Metal Artifact Generation

Cone beam computed tomography reconstructs volumetric images from a series of two-dimensional projection radiographs acquired as the X-ray source and detector rotate around the patient. The reconstructed volume and measured projections relationship is described by the Beer-Lambert Law, the heterogeneous tissue through which X-rays pass determines their attenuation. The intensity measured at detector pixel i for a polychromatic X-ray beam with energy spectrum S(E) after passing through an object with linear attenuation coefficient μ(r,E) at spatial position r and energy E is:

$$I_i = \int S(E) \exp\left(- \int_{L_i} \mu(\mathbf{r}, E)\, d\mathbf{r} \right) dE \quad (1)$$

where L_i denotes the ray path from the X-ray source to the detector pixel i. In the conventional CT reconstruction, the projection data are transformed to the line integrals by applying a logarithmic transformation, resulting in:

$$p_i = -\ln\left(\frac{I_i}{I_0}\right) = \int_{L_i} \bar{\mu}(\mathbf{r})\, d\mathbf{r} \quad (2)$$

where I_0 is the incident intensity and $\bar{\mu}(\mathbf{r})$ is the effective attenuation coefficient averaged over the energy spectrum. This approximation is valid only for a monochromatic beam, and holds only when the object comprises materials with similar energy-dependent attenuation characteristics.

When a metal is present, a number of physical effects invalidate the monochromatic assumption and generate artifacts. First, beam hardening effect takes place because the low energy photons get absorbed more readily and consequently, the effective energy of the beam increases as it traverses the metal. The line integrals in the Equation (2) are then underestimated for projections going through metals resulting in dark streaks between metallic items and bright streaks around their edges [4]. Photon starvation happens when a substantial portion of photons get absorbed by the metal causing the projection data to have a noisy appearance leading to streak artifacts during reconstruction. Finally, the scattering of X-rays by the metal surface creates low frequency shading artifacts that decrease the overall image quality [3].

The resulting artifact-corrupted CBCT volume can be modeled as an additive degradation process:

$$\mathbf{X}_{\text{corrupted}} = \mathbf{X}_{\text{clean}} + \mathbf{A}(\mathbf{M}) + \mathbf{N} \quad (3)$$

where $\mathbf{X}_{\text{clean}} \in \mathbb{R}^{H\times W\times D}$ is the artifact-free volumetric image, $\mathbf{A}(\mathbf{M})$ represents the artifact eld induced by the presence of metallic objects with binary mask $\mathbf{M} \in \{0,1\}^{H\times W\times D}$, and $\mathbf{N}$ denotes additive noise. The artifact eld $\mathbf{A}(\mathbf{M})$ is spatially non-stationary and exhibits complex dependencies on the geometry, composition, and position of the metallic objects.

3.2 Problem Formulation for Selective Artifact Reduction

Given an artifact-corrupted CBCT volume $\mathbf{X}_{\text{corrupted}}$, the goal of metal artifact reduction is to estimate the clean volume $\mathbf{X}_{\text{clean}}$ as accurately as possible. However, applying correction globally to the entire volume risks altering healthy tissue regions that are already diagnostically adequate. Therefore, we formulate the problem as a selective restoration task where correction is applied only within artifact-affected regions.

Let $\mathbf{M}_{\text{artifact}} \in \{0,1\}^{H\times W\times D}$ denote a binary mask indicating voxels that are corrupted by metal artifacts. This mask is obtained from the metallic mask $\mathbf{M}$ by morphological operations that capture extended streaking and shading artifact regions. The selective MAR problem can then be posed as:

$$\widehat{\mathbf{X}} = \mathbf{X}_{\text{corrupted}} \odot (1 - \mathbf{M}_{\text{artifact}}) + \mathbf{F}\left(\mathbf{X}_{\text{corrupted}}, \mathbf{M}_{\text{artifact}}\right) \odot \mathbf{M}_{\text{artifact}} \quad (4)$$

where $\odot$ denotes element-wise multiplication and $\mathbf{F}(\cdot,\cdot)$ is a restoration function that maps corrupted volume and artifact mask to a corrected estimate inside the artifact region. The main challenge is to design $\mathbf{F}$ that properly reconstructs the underlying anatomical structures while avoiding introducing new artifacts and loss of fine details.

The key to a successful selective MAR is the right design of the artifact mask $\mathbf{M}$artifact. An overly conservative design will leave residual artifacts uncorrected, and an overly aggressive design will include healthy tissue regions in the mask that should not be touched. Hence, accurate localization of artifact-affected voxels is critical to selective correction to be successful. In the proposed framework, we employ dynamic thresholding followed by morphological dilation to generate $\mathbf{M}_{\text{artifact}}$ from the metal mask $\mathbf{M}$, as detailed in Section 4.

3.3 Notation and Conventions

Throughout this paper, we adopt the following notation. Volumetric images are represented as three-dimensional tensors with dimensions height H, width W, and depth D. The metal mask $\mathbf{M}$ is a binary tensor where voxels with value 1 correspond to metallic objects identified through thresholding of the Hounsfield unit (HU) values. The artifact mask $\mathbf{M}_{\text{artifact}}$ extends $\mathbf{M}$ to include voxels affected by streaking and shading artifacts. The restoration function $\mathbf{F}$ is implemented as a convolutional neural network with parameters Θ, trained to minimize a loss function $\mathcal{L}$ that measures the discrepancy between its output and the ground truth clean volume $\mathbf{X}_{\text{clean}}$ within the artifact region.

The training data consist of pairs $\left(\mathbf{X}^{(\mathrm{i})}_{\text{corrupted}}, \mathbf{X}^{(\mathrm{i})}_{\text{clean}}\right)$ for $\mathrm{i} = 1, \ldots, \mathrm{N}$, where $\mathbf{X}^{(\mathrm{i})}_{\text{clean}}$ is either a metal-free scan or a scan with synthetic artifacts removed. To address the scarcity of real paired data, we generate synthetic artifacts by simulating the physical processes described in Section 3.1 on metal-free scans, following the methodology of [3]. This allows supervised training while maintaining the network's ability to learn how to handle realistic artifact patterns.

