Lose The Views: Limited Angle CT Reconstruction via Implicit Sinogram Completion

Rushil Anirudh, Hyojin Kim, Jayaraman J. Thiagarajan, K. Aditya Mohan,
Kyle Champley, Timo Bremer

Lawrence Livermore National Laboratory.

Abstract

Computed Tomography (CT) reconstruction is a fundamental component to a wide variety of applications ranging from security, to healthcare. The classical techniques require measuring projections, called sinograms, from a full 180° view of the object. However, obtaining a full-view is not always feasible, such as when scanning irregular objects that limit flexibility of scanner rotation. The resulting limited angle sinograms are known to produce highly artifact-laden reconstructions with existing techniques. In this paper, we propose to address this problem using CTNet – a system of 1D and 2D convolutional neural networks, that operates directly on a limited angle sinogram to predict the reconstruction. We use the x-ray transform on this prediction to obtain a "completed" sinogram, as if it came from a full 180°view. We feed this to standard analytical and iterative reconstruction techniques to obtain the final reconstruction. We show with extensive experimentation on a challenging real world dataset that this combined strategy outperforms many competitive baselines. We also propose a measure of confidence for the reconstruction that enables a practitioner to gauge the reliability of a prediction made by CTNet. We show that this measure is a strong indicator of quality as measured by the PSNR, while not requiring ground truth at test time. Finally, using a segmentation experiment, we show that our reconstruction also preserves the 3D structure of objects better than existing solutions.

1. Introduction

Computed Tomography (CT) is one of the most common imaging modalities used in industrial, healthcare, and security settings today. In a typical parallel-beam CT imaging system, x-ray measurements obtained from all viewing angles are effectively combined to produce a cross-sectional
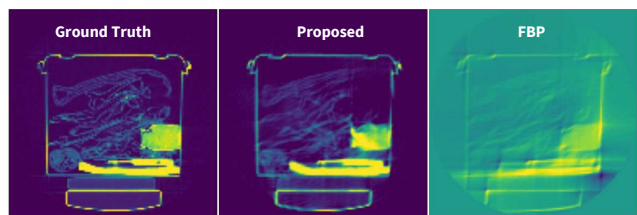

Corresponding author: anirudh1@llnl.gov.

Figure 1. We show the proposed limited angle reconstruction from just half the views (0-90°), in comparison to Filtered Back Projection (FBP) and ground truth. These are intensity normalized images.

image of 3D objects [16]. These x-ray measurements are collectively referred to as a *Sinogram*. The inverse problem of reconstructing cross-sectional images (or slices) from raw sinograms has been extensively studied by imaging researchers for several decades (Chapter 3 in [16]), the most popular technique being the Filtered Back Projection (FBP), which is derived from a discretization of the closed-form solution for the inverse x-ray transform. Alternatively, iterative techniques such as weighted least squares (WLS) have also been developed that improve upon FBP, in some cases, by making successive approximations of increasing accuracy to obtain the final image. In the traditional CT setting, one assumes access to measurements collected from the full range of views of an object, i.e. $\theta \in [0, 180°]$, but increasingly newer techniques are being developed that can recover images when a part of the views are missing, i.e. when $\theta \in [0, \theta_{max}], \theta_{max} < 180°$. These are referred to as limited angle projections, and reconstruction in such cases is highly ill-posed, as evidenced by the inferior performance of existing methods.

Need for Limited Angle Scans: The advantage of such a setup is that it can drastically reduce scan time by restricting the physical movement of the scanner. CT scans are being used to study organs such as the heart, and objects that are highly dynamic, implying that a slightly longer scan time introduces a lot of blurring into the image [26, 8]. Further, the limited angle setting can help limit the area of the

scan only to a region of interest like in healthcare applications such as breast [27], and dental [14] tomography. It can also support applications involving objects that have physical constraints restricting the angles from which they can be scanned, for example in electron microscopy [30, 29].

Recent works attempt to solve this problem through a variety of formulations – as explicit sinogram regression from limited view to full-view [13], reduction of artifacts obtained from FBP [9], and using convolutional neural networks to refine poorly initialized reconstructions obtained from FBP [36]. However, these techniques use simpler datasets with very less variability, and operate in regimes where a majority of the viewing angles are captured – for example $130°, 150°$ in [36], $140°$ in [30] and $170°$ in [13]. Instead, this paper performs CT reconstruction from just half the views during training, i.e. $90°$, from a challenging real world checked-in luggage dataset [1].

Challenges: Generally speaking, an edge of an object is recovered accurately in a CT image if an x-ray tangential to that edge is captured. When several such x-rays from a contiguous set of views are missing, as in the limited angle scenario, a significant amount of information regarding the scene is missing. A loose analogy in traditional computer vision, is like reconstructing a scene when it is partially occluded from the camera. In the last few years, data-driven approaches have made significant strides in solving similar challenging image recovery problems [5] such as image completion [28], image-inpainting [35], super-resolution [23], CS recovery [22]. These methods leverage the availability of large datasets, and the expressive power of deep learning to impose implicit constraints to the recovery problem.

