IEEE Open Journal of
Signal Processing

Received 11 February 2025; revised 3 July 2025; accepted 14 July 2025. Date of publication 28 July 2025; date of current version 8 August 2025. The review of this article was arranged by Associate Editor L. Chaari.

Digital Object Identifier 10.1109/OJSP.2025.3593189

Short Paper

Advancing the Frontiers of Deep Learning for Low-Dose 3D Cone-Beam CT Reconstruction

ANDER BIGURI [1], SUBHADIP MUKHERJEE [2] (Member, IEEE), XUZHI ZHAO [3], XI LIU [3], XINYI WANG [3], RUI YANG [3], YI DU [4,5], YAHUI PENG [3], MIKAEL BRUDFORS [6], MARK GRAHAM [7], HYUNGON RYU [6] (Member, IEEE), OLIVER KUTTER [6], ANDREAS HAUPTMANN [8,9] (Senior Member, IEEE), MUSTAFA AL-RUBAYE [10], MIIKA T. NIEMINEN [10,11], MIKAEL A. K. BRIX [10,11], AUSTIN YUNKER [12], RAJKUMAR KETTIMUTHU [12], JOHN C. ROESKE [13], SASIDHAR ALAVALA [14], SUBRAHMANYAM GORTHI [14], AND CAROLA-BIBIANE SCHÖNLIEB [1]

[1] Department of Applied Mathematics and Theoretical Physics, University of Cambridge, CB2 1TN Cambridge, U.K.
[2] Department of Electronics and Electrical Communication Engineering, IIT Kharagpur, Kharagpur 721302, India
[3] School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China
[4] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China
[5] Institute of Medical Technology, Peking University Health Science Center, Beijing 100191, China
[6] NVIDIA, Santa, Clara CA95051 USA
[7] School of Biomedical Engineering & Imaging Sciences, KCL, WC2R 2LS London, U.K.
[8] Research Unit of Mathematical Sciences, University of Oulu, 90570 Oulu, Finland
[9] Department of Computer Science, University College, WC1E 6BT London, U.K.
[10] Research Unit of Health Sciences and Technology, University of Oulu, 90570 Oulu, Finland
[11] Department of Diagnostic Radiology, Oulu University Hospital, 90220 Oulu, Finland
[12] Argonne National Laboratory, Lemont IL 60439 USA
[13] Loyola University Chicago, Maywood IL 60660 USA
[14] Department of Electrical Engineering, IIT Tirupati, Tirupati 517619, India

CORRESPONDING AUTHOR: ANDER BIGURI (e-mail: ab2860@cam.ac.UK).

(Ander Biguri and Subhadip Mukherjee contributed equally to this work.)

The work of Ander Biguri was supported in part by the Accelerate Program for Scientific Discovery and in part by EPSRC under Grant EP/V029428/1 and Grant EP/W004445/1. The work of Subhadip Mukherjee was supported by Faculty Start-up Research Grant (FSRG) through IIT Kharagpur under Project Code: RAI.

ABSTRACT X-ray computed tomography (CT) is an important noninvasive medical imaging modality for studying the structural details of internal organs. Image reconstruction in CT is an *inverse problem* of recovering an object's internal structure from the absorption profile of X-ray beams (*sinogram*) measured using a detector. The classical variational approach for CT reconstruction minimizes an energy functional using an appropriate iterative algorithm. Motivated by the success of deep learning (DL), researchers have begun to leverage training data and enhanced computing capabilities in recent years to produce high-fidelity reconstructed images. Nonetheless, much of the academic research in DL algorithms for CT has focused primarily on the two-dimensional setting (with simplified forward operators and noise model) for proofs-of-concept, and a comprehensive benchmarking of various classical and data-driven CT reconstruction approaches has not beenundertaken. The key objective of our CT reconstruction grand challenge was to promote methodological advancements for both classical and DL-based approaches for clinical CT with a reasonably accurately simulated 3D CT forward operator and noise model. We have utilized the publicly available LIDC-IDRI dataset and simulated sinograms and FDK images corresponding to two dose levels (clinical- and low-dose, constituting two tracks of the challenge) starting from the normal-dose images as the ground truth. In this paper, we summarize the motivation, context, and results of our challenge, and highlight the future research directions in DL for clinical CT.

INDEX TERMS X-ray tomography, inverse problems, cone-beam CT, deep learning.

I. INTRODUCTION

Image reconstruction in X-ray CT is an *inverse problem* of recovering an image from its projections along lines in the 3D space. The forward model in CT (without noise) can be expressed using the Beer-Lamberts law $T(\boldsymbol{x})(\ell) = \exp(-\mu \int_\ell \boldsymbol{x}(z)\,\mathrm{d}z)$, $\ell \in \mathcal{L}$, where $\boldsymbol{x}$ represents the function (i.e., the image) to be reconstructed, and μ denotes the mass attenuation coefficient. The set of lines $\mathcal{L}$ along which the projections are taken is determined by the acquisition geometry, often chosen as the 3D helical cone-beam geometry in clinical applications. By considering the measurement in the log domain, the CT reconstruction problem can be reasonably approximated as a *linear inverse problem* of recovering a function $\boldsymbol{x} : \Omega \to \mathbb{R}$, $\Omega \subset \mathbb{R}^3$, from its line integrals $\boldsymbol{y}(\ell) = \int_\ell \boldsymbol{x}(z)\,\mathrm{d}z$, $\ell \in \mathcal{L}$ (also called the *sinogram*). After discretizing and taking into account noise in the data, the measurement equation can be expressed as

$$y = Ax + w, \tag{1}$$

where $x \in \mathbb{R}^n$ is the discretized 3D volume to be reconstructed (with n voxels), $A \in \mathbb{R}^{m \times n}$ represents the discrete counterpart of the CT forward operator, and w denotes noise in the data. Without any prior knowledge about x, the reconstruction problem is ill-posed, since A is either not invertible or does not have a stable inverse. In other words, there could be many possible solutions x that are consistent with the operator (1), and a small amount of noise in the data can drastically alter the solution (instability).