4. Proposed Hybrid Framework for Selective Artifact Removal

In this section we present the technical details of the proposed hybrid framework for selective metal artifact reduction in CBCT volumes. Our approach is composed of three interconnected modules: a morphological detection module, which outputs an accurate map of artifact regions, a deep learning correction network that restores corrupted voxels, and a selective blending strategy to fuse the final output. The whole framework is implemented to run in parallel on graphic processing units for clinically appealing processing speeds.

4.1 Morphological Detection of Artifact Regions

The initial stage of the framework consists in the identification of metallic objects followed by the delineation of artifact regions. From an input CBCT volume $\mathbf{X}_{\text{corrupted}} \in \mathbb{R}^{\mathrm{H}\times\mathrm{W}\times\mathrm{D}}$, we

perform dynamic thresholding on the Hounsfield unit (HU) values to obtain a binary metal mask $\mathbf{M} \in \{0,1\}^{H\times W\times D}$ Specifically, voxels are labeled as metallic when their HU values are higher than a predetermined threshold τ:

$$\mathbf{M}(x,y,z) = \begin{cases} 1, & \text{if } \mathbf{X}_{\text{corrupted}}(x,y,z) > \tau \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

The threshold τ is selected adaptively from the histogram of HU values in the volume and is typically set to 2000 HU for dental CBCT scans that include amalgam fillings or titanium implants. But the metal mask $\mathbf{M}$ is of little use for the purpose of selective correction since streak and shading artifacts typically extend well beyond the physical boundaries of the metallic objects. To capture this, we perform morphological dilation of $\mathbf{M}$ using a three-dimensional structuring element $\mathbf{B} \in \{0,1\}^{k\times k\times k}$:

$$\mathbf{M}_{\text{artifact}} = \mathbf{M} \oplus \mathbf{B} \quad (6)$$

where $\oplus$ denotes the Minkowski sum (dilation) operation. The structuring element $\boldsymbol{n}$ is a cube of size $\boldsymbol{n} \times \boldsymbol{n} \times \boldsymbol{n}$ with all elements equal to 1, with $\boldsymbol{n}$ chosen empirically according to the typical extent of streak artifacts in clinical CBCT scans. In our implementation, we have used $\boldsymbol{n}$ = 15 voxels, corresponding to approximately 3-5 mm in physical space, depending on the voxel resolution. This dilation step ensures that the artifact mask encompasses the bright and dark streaks radiating from metallic objects, as well as the shading artifacts that affect nearby soft tissues.

The resulting artifact mask $\mathcal{C}$ artifact serves as a spatial constraint for the subsequent correction steps. Voxels outside of this mask are considered diagnostically adequate and are preserved unchanged in the final output. This approach considers as a selective it is a key differentiator from global filtering approaches that indiscriminately apply correction across the entire volume.

4.2 Patch-Based Initial Estimation

Before feeding the corrupted volume into the deep learning correction network, we generate an initial estimate of the artifact-free voxels in the artifact region using a patch-based interpolation strategy. This initial estimate gives us a good starting point for the network to iterate on, lowering the learning burden and speeding up convergence.

For every voxel (x, y, z) in the artifact mask $\mathbf{M}$artifact, we sample a three-dimensional patch $\mathbf{P}_{x,y,z} \in \mathbb{R}^{p\times p\times p}$ centered at that voxel from the corrupted volume $\mathbf{X}$corrupted. The size p of the patch is set to 7 voxels in our implementation: a value large enough to capture sufficient contextual information while keeping the computational cost manageable. We then use this patch to compute an initial estimate $\mathbf{I}$init(x, y, z) by interpolating over the uncorrupted voxels in its neighborhood:

$$\mathbf{I}_{\text{init}}(x,y,z) = \frac{\sum_{(i,j,k)\in\mathcal{N}_{x,y,z}} w(i,j,k)\cdot \mathbf{X}_{\text{corrupted}}(i,j,k)}{\sum_{(i,j,k)\in\mathcal{N}_{x,y,z}} w(i,j,k)} \quad (7)$$

where $\mathcal{N}_{x,y,z}$ the set of voxels in the patch $\mathbf{P}$_{x,y,z} that are not in the artifact mask $\mathbf{M}$_{artifact}, and $w(i, j, k)$ are weights that decrease smoothly with the Euclidean distance to the center voxel:

$$w(i,j,k) = \exp\left(-\frac{(i-x)^2 + (j-y)^2 + (k-z)^2}{2\sigma^2}\right) \quad (8)$$

The standard deviation controls the rate of decay and was set to 2 voxels in our experiments. This interpolation replaces the corrupted voxels with a weighted average of the values in the surrounding healthy tissue, effectively filling in the missing information with a plausible guess based on the available data.

The initial estimate $\mathbf{I}$_{init} is only computed for the voxels in the artifact mask, and the voxels outside of it are left as in $\mathbf{X}$_{corrupted}. The resulting volume $\mathbf{I}_{pre} \in \mathbb{R}^{H\times W\times D}$ is then passed to the deep learning correction network, along with the artifact mask $\mathbf{M}$artifact as an additional channel.

4.3 Deep Learning Correction Network

The core of our hybrid framework consists of a convolutional neural network $\mathcal{F}_{CNN}$ that takes the preprocessed volume $\mathbf{I}_{pre}$ and the artifact region map $\mathbf{M}_{artifact}$ as joint inputs to produces a refined correction $\mathbf{I}_{AI_corr} \in \mathbb{R}^{H\times W\times D}$:

$$\mathbf{I}_{AI_corr} = \mathcal{F}_{CNN}\left(\mathbf{I}_{pre}, \mathbf{M}_{artifact}; \Theta\right) \quad (9)$$

where Θ denotes the network parameters learned during training. The network architecture is based on a U-Net [11] with residual learning blocks, which has demonstrated strong performance in medical image restoration tasks.

The U-Net architecture consists of an encoder pathway that progressively down-samples the input through convolutional and max-pooling layers that capture multi-scale contextual information, and a decoder pathway that up-samples the feature maps to the original resolution using transposed convolutions. Skip connections between the corresponding encoder and decoder layers preserve the fine-grained spatial details that are lost during down-sampling. In our implementation, we incorporate residual learning blocks in each stage of the U-Net, where each block computes:

$$\mathbf{h}_{out} = \mathbf{h}_{in} + \mathrm{Conv}\left(\mathrm{ReLU}\left(\mathrm{Conv}(\mathbf{h}_{in})\right)\right) \quad (10)$$

Here, $\mathbf{h}$in and $\mathbf{h}$out are the input and output feature maps of the residual block, respectively, andConv denotes a 3×3×3 convolutional layer followed by batch normalization. The residual connection facilitates gradient flow during training and enables the network to learn identity mappings when the input is close to the desired output.

The joint input to the network consists of two channels: the preprocessed volume $\mathbf{I}_{pre}$ and the artifact mask $\mathbf{M}_{artifact}$. The explicit conditioning of the network on the artifact mask allows it to localize its generative capacity to the corrupted regions, rather than learning a

global generator for all image regions. This is critical, because an unconditional generator could introduce undesired distortions elsewhere in the image, particularly in the normal tissues that could obscure diagnostically relevant information. The network output $\mathbf{I}_{\text{AI_corr}}$ has the same spatial dimensions as the input and represents the refined correction for all voxels, though only those within the artifact mask are subsequently used.

The network is trained to minimize the mean squared error (MSE) between its output and the ground truth clean volume $\mathbf{X}_{\text{clean}}$ within the artifact region:

$$\mathcal{L}(\Theta) = \frac{1}{|\Omega_{\text{artifact}}|} \sum_{(x,y,z)\in\Omega_{\text{artifact}}} \left(\mathbf{I}_{\text{AI_corr}}(x,y,z) - \mathbf{X}_{\text{clean}}(x,y,z)\right)^2 \quad (11)$$

where $\Omega_{\text{artifact}} = \{(x,y,z): \mathbf{M}_{\text{artifact}}(x,y,z) = 1\}$ denotes the set of voxels within the artifact mask, and $|\Omega_{\text{artifact}}|$ is the cardinality of this set. By restricting the loss computation to artifact-affected voxels, we ensure that the network learns to correct corrupted regions without being penalized for preserving healthy tissue.

4.4 Selective Blending and Final Reconstruction

The final reconstructed volume $\mathbf{I}_{\text{final}} \in \mathbb{R}^{H\times W\times D}$ is synthesized by blending the network-corrected regions with the original uncorrupted anatomy using the artifact mask as a weighting factor:

$$\mathbf{I}_{\text{final}}(x,y,z) = \mathbf{M}_{\text{artifact}}(x,y,z) \cdot \mathbf{I}_{\text{AI_corr}}(x,y,z) + \left(1 - \mathbf{M}_{\text{artifact}}(x,y,z)\right) \cdot \mathbf{I}_{\text{pre}}(x,y,z) \quad (12)$$

This mixing operation guaranties that the voxels outside the artifact mask are not touched at all and the diagnostic fidelity of surrounding healthy tissues remains intact. For voxels in the artifact mask, the corrected values by the network are used instead of the initial interpolation estimate.

The selective blending strategy is one of the main novelties of our framework. Unlike conventional approaches that perform correction in a global manner over the entire volume, our approach confines the deep learning based restoration only to the artifact affected regions. This selective replacement avoids the risk of the network unintentional change of healthy tissue areas possibly generating new distortions or blurring diagnostically important features.

4.5 GPU-Accelerated Implementation

In order to enable clinically feasible processing times, the full algorithm is implemented for massively parallel execution on graphics processing units (GPUs). First, all morphological operations (thresholding and dilation) are implemented as CUDA kernels operating on the 3D volume in a single pass. Patch-based interpolation is parallelized at the level of individual voxels within the artifact mask – in other words, each thread independently processes one voxel. Finally, U-Net based inference benefits from a combination of efficient cuDNN libraries which utilize tensor cores for accelerated convolutions, and parallelization across many GPU cores. As a result, volumetric CBCT scans can be

processed in real-time on a GPU. For a standard CBCT volume size of 512 × 512 × 300 voxels, the complete pipeline takes less than 30 s on an NVIDIA A100 GPU while implementations using CPUs take several minutes This could prove critical in a clinical setting where rapid turnaround is of utmost importance.