Traditionally, variational regularization (VR) [1] has been the most successful framework for addressing this inherent ill-posedness. In VR, one formulates an optimization

$$x_\lambda(y) \in \arg\min_x f(y, Ax) + \lambda\, g(x), \tag{2}$$

where f measures data fidelity (typically using the squared ℓ_2 distance), g is a regularizer promoting certain smoothness properties (such as a piecewise constant structure or sparsity in the wavelet basis), and the penalty parameter $\lambda > 0$ trades off data fidelity with regularization. For a large class of fidelity and regularizers, (2) is a convex program and can be solved using gradient-based algorithms, e.g., proximal gradient-descent (PGD) [2], *alternating directions method of multipliers* (ADMM) [2], and primal-dual techniques [3].

Despite theoretical guarantees, purely optimization-based approaches with handcrafted regularizers are limited in their ability to adapt to the images over a wide range of CT reconstruction tasks, which might lead to reconstructed images with suboptimal objective quality metrics, thereby affecting subsequent diagnostic accuracy. Moreover, iterative algorithms such as PGD or ADMM can take a few thousand iterations to converge, which could be unacceptably slow where the sinogram has a large number of projections and the image has a large number of voxels.

By training powerful deep neural networks (DNNs) on large corpora of data, researchers in recent years have developed faster and more efficient data-driven reconstruction methods that produce higher-quality reconstruction as compared to their classical counterparts. However, regularization approaches based on generative machine learning techniques [4], [5], [6] have the propensity of hallucinating undesirable artifacts, having serious consequences for clinical diagnosis. Further, the deep learning (DL)-based CT reconstruction techniques in the academic literature are often applied to highly pre-processed simulated data (primarily in the 2D setting) with highly simplistic simulations of forward operators and noise models. This can often lead to overly optimistic results, a phenomenon known as *implicit data crime* [7].

The key motivation for our 3D CT challenge was to facilitate a fair comparison of different data-driven and classical techniques for realistic 3D clinical CT reconstruction, providing new insights about their performance limits while indicating the avenues for new algorithm development. In the following, we briefly review recent DL approaches for CT reconstruction to put the challenge objectives in perspective.

II. DEEP LEARNING FOR CT: RECENT ADVANCES

The impressive success of deep learning in recent years has motivated researchers to go beyond the model-based variational framework for CT and develop data-driven strategies that leverage DNNs. We summarize some key deep learning-based CT reconstruction methods in the following. Some of these methods apply to a broader class of imaging inverse problems apart from CT image reconstruction.

A. LEARNED POST-PROCESSING

A two-step reconstruction approach (referred to as *FBPConvNet*) was proposed in [8], in which the filter back-projection (FBP) images were enhanced using a convolutional neural network (CNN). The noise and under-sampling artifacts in the FBP images can be removed using a CNN R_θ (θ representing learnable parameters) trained on pairs of FBP and the corresponding target images to minimize a supervised mean-squared error (MSE) loss. Another notable method in this category is the residual encoder–decoder convolutional neural network (RED-CNN) for low-dose CT proposed by Chen et al. [9]. These methods need more training data to generalize and suffer from data inconsistency [10].

B. DEEP UNROLLING OF OPTIMIZATION ALGORITHMS

The *algorithm unrolling* framework [11] achieves data efficiency by incorporating A and $A^\top$ into the reconstruction network. In unrolling, one considers a proximal splitting method (e.g., PGD or ADMM) for solving (2) and then replaces the proximal operators with a trainable shallow CNN. The learned primal-dual (LPD) framework [12] (constructed by unrolling the primal-dual algorithm [3]) is a notable example of an empirically successful unrolling approach for CT imaging, offering the flexibility of learning the proximal operators in both image and data spaces. A trained LPD model is significantly faster than its model-driven variant, but training an LPD network is computation-intensive. This is because A and $A^\top$ need to be computed as many times as the number of layers (typically ~ 10) during the forward and the backward passes while training, making it infeasible for 3D clinical CT. One major shortcoming of unrolling is that an unrolled network trained for L iterations tends to diverge when more than L iterations are executed during reconstruction. That is, the unrolled networks do not inherit the convergence properties of the corresponding optimization scheme. This problem is alleviated by the deep equilibrium (DEQ) model [13], which extends the concept of unrolling up to an infinite number of iterations.

To achieve memory and time efficiency, a promising alternative is to employ stochastic unrolling [14], wherein each layer needs to compute only a fraction of the forward operator (and the corresponding adjoint). The learned stochastic primal-dual (LSPD) approach [14] leads to a provably convergent unrolling scheme which is significantly more efficient than LPD while preserving the image quality.

C. PLUG-AND-PLAY (PNP) DENOISING

In the PnP framework (proposed in [15], see [16] for a recent review), off-the-shelf image denoisers are utilized as implicit regularizers for solving general ill-posed inverse problems. The key idea of PnP denoising is to replace the proximal operator with a Gaussian denoiser within a proximal splitting algorithm. This is motivated by the observation that the MAP denoiser corresponding to Gaussian noise contamination, given by $\hat{x} = \arg\min_x \frac{1}{2}\|x_\sigma - x\|_2^2 - \sigma^2 \log p(x)$, where $x_\sigma = x + \sigma\, w$ is the noisy image and $w \sim \mathcal{N}(0, I)$ denotes additive Gaussian noise, is essentially the proximal operator corresponding to $g(x) = -\sigma^2 \log p(x)$. Another motivation for using Gaussian denoisers as image priors comes from Tweedie's formula [17]: $\mathbb{E}[x|x_\sigma] - x_\sigma = \sigma^2\, \nabla_{x_\sigma} \log p_\sigma(x_\sigma)$, which establishes a direct relationship between the optimal minimum mean-squared error (MMSE) Gaussian denoisers and the score function of the noisy image (that serves as a good proxy for the true image prior). A popular instance of PnP denoising is PnP-PGD, with an iterative scheme $x_{k+1} = D(x^{(k)} - \eta_k \nabla f(y, Ax^{(k)}))$, where D is a Gaussian denoiser. Besides proximal PnP methods, there exists an alternative class of PnP techniques, referred to as the *regularization-by-denoising* (RED) approach [18], [19], in which a denoiser is used to construct an explicit regularizer.