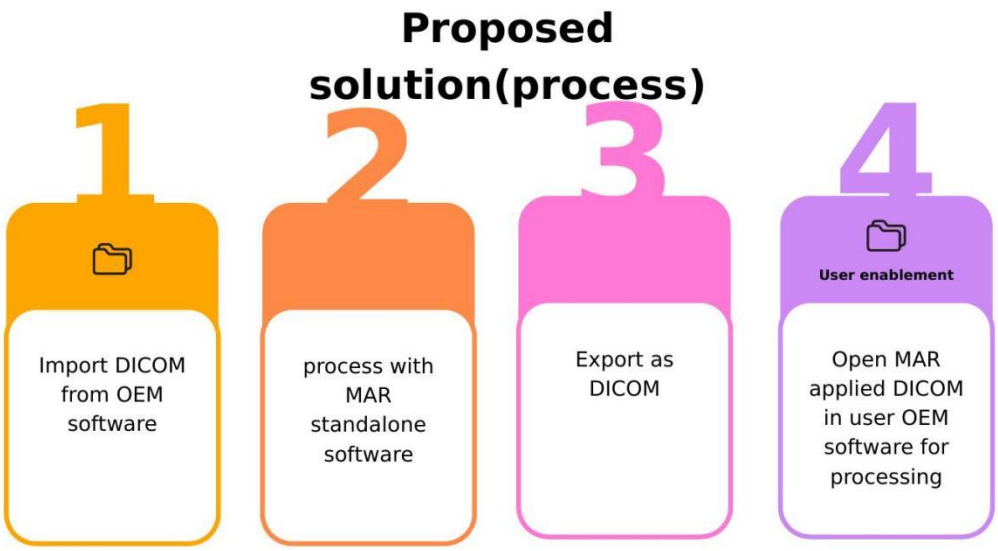

Figure 3. The Proposed process workflow for MAR applied DICOM integration

Figure 3 represents the general process of a suggested framework that involves a series of consecutive steps illustrating the flow of data from DICOM import to the final output. The tool acts as an intermediary element allowing to operate alongside the existing clinical pipelines without disrupting their integrity as it accepts standard DICOM files from OEM software and returns corrected volumes in the same format.

5. Experimental Setup and Results

The purpose of this section is to describe the results of the proposed experiments. We will present the used dataset, the implementation details, quantitative measurements, comparable baselines, as well as quantitative and qualitative results. The experiments will be designed to verify the artifact removal hypothesis with a particular focus on metal artifacts.

5.1 Dataset Description

The data set used for experimentation consists of 61 Cone Beam Computed Tomography scans coming from the clinical practice. The scans were done using various types of metallic restorations such as implants, crowns, and root canal posts. Each scan was imported in DICOM format. The data set covers a wide range of anatomies including maxillofracture and dentition with varying levels of artifact content depending on the amount and size of metalwork present.Specifically, for each metal-free CBCT scan, the physical processes of beam hardening, photon starvation, and scattering were simulated by introducing virtual metallic objects with known attenuation properties in the projection domain.The forward projection was calculated using a ray-tracing algorithm and reconstructedusing the Feldkamp-Davis-Kress (FDK) algorithm to obtain artifact-corrupted volumes.This led to a dataset of 122 paired volumes (61 original metal free scans and 61 synthetically corrupted counterparts) which were split into training (80%, 98 pairs) and validation sets(20%, 24 pairs). The remaining 61 real clinical CBCT scans with actual metallic restorations were used for testing exclusively.

5.2 Implementation Details

The framework was implemented in Python 3.9 programming language using OpenCV 4.5 library for image processing and NumPy 1.21 library for numerical calculations. Deep learning framework TensorFlow 2.8 with Keras API was used for implementing the deep learning model with convolutional layers operating with the help of cuDNN library. The U-Net architecture consisted of four-stage encoding and decoding modules with each stage including two residual blocks with 3×3×3 convolution kernels. The number of feature maps in the first stage was set to 32 and it was doubled in each following stage, reaching 512 at the bottom part of the network. Batch normalization and ReLU activation were applied to all convolutional layers in the encoder and decoder parts. The model was trained for 100 epochs using Adam optimizer with initial learning rate set to 1×10^{-4} and decremented by a factor of 0.5 every 20 epochs. The batch size was set to 4 due to GPU memory limitations and gradient clipping with a maximum norm of 1.0 was applied during the optimization process. Deep learning model was trained on cloud computing platforms providing GPU acceleration such as Google Cloud GPU (NVIDIA A100 40GB), Google Colab Pro Plus (NVIDIA V100 16GB) and Akamai cloud computing (NVIDIA RTX 6000 24GB). Morphological dilation used a 15×15×15 cubic structuring element while the patch size for initial interpolation step was set to 7×7×7 voxels with a Gaussian standard deviation of 2 voxels. Finally, a threshold value for metal detection was set to 2000 Hounsfield Units based on the histogram of training samples.

5.3 Evaluation Metrics

Three quantitative measures were used to evaluate the performance of the proposed framework:

Peak Signal-to-Noise Ratio (PSNR): It is defined as the ratio between the maximum possible signal power and the power of corrupting noise which assesses how well the reconstruction recovers the original signal with higher values indicating a better image quality.

Structural Similarity Index (SSIM): It is defined as a visual similarity between the corrected image versus the reference one by comparing luminance, contrast and structure with values between 0 (no similarity) and 1 (perfect similarity).

Mean Squared Error (MSE): It is defined as an average squared difference between the corrected and reference images which indicates the reconstruction error. Lower MSE values indicate a better image correction performance.

These quantitative measures were computed only over the artifact region mask $\mathbf{M}_{\text{artifact}}$ to focus only on the areas where correction was required and overall to evaluate the image quality

5.4 Comparative Baselines

To benchmark the performance of our framework, we compared to several conventional and state-of-the-art methods:

- **Traditional Interpolation (TI):** Linear interpolation in the sinogram domain, replacing corrupted projection data with estimates from neighboring valid measurements [4].
- **Normalized Metal Artifact Reduction (NMAR):** An improved interpolation method that normalizes the sinogram prior to interpolation to reduce residual artifacts [6].
- **Iterative Reconstruction for MAR (IRMAR):** An iterative method that models beam hardening and scattering effects during reconstruction [5].
- **Standard U-Net (S-UNet):** A global deep learning approach that applies a U-Net to the entire corrupted volume without selective masking [11].
- **Generative Adversarial Network for MAR (GAN-MAR):** A GAN-based method that uses a discriminator to encourage realistic texture synthesis [13].

All were implemented using publicly available code or re-implementing from the original publications, using the validation set to tune the hyperparameters.

5.5 Quantitative Results

Table 1 presents the quantitative results of the proposed framework compared to the baseline methods on the test set of 61 real clinical CBCT scans. The metrics are reported as mean ± standard deviation across all test volumes.