D. LEARNED DNN-BASED REGULARIZERS

DNNs can be utilized to directly parameterize and learn the regularizer in a data-driven fashion. *Adversarial regularization* (AR) [20] and its convex variant (adversarial convex regularizer (ACR)) [21] are notable approaches in this class. Adversarial regularizers are trained as *critics* to discern a set of noise-free images $(x^{(i)})_{i=1}^N$ from FBP reconstructions $(A^\dagger y^{(j)})_{j=1}^N$ having under-sampling artifacts and noise. A regularizer $g_{\hat{\theta}}$ trained this way penalizes noise and artifacts generated by the FBP operator $A^\dagger$ and promotes solutions that are distributionally close to the clean ground-truth images. The optimal 1-Lipschitz regularizer approximates the Wasserstein-1 distance between π_x and $\pi_\dagger$, the distributions of clean and FBP images, respectively. For a new noisy sinogram y, the trained regularizer g_θ is incorporated into the variational framework and one minimizes $J(x) = \|y - Ax\|_2^2 + \lambda\, g_\theta(x)$ for reconstruction, with an appropriately chosen $\lambda > 0$.

Constructing the regularizer g_θ using a generic CNN leads to a nonconvex variational problem for reconstruction, leading to the issue of non-uniqueness of the solution and associated algorithmic complexities to find it. This can be circumvented by modeling g_θ such that $g_\theta(x)$ is a convex function of x. This parameterization is known as *input-convex neural networks* (ICNNs) [22] and is utilized in deriving the convex counterpart of AR [21], which is shown to provide improved robustness for limited-angle CT reconstruction. It was shown recently that similar convergence results can be obtained with an input-weakly-convex regularizer g_θ, thus bridging the gap between AR and its convex variant [23].

The Network Tikhonov (NETT) framework [24] models the regularizer as $g_\theta(x) = \varphi(\phi_\theta(x))$, where ϕ_θ is an L-layer DNN with learnable parameters θ, and φ is a scalar-valued functional. Specifically, the regularizer in [24] is a nonlinear extension of the ℓ_q regularizer: $g_\theta(x) = \sum_i \beta_i |[\phi_\theta(x)]_i|^q$, where $\beta_i > 0$. To ensure that the regularizer is coercive (which is needed for the theoretical guarantees), residual connections are added to the layers of ϕ_θ. The parametric model $\phi_\theta(x)$ used in constructing the regularizer is learned such that g_θ is small for artifact-free images and large for images with artifacts (similar to AR, but achieved with an encoder-decoder-based training approach).

E. SELF-SUPERVISED METHODS

Learned post-processing and unrolling require supervision, i.e., pairs of clean images and sinograms (or the FBP images), which might be difficult to obtain in practice. Self-supervised methods do not rely on the availability of any target ground-truth images and only make use of the noisy sinogram data (or FBP) to learn a reconstruction operator, thus making them more flexible than supervised methods. Self-supervised approaches for CT are a rapidly expanding area of research, with several new algorithms proposed in recent years (see, for instance, [25]). In the following, we review some notable self-supervised deep learning approaches in the context of CT.

Deep image prior: An empirically successful self-supervised approach for imaging is the deep image prior (DIP) method [26]. Surprisingly, DIP requires no training data, relying completely on the regularization effect of the architecture of the deep CNNs and the implicit regularization of the gradient-based optimizers [27]. Let G_θ be a deep convolutional generator, which can be either untrained or pretrained. For an arbitrary latent vector z, the DIP scheme seeks to *approximately* minimize the data consistency loss $\|y - AG_\theta(z)\|_2^2$ using some first-order methods such as Adam, with early-stopping to avoid overfitting. The final reconstruction is then computed as $x^\star = G_{\theta^\star}(z)$, where θ^* is the learned parameter. A major disadvantage of DIP is that one needs to train G_θ to reconstruct for each new y.

Equivariant imaging: Learning a reconstruction operator R_θ by minimizing the unsupervised training loss $J_{\text{unsup}}(\theta) = \frac{1}{N}\sum_{i=1}^N \|y^{(i)} - AR_\theta(y^{(i)})\|_2^2$ directly does not work due to the highly non-trivial null-space of A. To mitigate this, Chen et al. [28] proposed the *equivariant imaging* (EI) framework, utilizing the equivariance of the forward operator to improve the performance of unsupervised training. More precisely, for CT reconstruction, the plausible set of images $\mathcal{I}$ are invariant to a certain group of transformations $\mathcal{G} = \{g_1, g_2, \ldots, g_{|\mathcal{G}|}\}$ with actions T_g such that $T_g x \in \mathcal{I}$ for all $x \in \mathcal{I}$. For example, CT images are usually rotation invariant, and the desired reconstruction operator should approximately satisfy $R_\theta(AT_g x) = T_g R_\theta(Ax)$, i.e., $R_\theta \circ A$ should be equivariant under T_g, which is enforced as a penalty to regularize the unsupervised training loss. Although EI is more expensive to train and requires more memory, this framework demonstrates remarkable empirical potential and can closely match the accuracy of fully supervised approaches [28].

Stein's unbiased risk estimation (SURE): A self-supervised approach based on SURE [29] was proposed by Metzler et al. [30]. The reconstruction problem considered in [30] was that of recovering an image $x \in \mathbb{R}^n$ from its linearly degraded measurement $y = Ax + w$, where $w \sim \mathcal{N}(0, \sigma_w^2 I)$. It is shown in [30, equation (5)] that one can construct a surrogate quantity that depends only on y and the learnable parameters θ such that its expectation is equal to the MSE of reconstruction, which results in a self-supervised learning framework. Notably, the EI framework can be made robust to large noise by incorporating SURE into the EI loss [31].

III. PREVIOUS CT RECONSTRUCTION CHALLENGES

Tomographic reconstruction challenges have been routinely held in the community (e.g., visit www.aapm.org/GrandChallenge/ for challenges organized by the *American Association of Physicists in Medicine* (AAPM) in recent years). In the interest of brevity, we will mention below three such representative challenges (which are also closely related to ours) conducted over the past few years focusing on (i) low-dose, (ii) sparse-view, and (iii) limited-angle CT. We will also highlight the main differences between these challenges and ours in terms of the settings and objectives.