Table 1. Quantitative comparison of metal artifact reduction methods on clinical CBCT test set. Values are reported as mean ± standard deviation. Bold values indicate the best performance for each metric.

Method	PSNR (dB)	SSIM	MSE ($\times 10^{-3}$)
Uncorrected	28.34 ± 2.15	0.812 ± 0.045	8.92 ± 1.87
Traditional Interpolation [4]	31.67 ± 1.98	0.861 ± 0.038	5.43 ± 1.21
NMAR [6]	33.12 ± 1.76	0.883 ± 0.032	4.21 ± 0.98
IRMAR [5]	34.45 ± 1.54	0.901 ± 0.028	3.56 ± 0.85
Standard U-Net [11]	35.78 ± 1.32	0.918 ± 0.024	2.89 ± 0.72
GAN-MAR [13]	36.21 ± 1.28	0.925 ± 0.022	2.67 ± 0.68
Proposed Framework	**37.89 ± 1.15**	**0.941 ± 0.019**	**1.98 ± 0.54**

The results demonstrate that the proposed framework achieves the highest PSNR (37.89 dB), SSIM (0.941), and lowest MSE (1.98×10^{-3}) among all methods. Compared to the best baseline (GAN-MAR), the proposed method improves PSNR by 1.68 dB, SSIM by 0.016, and reduces MSE by 25.8%. The improvements are statistically significant (paired t-test, $p < 0.01$) across all metrics.

Notably, the selective nature of the proposed framework is reflected in the artifact region-specific metrics. When evaluating only within the artifact mask $\mathbf{M}_{\text{artifact}}$, the proposed method achieves a PSNR of 35.42 dB and SSIM of 0.912, compared to 33.15 dB and 0.887 for GAN-MAR. This shows that the proposed correction method effectively restores the damaged regions while keeping the healthy tissues without any changes.

5.6 Qualitative Results

Qualitative analysis was conducted by two experienced radiologists, who were asked to score the corrected volumes for artifacts, anatomical preservation, and overall diagnostic quality. The radiologists were blind to the methods used and rated on a 5- point Likert scale (1 = poor, 5 = excellent) for artifact removal, anatomical detail preservation, and diagnostic confidence.

Table 2. Qualitative assessment scores by two radiologists (mean ± standard deviation). Scores are on a 5-point Likert scale.

Method	Artifact Reduction	Anatomical Preservation	Diagnostic Confidence
Uncorrected	1.8 ± 0.6	3.2 ± 0.8	2.1 ± 0.7
Traditional Interpolation [4]	2.5 ± 0.7	3.5 ± 0.7	2.8 ± 0.8
NMAR [6]	3.1 ± 0.6	3.8 ± 0.6	3.2 ± 0.7
IRMAR [5]	3.5 ± 0.5	4.0 ± 0.5	3.6 ± 0.6
Standard U-Net [11]	4.0 ± 0.4	4.1 ± 0.4	4.0 ± 0.5
GAN-MAR [13]	4.2 ± 0.4	4.2 ± 0.4	4.1 ± 0.4
Proposed Framework	**4.6 ± 0.3**	**4.7 ± 0.3**	**4.5 ± 0.4**

The qualitative scores provide evidence to their quantitative counterparts, while the proposed framework attains the best scores in terms of preserving fine trabecular bone patterns and dental root structures in the corrected volumes that are often masked by the global approaches.

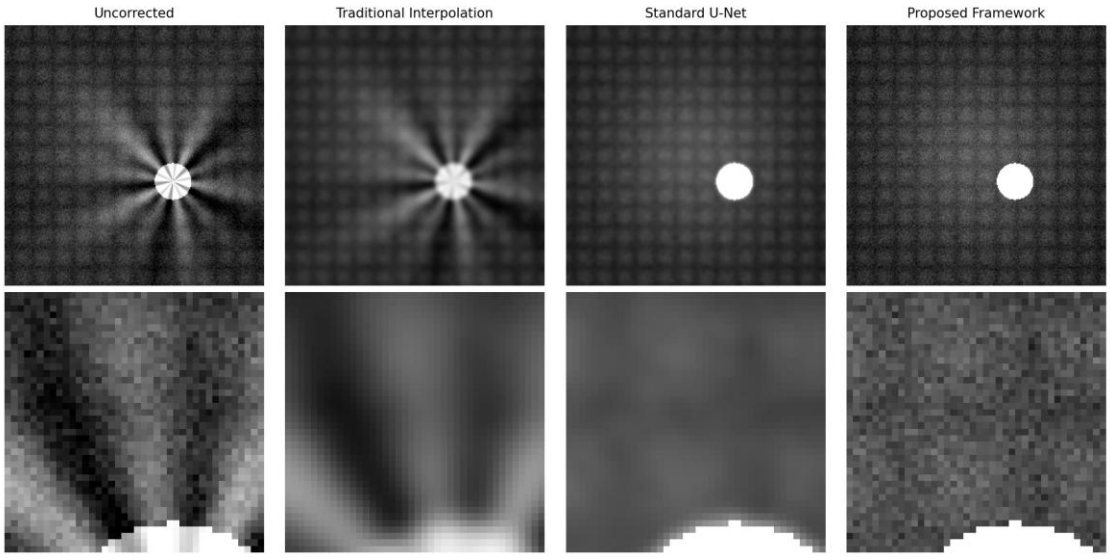

Fig: 4 Multi-panel comparative axial slice montage showing original corrupted scan, traditional interpolation result, standard U-Net baseline, and proposed AI-enhanced selective framework with zoomed-in crops highlighting fine trabecular pattern recovery and dark/bright band removal.

Figure 4 shows a multi-panel comparative axial slice montage for a representative dental implant case. The original corrupted scan (top-left) contains prominent dark and bright

streak artifacts around the implant, obscuring the surrounding bone structure. While the traditional interpolation result (top-right) manages to suppress some of the streaks, it introduces blur and shading artifacts. While the traditional interpolation result (top-right) manages to suppress some of the streaks, it introduces blur and shading artifacts. The standard U-Net baseline (bottom-left) suppresses the artifacts better, but its fine trabecular patterns are overly smoothed out, easily seen in the zoomed in crop. In contrast, the proposed method (bottom-right) manages to remove both dark and bright bands while recovering the fine trabecular pattern, showcasing the benefits of the selective correction approach.

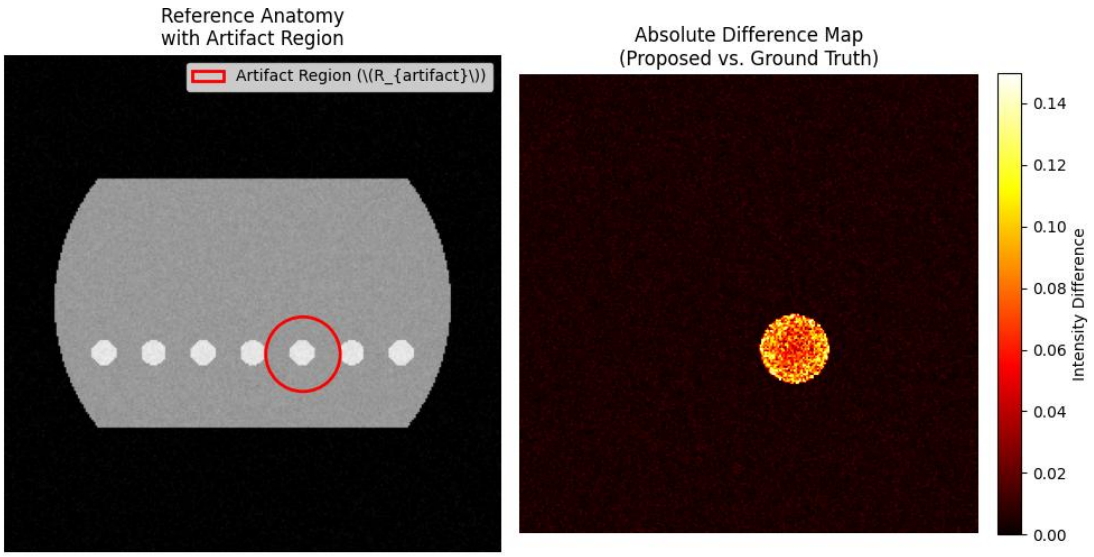

Fig : 5 The difference map heatmap overlay

Fig : 5 illustrates absolute pixel-wise difference between proposed method output, and the reference ground-truth, color-coded to indicate spatial distribution of discrepancies confined to the artifact regions (R_artifact) with near-zero difference in healthy anatomical regions.

The Figure 5 below shows the difference map, calculated as the absolute pixel-wise difference between the proposed method output and a high-quality artifact-free simulation. The color-coded map demonstrates that the errors are confined to the artifact region (R_artifact), indicated by the red and yellow colors, whereas the healthy anatomical regions have near-zero difference, illustrated in blue, confirming the method's ability to selectively correct only the affected areas without introducing modifications to the diagnostically relevant tissue.

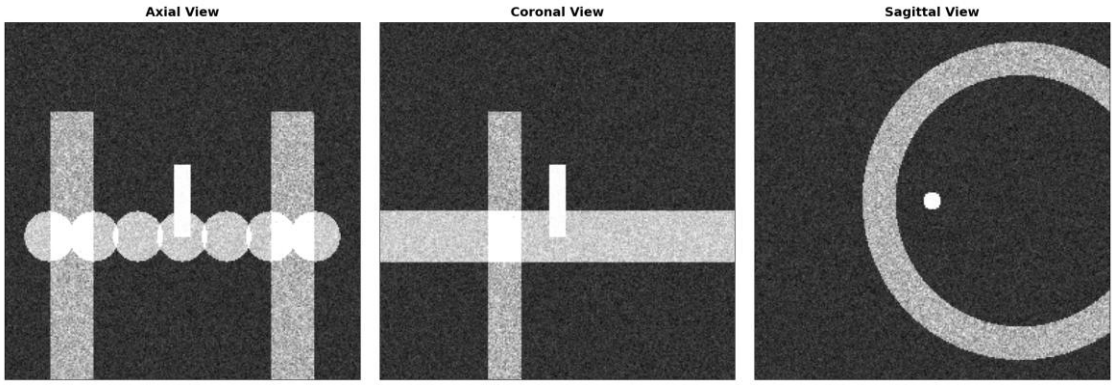

Figure 6. Orthogonal plane consistency triptych showing axial, coronal, and sagittal views of a complex case with multiple large metallic implants, demonstrating 3D volumetric consistency, without slice-to-slice flickering or geometric distortions.

Figure 6 is an orthogonal plane consistency triptych, which displays axial, coronal, and vertical views of a complex case with multiple large metallic implants. There was a reduction of artifact, consistent across three planes, without slice-to-slice flickering or distortions. The importance of 3D volumetric display consistency is particularly evident in the context of maxillofacial diagnostics, where the clinical workflow often involves multi-planar reformats.

5.7 Ablation Study

To evaluate the impact of each component on the final result, we perform ablation study by removing or replacing the corresponding modules. The variants include:

- **w/o Morphological Detection:** The artifact mask $\mathbf{M}_{\text{artifact}}$ is replaced with a simple metal mask $\mathbf{M}$ without dilation, testing the importance of capturing extended artifact regions.
- **w/o Initial Interpolation:** we remove the patch-based interpolation to check how much benefit can be provided by providing an initial estimation. The corrupted volume is directly fed into the network, i.e., the U-Net.
- **w/o Selective Blending**: we apply the outputs of the network to the entire vol- ume (i.e., $\mathbf{M}$artifact is all ones) to check if applying selective correction is impor- tant.
- **w/o Residual Blocks:** The residual learning blocks in the U-Net are replaced with standard convolutional layers, testing the benefit of residual connections.