The AAPM low-dose CT grand challenge was conducted by the *American Association of Physicists in Medicine* (AAPM) in 2016 to perform a quantitative evaluation and comparison of denoising and iterative reconstruction algorithms for low-dose CT.[1] The overall dataset consisted of abdominal CT scans of 30 patient cases, of which 10 were utilized for training. The participants were given the patient data with full- and quarter-dose sinograms and the FBP images for algorithm development. The low-dose data were simulated by introducing Poisson noise corresponding to one-fourth of the clinical dose level. The two best-performing methods utilized a patch-based similarity constraint with a spatially variant penalty [32] and a deep CNN on the directional wavelet transform coefficients of the low-dose images [33]. While this challenge is by far the closest one in spirit to ours, it was conducted using fewer training scans than our challenge (thus restricting the generalizability of the methods). Moreover, DL research for CT has progressed considerably since this challenge, necessitating the establishment of new benchmarks.

The AAPM sparse-view CT challenge [34] was launched to gather evidence as to whether deep learning methods outperform their classical counterparts (in particular, TV minimization) for solving the sparse-view CT inverse problem in the idealized scenario (no measurement noise). For training, 4000 simulated 2D phantoms and their noiseless sinograms (and the FBP images) with projections along 128 angles were made available to the participants. The evaluation metric was the root-mean-squared error (RMSE) on 100 test images. The winning method [35] (which improved the RMSE of a previous CNN-based reconstruction by two orders of magnitude), adopted an approach that first estimated the acquisition geometry followed by an iterative unrolling strategy. This challenge convincingly established the promise of deep learning for the CT inverse problem. However, the reconstruction problem was formulated in 2D with sparse-view projections and no noise, which are unrealistic from a clinical point of view.

The Helsinki tomography challenge (HTC), organized by the *Finnish Inverse Problems Society* (FIPS), considered the task of limited-angle tomography, wherein the projection data are recorded over a limited angular region (instead of the usual full angular coverage). The HTC also considers the task of 2D object recovery from their limited-view sinograms. The challenge data consists of 21 phantoms, arranged into seven groups of progressively increasing difficulty (with a shorter field-of-view). The winning strategy [36] adopted a data augmentation approach (considering the small amount of training data) and generated synthetic phantoms and their sinograms using the known forward projection. Subsequently, a neural network was trained to predict the image directly from the sinogram. The HTC was also conducted for 2D reconstruction with a small number of training images.

IV. THE OBJECTIVES OF THE CHALLENGE

Besides the postprocessing and unrolling approaches, several novel data-adaptive methods for regularization have been proposed for imaging inverse problems (see [37] and references therein), and for CT reconstruction in particular. Notably, most of the academic papers demonstrate proof-of-concept for new data-driven reconstruction methods for the 2D CT imaging problem, and no systematic comparison of different deep-learning approaches for clinical cone-beam CT is available in the literature. This served as the primary motivation behind conducting this challenge, through which we sought answers to the following questions: (1) Are deep learning approaches competitive or superior to model-driven iterative methods for 3D clinical CT? (2) Do these methods scale well to 3D? (3) How robust are deep learning-based algorithms when accurate knowledge of the forward operator is not available? This is crucial as a learned approach might face operator mismatch when used in practice. We believe that addressing these issues will bridge the gap between the methodological advances in deep learning and their application to real-world 3D clinical CT.

A. DATASET PREPARATION

We used an instance of the LIDC-IDRI dataset, which consists of 1010 chest cone-beam CT (CBCT) scans for lung nodule analysis. These images were used as *ground truth* to simulate CBCT sinograms using the *Tomosipo* library [38]. Considering the size of the dataset, simplified (yet reasonably realistic) simulations are produced (i.e., not using Monte Carlo physics-based simulations), taking into account first-order approximations of expected noise. The complete detector model [39] is implemented as

$$
\begin{aligned}
y_i &= f_{\text{conv}} \sum_k e_k \cdot \mathcal{P}(\mathrm{DQE}_{ik} \cdot (a_{ik} + s_{ik})) + \mathcal{N}(0, \sigma^2_{\text{elec}}),\\
y + w &= \Gamma_{\sigma_{\text{cross-talk}}}[y_1, y_2, \ldots, y_I]^\top
\end{aligned}
\tag{3}
$$

where i is the pixel index of the full detector I, e_k is the energy level at energy k, a_{ik}s are the incident photons in the pixel, and s_{ik}s are the detected scattered photons in the same pixel. The detector quantum efficiency DQE_{ik} is assumed to have a value of one in this work, and f_{conv} represents the energy to electron conversion rate, which has been modeled using an 8-bit ADC. The noise process is described by $\mathcal{P}$, denoting a Poisson random generator, and $\mathcal{N}(\sigma^2_{0,\text{elec}})$ denotes Gaussian noise with zero mean and variance σ^2_{elec}. Finally, $\Gamma_{\sigma_{\text{cross-talk}}}$ is a $D \times D$ matrix that models detector cross-talk, defined as a fraction of the signal $\sigma_{\text{cross-talk}}$ that is shared between adjacent pixels, taken as 5% shared. In particular, A_{ik} is modeled by clean forward projection and reverse-log-transformation following the Beer-Lambert law, assuming 10^5 and 10^4 photons in the air for the clinical- and the low-dose, respectively, followed by a reverse flat-field correction using only light fields that reflect the cone beam geometry. To model s_{ik}, a simple scatter simulation is added by thresholding the image at the bone level Hounsfield Units, such that only the highly scattered tissues are kept. Then, using only this highly attenuating tissue, a sinogram is produced, and convolved with a large Gaussian kernel (size 100×100 and $\sigma = 45$), producing a first-order approximation of scatter. Finally, the detector information I is log-transformed and flat-field corrected.

We provided the data corresponding to two doses, one with the clinical dose and the other for approximately 10% of the clinical dose. The simulations are divided into 800 training samples, 100 validation samples, and 110 test samples. We provided the training and validation datasets only, with the ground truth data, two sinograms, and two noisy reconstructions using the FDK algorithm corresponding to two levels of noise. The acquisition geometry was specified, although we did not provide the exact code for simulating the sinograms and the FDK images. The rationale behind this was to ensure that the methods were not sensitive to inaccuracies in the modeling of the forward operator. All data can be accessed at https://zenodo.org/communities/icassp_2024_cbct_challenge.

[1] See www.aapm.org/GrandChallenge/LowDoseCT for more details on the objectives, data, and results.