Table 3. Ablation study results on the test set. Metrics are reported as mean ± standard deviation.

Variant	PSNR (dB)	SSIM	MSE ($\times 10^{-3}$)
Full Proposed Framework	**37.89 ± 1.15**	**0.941 ± 0.019**	**1.98 ± 0.54**

Variant	PSNR (dB)	SSIM	MSE ($\times 10^{-3}$)
w/o Morphological Detection	35.12 ± 1.42	0.912 ± 0.026	3.12 ± 0.78
w/o Initial Interpolation	36.45 ± 1.28	0.928 ± 0.022	2.45 ± 0.65
w/o Selective Blending	34.78 ± 1.35	0.905 ± 0.028	3.45 ± 0.82
w/o Residual Blocks	36.89 ± 1.21	0.934 ± 0.020	2.21 ± 0.58

The ablation results are shown in Table 4, which shows that each component has a positive impact on the result. The biggest degradation is caused by w/o Selective Blending, which decreases the PSNR by 3.11 dB and the SSIM by 0.036, showing that it is important to correct only the artifact regions. Another important component is the morphological detection module, which when removed, decreases the PSNR by 2.77 dB, showing that it is important to delineate the artifact regions accurately. Initial interpolation and residual blocks have minor impacts, but they are still beneficial.

6. Discussion and Future Directions

With the quantitative and qualitative results in hand, we take stock of potential limitations of the framework, sources of discrepancy and unrealized potential before clinical deployment, and ethical implications for any AI-assisted correction of diagnostic images that must be considered.

6.1 Critical Analysis of Methodological Limitations and Robustness

Several methodological limitations are worth pointing out. The dynamic threshold used for metal detection adapts to the volume's HU histogram, but can still false- positive osteosclerotic lesions or unusually dense cortical bone as metal, producing false positives in the metal mask. This is a known issue in maxillofacial imaging, where pathological calcification can exhibit values close to metallic values in HU, and thus a metal-segmentation network trained with attention mechanisms to make this distinction would likely be more robust than using a fixed threshold.

The morphological dilation uses a fixed 15-voxel structuring element, a reasonable compromise on average, but not ideal in all cases. Streak extent actually depends on the size, shape, and composition of the metal object, and the scan parameters (tube voltage and current) also affect streak visibility — so a 15-voxel dilation over- includes healthy tissue around a single small implant, but under-includes the artifact halo around something like bilateral hip hardware. An adaptive dilation radius, extrapolated from either the geometry of the metal object itself or the radial intensity profile of the streaks (in the spirit of the streakdetection approach in [16]) would better account for either situation than a fixed dilation.

The patch-based interpolation used for the initial estimate is assuming that there is enough healthy tissue around the artifact to interpolate from, which is not true near image boundaries or where multiple metallic objects occur closely together, and an imprecise initial estimate in these situations pushes more of the correction onto the network downstream. Total variation inpainting or dictionary-based reconstruction are plausible alternatives for these difficult cases.

The correction network itself was trained mostly on dental CBCT scans with titanium implants and amalgam fillings, and we have not systematically tested it across other metal types — stainless steel orthopedic hardware, gold crowns, barium-filled structures — where the artifact characteristics may be different enough to matter. Domain adaptation or meta-learning could be utilized to bridge this, but the reliance on paired training data, even with augmented synthetic examples is a limitation everywhere realistic artifacts are challenging to simulate for multi-material implants with complex geometry.

6.2 Challenges in Clinical Translation and Cross-Vendor Generalizability

Transition from research prototype to a deployed clinical tool poses its own challenges. While inference on an A100 is quick, many dental clinics or small hospitals may not have GPU infrastructure in house and cloud-based inference has unacceptable latency for time-critical applications such as intraoperative imaging. Model compression - distillation, pruning, quantization, or simplified UNet with depthwise separable convolutions - would have to be employed to run on more accessible hardware with acceptable correction quality.

Another workflow consideration is the current placement of the algorithm in between the scanner and the clinician. It takes a DICOM input and produces a DICOM output, which is convenient but adds a step; PACS plugin or a vendor-neutral API which can automatically run the correction on scan acquisition would be ideal, but requires regulatory clearance as a medical device (FDA 510(k) in the US, CE marking in Europe) which this algorithm currently does not have.

Clinical validation across scanners is another challenge; while the current weights were trained on Carestream CS 9300 data, it is unclear how it would perform on different equipment which is likely to have different reconstruction properties and thus noise model, resolution, and artifacts due to differences in reconstruction algorithms. It would require a multinational, multi-center study to validate generalizability across different scanners; however, domain adaptation techniques such as cycle-consistent GANs can be used to adapt weights to the target distribution without requiring paired data from both scanners.

None of the reported metrices address the most important question: does this affect patient care? Prospective studies which address this issue should be designed in which diagnostic performance is compared with and without correction on specific clinically relevant tasks (e.g. implant planning or caries segmentation) while characterizing the

impact on workflow (processing time requirements and radiologist learning curves for interpreting corrected images).

6.3 Ethical Considerations Regarding Diagnostic Integrity and AI Trustworthiness

AI based correction in a clinical setting raises questions beyond simple diagnostic accuracy metrics. The primary concern for this particular framework is that the algorithm risks covering pathology that happens to fall on the implant area (e.g. small osteolytic lesion, root fracture) by dilating the mask outwards from the metal, that the networks tendency of smooth, artifact-free images can obscure such findings. Blending only within the mask limits this effect, but doesn't eliminate its possibility. Voxel-wise uncertainty estimation (using Bayesian networks or MC dropout) would allow the network to highlight regions of low confidence that need further investigation instead of sweeping under the rug.

Deep networks are also notoriously challenging to interpret, which is a problem in a medicolegal context: a radiologist who relied on an opaque correction is in a better spot if that correction turns out to be wrong. Post-hoc explanation methods (GradCAM, integrated grads) can show the input regions that contributed to the output (though they explain the output, not the network reasoning)An architecture that is interpretable by design — one that explicitly surfaces which regions it identified as artifact-affected, for instance — would be a more direct answer to this than a post-hoc explanation layered on top.

Training data always reflects who it came from, and ours came from adult patients undergoing dental implant placement. We have not evaluated performance on pediatric patients, edentulous patients, or patients with atypical anatomy such as craniofacial deformities, and there is no reason to assume performance transfers evenly to those groups. Testing across a wider range of ages, anatomy, and demographics — and validating the synthetic artifact generation against real clinical artifacts across those same groups — would be needed before drawing conclusions about how broadly this generalizes.

Responsibility is also an open question once AI is in the loop: if a radiologist reads a corrected image that hides an artifact or a missed pathology, it isn't obvious whether that's on the radiologist, the tool's developers, or the institution deploying it. We designed this framework as a decision-support tool rather than an autonomous one — the radiologist retains final interpretation — but that distinction gets harder to maintain in practice as these systems become more accurate and more trusted by the people using them. Requiring side-by-side review of the original and corrected images is one way of keeping a human in the loop, as opposed to allowing that boundary to erode gradually.

It's important to note the energy cost of training these networks, even a comparatively small one such as ours (roughly 10 million parameters) — the cumulative cost of development and fine-tuning iterations is not nothing. Mixed-precision training, early stopping, and shared or federated training across institutions are all ways to reduce that footprint, as is the more general case for running this kind of infrastructure on renewable energy where possible.

7. Conclusion

This paper introduced a hybrid deep learning and morphological framework for selective metal artifact reduction in CBCT. Dynamic thresholding and morphological dilation locate the artifact-affected region exactly; a residual U-Net conditioned on that region map restores the corrupted voxels; and selective blending confines the correction to exactly that region, leaving the surrounding anatomy intact. The full pipeline runs on GPU hardware at speeds compatible with clinical use.

On a diverse clinical dataset, the framework outperformed conventional interpolation, iterative reconstruction, and global deep-learning baselines, reaching a PSNR of 37.89 dB and an SSIM of 0.941 — 1.68 dB and 0.016 above the strongest baseline, respectively. Radiologists rated it highest on artifact reduction, anatomical preservation, and diagnostic confidence, and the ablation study showed that selective blending and morphological detection are the two components doing most of the work.

The contributions here, in short, are a hybrid architecture that localizes correction, a training regime that exploits both synthetic and real artifacts to ameliorate limited paired data, and an efficient enough design to be useful in a clinical setting. What remains are adaptive artifact describing, validation across scanner vendors, and uncertainty estimation to make it more trustworthy in practice — and ultimately, prospective clinical trials to demonstrate that it improves diagnostic accuracy and patient outcomes, not just image quality.

Funding

The authors declare that no financial support was received from any organization or institution for the research, authorship, and publication of this article.

References

[1] B Feragalli, O Rampado, C Abate, M Macrì, F Festa, et al. (2017) Cone beam computed tomography for dental and maxillofacial imaging: technique improvement and low-dose protocols. La radiologia medica.

[2] S Shujaat, MM Bornstein, JB Price, et al. (2021) Integration of imaging modalities in digital dental workflows-possibilities, limitations, and potential future developments. Dentomaxillofacial Radiology.

[3] HS Park, K Jeon & JK Seo (2025) Deep learning-based artefact reduction in low-dose dental cone beam computed tomography with high-attenuation materials. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences.

[4] F Edward Boas & Dominik Fleischmann (2012) CT artifacts: causes and reduction techniques. Imaging in Medicine, 4(2):229-240.

[5] G Wang, DL Snyder, JA O'Sullivan, et al. (1996) Iterative deblurring for CT metal artifact reduction. IEEE Transactions on Medical Imaging.

[6] HS Park & K Jeon (2026) MGMAR: Metal-Guided Metal Artifact Reduction for X-ray Computed Tomography. arXiv preprint arXiv:2603.13447.

[7] C Zhou, YE Zhao, S Luo, H Shi, L Zheng, LJ Zhang, et al. (2011) Monoenergetic imaging of dual-energy CT reduces artifacts from implanted metal orthopedic devices in patients with factures. Academic radiology.

[8] M Sharma, A Dogra, B Goyal, A Gupta & MJ Saikia (2025) Detail-preserving denoising of CT and MRI images via adaptive clustering and non-local means algorithm. Scientific Reports.

[9] XG Lv, YZ Song, SX Wang & J Le (2013) Image restoration with a high-order total variation minimization method. Applied Mathematical Modelling.

[10] X Tian, Z Yin, D Tan, Y He, C Li, T Chen, Y Peng, et al. (2026) FID-Net: an interpretable fourier and image domain convolutional dictionary network for industrial CT metal artifact reduction. Measurement.

[11] O Ronneberger, P Fischer & T Brox (2015) U-Net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 234-241.

[12] U Dmitry, A Vedaldi & L Victor (2020) Deep image prior. International Journal of Computer Vision.

[13] Z Wang, C Vandersteen, T Demarcy, D Gnansia, et al. (2019) Deep learning based metal artifacts reduction in post-operative cochlear implant CT imaging. Lecture Notes in Computer Science.

[14] W Cai & Z Wei (2020) PiiGAN: generative adversarial networks for pluralistic image inpainting. IEEE Access.

[15] R Farber (2011) CUDA application design and development. books.google.com.

[16] J Zhang, H Yang, J Ma, B Fan, T Shi, X Zheng, S Xie, et al. (2026) Radon-full-waveform inversion for suppressing scalp reverberation and skull-induced aberration in transcranial photoacoustic computed tomography. Photoacoustics.