TABLE 1. A Summary of the Top-Five Methods and Their Corresponding MSE Values for Both Clinical- and Low-Dose Settings

clinical dose		low dose	
Method	**MSE** $\times 10^{-6}$	**Method**	**MSE** $\times 10^{-6}$
Postprocessing FDK via 3D ResUNet [40]	0770	Singoram PDNet, FDK and IDNet denoiser [40]	1449
Postprocessing FDK via SegResNet [41]	0841	Postprocessing FDK via SegResNet [41]	1476
Multi-filter and multi-scale U-Net for FDK post-processing [42]	0966	3D U-Net for FDK postprocessing [42]	1679
3D U-Net for FDK postprocessing [43]	1044	multi-filter and multi-scale U-Net for FDK post-processing [42]	1706
SwinIR-based sinogram and image enhancement [44]	4480	3D U-Net for FDK postprocessing [43]	1939

TABLE 2. Training and Inference Times With Computational Resources for Each Submission. Memory Refers to the Final Model At Inference

Submission	**Training time**	**Inference time**	**Memory**	**GPU requirement**
Postprocessing FDK via 3D ResUNet [40]	53h/47h	5s	7/2GB	2× NVIDIA RTX 4090
Postprocessing FDK via SegResNet [41]	24h	5s	12GB	1× NVIDIA A100
Multi-filter and multi-scale U-Net for FDK post-processing [42]	120h/96h	13s/10s	8GB	1× NVIDIA Quadro RTX 6000
3D U-Net for FDK postprocessing [43]	18h/14h	4s	10.15GB	8× NVIDIA H100
SwinIR-based sinogram and image enhancement [44]	26h	1.7s	0.3GB	1x NVIDIA A100

V. SUBMISSIONS, EVALUATION, AND RESULTS

We received a total of seven submissions in the clinical dose category and nine submissions in the low-dose category. The MSEs (between the reconstructed volume and the corresponding target ground truth) for the top-five submissions in both categories, along with a short description of the methods, are provided in Table 1 and Table 2 highlights the computational resources and training times for each method. The methods mainly comprise deep neural network-based denoising in the image and sinogram domains (or some combination thereof). We give a concise summary of the methods in Section VI. More details on their implementation can be found in the accompanying papers for the respective submissions.

We believe that the challenge will expedite algorithmic developments in deep learning tailored for the 3D clinical CT, while more efforts will be invested to incorporate robustness to simplistic or imprecisely simulated imaging operators, noise models, and mismatch in the prior during training and testing, leading to more sophisticated and scalable models for clinical practice. It is notable that most (except two) submissions to the challenge are fully postprocessing based. This an important observation as in 2D, often unrolled methods and other variational regularization-based methods outperform post-processing, yet no submissions that fit these categories exist, and thus the hypothesys that they would behave similarly in 3D has not been tested.

Region-of-interest (ROI)-only evaluation: As an alternative evaluation, the CBCT images (both ground truth and the model outputs) are cropped to only account for the *region of Interest* (ROI) within the cylinder that defines the full-view CBCT. This evaluation favors methods that focus on improving the meaningful areas of the image, while penalizing methods that focus on cleaning the artifacts in the areas of the image with no signal. The results are similar to Table 1, except that the second winner becomes the fourth and the fifth for clinical and low-dose, respectively.

VI. DETAILS OF THE TOP FIVE METHODS

We outline in descending order the best five methods (in terms of MSE for both clinical and low-dose. Interested readers may refer to the respective summary papers for details (model architecture(s), training hyperparameters, etc.). Notably, the top methods in our challenge are based on the post-processing/denoising of the FDK images (and the sinogram), and not based on more advanced deep reconstruction methods such as algorithm unrolling or plug-and-play denoising. We believe that this is largely due to their high memory footprint and computational cost for training and inference, which restricts the number of iterations and network complexity. To mitigate these limitations, several promising strategies, such as stochastic unrolling and operator sketching [14], invertible unrolling [45], and deep equilibrium models [13], can be explored. Incorporating such approaches could enable more effective and scalable 3D reconstruction frameworks, and we view this as a compelling direction for future work.

A. DUAL-DOMAIN NETWORKS

For clinical dose CBCT, images were first reconstructed using the Feldkamp, Davis, and Kress (FDK) algorithm [38], followed by IDCNN-based denoising. IDCNN is based on a 3D-ResUNet architecture, consisting of four encoder-down and four decoder-up layers. In each encoder (decoder) layer, a residual convolution block is followed by a down-sampling (up-sampling) block. Additionally, the model adopts residual learning by adding the input directly to the output to produce the final reconstruction. For low-dose CBCT, the low-dose projections were first denoised using the PDNet model, the output of which is then used to reconstruct images with the FDK algorithm. The reconstructed images were then refined using the IDNet model. The PDNet model shares the same architecture as IDCNN and is tailored to reduce noise in 3D low-dose projections. The IDNet architecture, modified from the 2D U-Net model, incorporates residual learning similar to IDCNN and PDNet, adding the model's input directly to its output. The IDNet model focuses on preserving fine details to improve the quality of 2D reconstructed images.

B. MONAI'S SEGRESNET

MONAI [46] is used to train two separate denoising networks, one on the clinical dose and the other on low-dose reconstructions. The reconstructions are created by applying the FDK algorithm [38] to

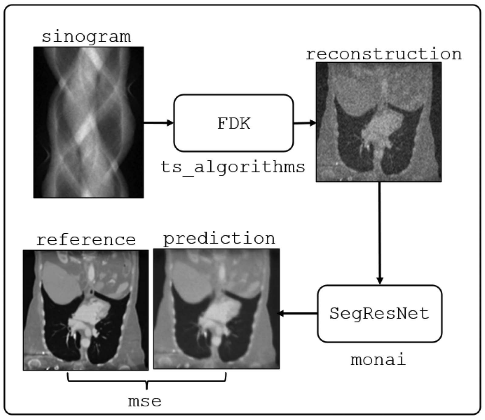

FIGURE 1. MONAI-based SegResNet for FDK post-processing.

the sinograms. For both networks, we use the SegResNet architecture [47], with 32 initial filters, (1, 2, 2,4) downsampling blocks, (1, 1, 1) upsampling blocks, and a dropout probability of 0.2. The training was done with the mean-squared error (MSE) loss, comparing the ground-truth FDK reconstruction with the network prediction. The Adam optimizer was used to train both networks, with a learning rate of 10^{-4} and a batch size of one. The model is trained for 400 epochs, and the model with the lowest average validation MSE is selected. The training was augmented with random affine transformations. A schematic of the proposed solution is shown in Fig. 1. The implementation will be made available in the MONAI tutorial repository to facilitate easy experimentation with different architectures and hyperparameters.

C. MULTI-FILTER AND MULTI-SCALE UNET

The initial reconstruction of CBCT projection data (y) is achieved using the FDK algorithm. If y is undersampled or noisy, the FDK reconstructions exhibit artifacts. Additionally, the cone-beam geometry introduces further artifacts in areas where only a partial set of rays passes through the target. One effective method to remove these artifacts is to use a post-processing network Λ_θ, such as a CNN with parameters θ, trained to enhance the initial reconstructions by mitigating artifacts and noise. An approach to enhance network performance by providing a more informative input to the post-processing network is proposed, inspired by [48]. This is achieved by performing a multi-filter FDK reconstruction using several frequency cut-offs for the frequency filter. Additionally, the discretization invariance of the ray transform is leveraged to create initial reconstructions on multiple scales. This approach provides richer information to the network, simplifying the learning task. The resulting network uses a set of reconstructions, as illustrated in Fig. 2. The best performance was achieved by using four filters with decreasing frequency cut-offs and two scales (full and half resolution). The proposed multi-filter and multi-scale U-Net [49] and a baseline U-Net achieved third place in the challenge. A visual assessment of results showcased that the multi U-Net could reduce artifacts and noise more effectively in the clinical dose scenario, while the baseline U-Net performed better in the low-dose setting.

D. 3D UNET

This approach also utilizes the U-Net architecture, consisting of a down-sampling block, a bottleneck block, and an up-sampling block. In the down-sampling block, the input is passed through repeated applications of two 3 × 3 convolutions, each followed by a rectified linear unit (ReLU) and a 2 × 2 max-pooling operation with a stride of two pixels for down-sampling. In each down-sampling step, the number of feature channels are doubled. The input is then passed through the bottleneck layer, which does not change the input shape. From there, the input is passed through the upsampling block that consists of an upsampling of the feature map followed by a 2 × 2 convolution that halves the number of feature channels, two 3 × 3 convolutions, each followed by ReLU activation. A key component of these models is the skip connections, which combine deep, semantic, and coarse-grained feature maps from the decoder sub-network with shallow, low-level, fine-grained features from the encoder sub-network. For 3D LD-CBCT denoising, the following changes are made: 3D convolutional layers are used instead of 2D, only three down- and upsampling layers are used to reduce the memory usage, use of Leaky ReLU instead of ReLU, and the final output channel size is one. Separate models for the clinical- and low-dose datasets are trained, with have 520697 trainable parameters each. The models are trained using the Adam optimizer with a learning rate of 0.001 with a mini-batch size of two using eight H100 GPUs. Final models are saved based on the lowest validation error.

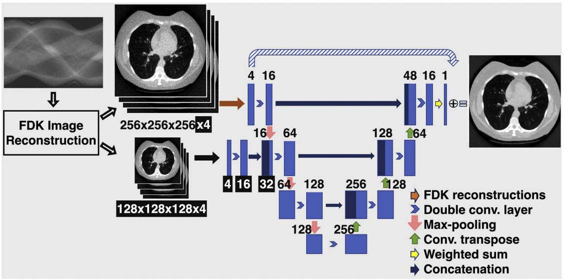

FIGURE 2. The projection data is reconstructed using the FDK algorithm with four different frequency cut-offs. The multi-filter and multi-scale U-Net, shown in white font, include processing the input images at both full-scale (256 × 256 × 256) and downscaled (128 × 128 × 128) resolutions. The baseline U-Net architecture takes a single reconstruction (256 × 256 × 256).

E. SWINIR-BASED SINOGRAM AND IMAGE ENHANCEMENT

The approach differs to an extent from its competitors in that it integrates the Swin image restoration (SwinIR)–based sinogram and image enhancement modules (SEM and IEM in short, see Fig. 5 for a block-diagram) for reconstruction. The overall method is philosophically similar to the approach for low-dose reconstruction proposed by Xuzhi et al., but differs in the specific ways in which the sinogram and the image are enhanced. The SEM first denoises the sinogram while capturing fine details, which is then utilized to solve a least-squares minimization problem using Nesterov's accelerated gradient algorithm to produce a CBCT reconstruction. The IEM acts on this reconstruction to produce the final output. The SwinIR [50] architecture comprises three modules: (1) Shallow feature extraction module: It consists of a convolution layer to extract the low-level features. (2) Deep feature extraction module: It consists of a series of residual swin transformer blocks (RSTB) followed by a convolution layer. Each RSTB comprises Swin-transformer layers with a residual connection. It extracts the high-level features that are not captured in the earlier module. (3) Image reconstruction module: It consists of a convolution layer for reconstructing the CT image by

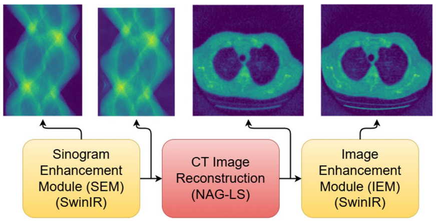
Sinogram Enhancement Module (SEM) (SwinIR)
CT Image Reconstruction (NAG-LS)
Image Enhancement Module (IEM) (SwinIR)

FIGURE 3. The (SwinIR)–based sinogram and image enhancement.

aggregating both the output of shallow and deep feature extraction modules. Integrating the sinogram and image enhancement modules aims to enhance image clarity and preserve fine details (see Fig. 3), offering a promising solution for low-dose and clinical-dose CBCT reconstruction. Despite the occurrence of considerable blurring due to the brute-force least-squares method, the IEM effectively mitigates this effect by enhancing the image while accentuating sharp features.

VII. CONCLUSIONS, OUTLOOK, AND FUTURE DIRECTIONS

The CT reconstruction problem in the 3D clinical setting is an extremely important inverse problem, both theoretically and practically, and it is imperative to make methodological advances in this area to improve the reconstruction image quality for facilitating accurate clinical diagnosis. While deep learning has shown enormous promise to bring about significant improvement in the image quality for CT, the algorithmic research in this area has remained somewhat disconnected from the requirements on the clinical front. Most of the deep learning-based approaches for CT are usually tested and benchmarked in the 2D setting, primarily due to the heavy demands on data and computation for training these models in 3D. While physics-aware deep learning algorithms such as unrolling have been demonstrated to have excellent image recovery performance in 2D, their applicability for 3D CT is severely restricted due to the presence of the 3D CT forward operator and its adjoint in the reconstruction network. Further research in strategies such as greedy layer-wise learning, stochastic unrolling, and operator sketching is needed to unleash the full power of algorithm unrolling for 3D CT. Post-processing-based approaches are relatively less computationally demanding in 3D, as the training process just requires learning the parameters of a 3D CNN in a supervised manner (without having to incorporate A and $A^\top$ iteratively). This perhaps explains why most of the top five methods in our challenge leveraged different variants of a 3D U-net for FDK post-processing. Nonetheless, it remains to be investigated how such post-processing schemes generalize to slight variations in the forward operator (which is inevitable while training on simulated FDK images) and the image prior when trained on a limited number of 3D volumes. Different well-known strategies for designing memory-efficient networks (e.g., using separable convolutions, introducing invertible layers, etc.) also need to be thoroughly investigated.

This study was motivated by three key questions: whether deep learning methods are competitive with traditional model-based reconstruction in 3D CT, whether such methods scale effectively to volumetric settings, and how robust they are to inaccuracies in the forward model. Our results show that deep learning-based post-processing approaches, particularly 3D U-Net and SegResNet, outperform or match classical algorithms like FDK, demonstrating strong reconstruction quality. These methods also scale well to 3D, both in terms of computational feasibility and performance. In contrast, hybrid approaches (such as PnP or iterative unrolling that combines CT physics with deep learning) were absent among the top submissions, likely due to their high memory and training costs, preventing any firm conclusion on their scalability. Furthermore, as the precise forward and noise models were intentionally withheld from participants, the observed performance reflects the methods' resilience to moderate operator mismatch. In particular, post-processing models appear robust under these conditions, although thorough validation in clinical environments remains an important direction for future work.

Interpretability and diagnostic reliability remain critical concerns in the application of deep learning to medical image reconstruction. In particular, post-processing methods may introduce hallucinated structures due to their limited ability to enforce data consistency, posing risks in clinical settings. While we have not performed a systematic interpretability analysis or clinical feedback, we acknowledge the importance of evaluating false positives, false negatives, and perceived diagnostic quality through radiologist assessment. Techniques such as deep null space learning [10], which explicitly constrain the learned solution to remain within the data-consistent manifold, offer promising directions to mitigate hallucination artifacts. A thorough clinical evaluation, including qualitative scoring and expert feedback, is an essential next step and lies beyond the current study's scope.

As a final note, one should be careful about the clinical relevance of the evaluation criteria. While MSE and SSIM are the most commonly used parameters for full reference image evaluation (PSNR in particular, just being a scaled MSE), these metrics do not always select the best clinical image [51] as they were designed for natural images, and not medical ones. Recently, a comprehensive study of image quality metrics performed using clinicians' input for 'best' image [52] and HaarPSI [53] showed the highest correlation with expert choice. Future challenges should consider this aspect, and readers should be careful while concluding that the best results for this (and other) challenges are clinically relevant or generalize to other experiments.

This challenge posed three big questions that are open for the learned CBCT reconstruction community. This challenge submission undoubtedly pushed the frontiers of the field, however we believe that the question are still open, and further research is required if these models are going to be used in clinical scenarios.

REFERENCES

[1] M. Benning and M. Burger, "Modern regularization methods for inverse problems," *Acta Numerica*, vol. 27, pp. 1–111, May 2018.

[2] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[3] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, Dec. 2010.

[4] S. Mukherjee, O. Öktem, and C.-B. Schönlieb, "Adversarially learned iterative reconstruction for imaging inverse problems," in *Proc. Int. Conf. Scale Space Variational Methods Comput. Vis.*, Apr. 2021, pp. 540–552.

[5] V. Shah and C. Hegde, "Solving linear inverse problems using gan priors: An algorithm with provable guarantees," in *2018 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 4609–4613.

[6] S. Mukherjee, M. Carioni, O. Öktem, and C.-B. Schönlieb, "End-to-end reconstruction meets data-driven regularization for inverse problems," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 21413–21425.

[7] E. Shimron, J. I. Tamir, K. Wang, and M. Lustig, "Implicit data crimes: Machine learning bias arising from misuse of public data," *Proc. Nat. Acad. Sci.*, vol. 119, no. 13, Feb. 2022, Art. no. e2117203119.

[8] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Proc.*, vol. 26, no. 9, pp. 4509–4522, Jun. 2017.

[9] H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[10] J. Schwab, S. Antholzer, and M. Haltmeier, "Deep null space learning for inverse problems: Convergence analysis and rates," *Inverse Problems*, vol. 35, no. 2, Jan. 2019, Art. no. 025008.

[11] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38 no. 2, pp. 18–44, Mar. 2021.

[12] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jan. 2018.

[13] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1123–1133, 2021.

[14] J. Tang, S. Mukherjee, and C.-B. Schönlieb, "Stochastic primal-dual deep unrolling," *arXiv:2110.10093v4*.

[15] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. 2013 IEEE Glob. Conf. Signal Inf. Process.*, Dec. 2013, pp. 945–948.

[16] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 85–97, Jan. 2023.

[17] B. Efron, "Tweedie's formula and selection bias," *J. Amer. Stat. Assoc.*, vol. 106, no. 496, pp. 1602–1614, Mar. 2011.

[18] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.

[19] E. T. Reehorst and P. Schniter, "Regularization by denoising: Clarifications and new interpretations," *IEEE Trans. Comput. Imag.*, vol. 5, no. 1, pp. 52–67, Mar. 2019.

[20] S. Lunz, O. Öktem, and C.-B. Schönlieb, "Adversarial regularizers in inverse problems," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 8507–8516.

[21] S. Mukherjee, S. Dittmer, Z. Shumaylov, S. Lunz, O. Öktem, and C.-B. Schönlieb, "Learned convex regularizers for inverse problems," Mar. 2021, *arXiv:2008.02839*.

[22] B. Amos, L. Xu, and J. Z. Kolter, "Input convex neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 146–155.

[23] Z. Shumaylov, J. Budd, S. Mukherjee, and C.-B. Schönlieb, "Weakly convex regularisers for inverse problems: Convergence of critical points and primal-dual optimisation," in *Proc. 2024 Int. Conf. Mach. Learn. (ICML)*, Jun. 2024, pp. 45286–45314.

[24] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier, "NETT: Solving inverse problems with deep neural networks," Dec. 2019, *arXiv:1803.00092*. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6420/ab6d57

[25] D. Evangelista, E. Morotti, and E. Loli Piccolomini, "RISING: A new framework for model-based few-view CT image reconstruction with deep learning," *Computerized Med. Imag. Graph.*, vol. 103, Jan. 2023, Art. no. 102156. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611122001264

[26] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2018, pp. 9446–9454.

[27] J. Tachella, J. Tang, and M. Davies, "The neural tangent link between CNN denoisers and non-local filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 8618–8627.

[28] D. Chen, J. Tachella, and M. E. Davies, "Equivariant imaging: Learning beyond the range space," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4379–4388.

[29] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981, doi:10.1214/aos/1176345632.

[30] C. A. Metzler, A. Mousavi, R. Heckel, and R. G. Baraniuk, "Unsupervised Learning With Stein's Unbiased Risk Estimator," Jul. 2020. [Online]. Available: https://arxiv.org/abs/1805.10531

[31] D. Chen, J. Tachella, and M. E. Davies, "Robust equivariant imaging: A fully unsupervised framework for learning to image from noisy and partial measurements," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 5647–5656.

[32] K. Kim, G. E. Fakhri, and Q. Li, "Low-dose CT reconstruction using spatially encoded nonlocal penalty," *Med. Phys.*, vol. 44, no. 10, pp. 376–390, Oct. 2017.

[33] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. 360–375, Oct. 2017.

[34] E. Y. Sidky, I. Lorente, J. G. Brankov, and X. Pan, "Do CNNs solve the CT inverse problem?," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 6, pp. 1799–1810, Jun. 2021.

[35] M. Genzel, J. MacDonald, and M. März, "AAPM DL-sparse-view CT challenge submission report: Designing an iterative network for fanbeam-CT with unknown geometry," *CoRR*, Jun. 2021, *arXiv:2106.00280*.

[36] T. Germer, J. Robine, S. Konietzny, S. Harmeling, and T. Uelwer, "Limited-angle tomography reconstruction via deep end-to-end learning on synthetic data," Sep. 2023, *arXiv:2309.06948*. [Online]. Available: https://www.aimsciences.org/article/doi/10.3934/ammc.2023006

[37] S. Mukherjee, A. Hauptmann, O. Öktem, M. Pereyra, and C.-B. Schönlieb, "Learned reconstruction methods with convergence guarantees: A survey of concepts and applications," *IEEE Signal Process. Mag.*, vol. 40, no. 1, pp. 164–182, Jan. 2023. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10004773

[38] A. A. Hendriksen et al., "Tomosipo: Fast, flexible, and convenient 3D tomography for complex scanning geometries in Python," *Opt. Exp.*, vol. 29, no. 24, pp. 40494–40513, 2021.

[39] M. Wu et al., "XCIST—An open access X-ray/CT simulation toolkit," *Phys. Med. Biol.*, vol. 67, no. 19, Sep. 2022, Art. no. 194002.

[40] Z. Xuzhi, X. Liu, X. Wang, R. Yang, Y. Du, and Y. Peng, "Dual-domain neural networks for clinical and low-dose CBCT reconstruction," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Aug. 2024, pp.,17–18.

[41] M. Brudfors, M. Graham, H. Ryu, and O. Kutter, "MONAI for deep-learning based CBCT reconstruction," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Aug. 2024, pp. 75–76.

[42] A. Hauptmann, M. Al-Rubaye, M. T. Nieminen, and M. Brix, "A multi-filter and multi-scale U-Net for cone-beam computed tomography with hardware constraints," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Aug. 2024, pp. 69–70.

[43] A. A. Yunker, B. R. Kettimuthu, and C. J. C. Roeske, "Low Dose CBCT Denoising Using a 3D U-Net," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Aug. 2024, pp. 85–86.

[44] S. Alavala and S. Gorthi, "3D-CBCT Challenge 2024: Improved cone beam CT reconstruction using SwinIR-based sinogram and image enhancement," in *Proc. 2024 IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2024, pp. 101–102.

[45] J. Rudzusika, B. Bajić, T. Koehler, and O. Öktem, "3D helical CT reconstruction with a memory efficient learned primal-dual architecture," 2023, *arXiv:2205.11952*. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10683984

[46] M. J. Cardoso et al., "Monai: An open-source framework for deep learning in healthcare," Nov. 2022, *arXiv:2211.02701*.

[47] A. Myronenko, "3D MRI brain tumor segmentation using autoencoder regularization," in *Proc. Brainlesion: Glioma, Mult. Sclerosis, Stroke Traumatic Brain Injuries,* Jan. 2019, pp. 311–320.

[48] A. Hauptmann, J. Adler, S. Arridge, and O. Öktem, "Multi-scale learned iterative reconstruction," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 843–856, Apr. 2020.

[49] A. Hauptmann, M. Al-Rubaye, M. T. Nieminen, and M. A. Brix, "A multi-filter and multi-scale u-net for cone-beam computed tomography with hardware constraints," in *Proc. 2024 IEEE Int. Conf. Acoustics, Speech, Signal Process. Workshops (ICASSPW)*, Aug. 2024, pp. 69–70.

[50] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. 2021 IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, 2021, pp. 1833–1844.

[51] A. Breger et al., "A study of why we need to reassess full reference image quality assessment with medical images", 2024, *arXiv:2405.19097*. [Online]. Available: https://link.springer.com/article/10.1007/s10278-025-01462-1

[52] A. Breger et al., "A study on the adequacy of common IQA Measures for Medical Images," 2024, *arXiv:2405.19097*. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-96-3863-5_41

[53] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A haar wavelet-based perceptual similarity index for image quality assessment," *Signal Process.: Image Commun.*, vol. 61, pp. 33–43, Feb. 2018.