However, CT reconstruction presents several additional challenges. Unlike standard images, CT images of transportation luggage, cargo, etc. can be very complex with no apparent low dimensional structure. As a result, even under the classical CT setting of full-view scans, training a neural network end-to-end to predict the final image is challenging. This is further exacerbated by the fact that collecting a large dataset of CT images and their corresponding projections is significantly harder. Consequently, many state-of-the-art methods rely on analytical techniques, e.g. FBP, to provide an initial coarse estimate [13, 6], which is then refined using deep neural networks. In the extreme setting as considered here, FBP can be highly misleading, rendering subsequent techniques ineffective. An example of FBP under a limited angle setting $(0, 90°)$ is shown in Figure 1. Finally, CT is typically used in critical applications, which necessitates the need for practitioners to understand the confidence or reliability of reconstructions at test time.

Proposed Work: In this paper we address these challenges, with **CTNet** – a system consisting of 1D and 2D convolutional neural networks (CNN) coupled with adver-

sarial training to recover CT slices from limited angle sinograms. Since sinograms have certain consistency conditions [33, 25] that are hard to enforce directly within a neural network, we propose to solve this problem by completing sinograms from limited angle to full view ($180°$), implicitly in the image space. In other words, we employ a three-stage approach – first CTNet produces a reconstruction based on a limited-angle sinogram. Next, we project this image into sinogram space as if it came from full-view measurements using the x-ray transform. Lastly, we use existing techniques such as FBP or WLS to obtain the final image.

We train our network with sinograms containing only half the viewing angles to directly predict ground truth consisting of reconstructions obtained from full-view measurements. Inspired by the success of 1D CNNs in language modeling [19], our network interprets the sinogram as a "sequence". This formulation allows us to model projections from individual views, while also enabling us to capture relationships across views through a simple attention model. Consequently, our approach supports the use of a different number of views at test time, to even lower viewing angles than $90°$. As seen in Figure 1, the proposed sinogram completion strategy is able to recover CT slices with high fidelity much better than FBP.

Finally, in order to generate a confidence measure for the recovery process, we propose to estimate per-pixel variabilities to perturbations in the latent space of sinograms, and compute an aggregated confidence score. Interestingly, the proposed score is highly correlated to the actual reconstruction quality measured with respect to the ground truth.

Our main contributions can be summarized as follows:

1. We propose the first deep learning solution to recover CT images from limited angle or incomplete-view sinograms.

2. We propose to utilize 1D CNNs to process sinograms, which enables generalization to different number of views during training and testing.

3. We develop a confidence metric for the recovered images, and show that it is a strong indicator of reconstruction quality, as measured by PSNR.

4. We demonstrate that our method significantly outperforms state-of-the-practice approaches on a challenging transportation security dataset.

5. Using 3D semantic segmentation experiments on the resulting reconstructions, we illustrate that the proposed approach preserves the 3D structure effectively.

2. Other Related Work

There have been other studies addressing limited angle reconstruction in different contexts; we refer the reader to
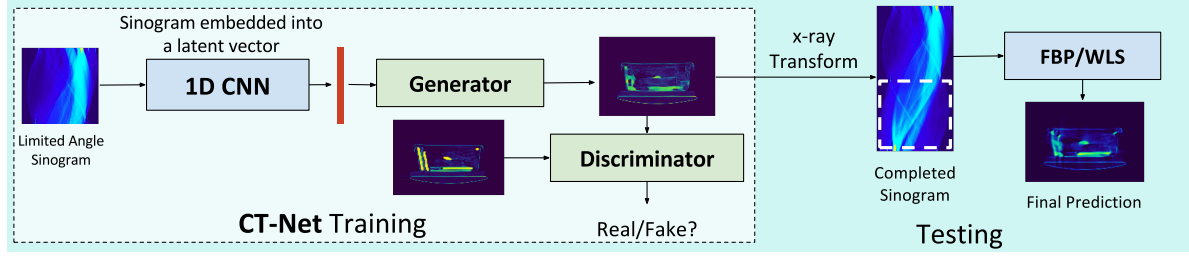
Figure 2. Overview of the proposed approach. During training, we train CTNet to predict the CT slice directly from limited angle sinograms. We use a mix of mean squared error and an adversarial loss. During test time, we forward project the ouput of CTNet using the x-ray transform to *complete* the sinogram. Next, we use WLS or FBP on the completed sinogram to obtain the final reconstruction. The architectures for the 1D CNN, Generator and Discriminator are described in the supplementary material.

[9] for a detailed list of them. A related but different problem is the few-view (also called sparse-view) CT reconstruction, which has been of significant interest. It differs from the limited-view problem in that, it reduces the number of viewing angles by *uniformly sampling* in the possible range of angles, $(0°, 180°)$ [11, 17, 15]. Zhao *et al.* proposed a convolutional neural network framework to recover poorly reconstructed images [37], and Chen *et al.* used a similar approach to denoise images from low-dose CT [6]. This recovery process closely resembles the techniques used for solving inverse problems in vision recently such as super-resolution [23], recovering images from compressive measurements [22], and other linear inverse problems [5]. When compared to sparse-view reconstruction, the limited angle problem is more challenging, as it is equivalent to extrapolation in the sinogram space.

In the tomographic reconstruction community, numerous studies have focused on algebraic approaches to inverse problems, in particular utilizing Algebraic Reconstruction Techniques (ART) and its variants such as Multiplicative ART (MART). Examples include the work on MART-AP [21] and simultaneous MART [31]. Further, Chen *et al.* proposed an adaptive Non-Local Means (NLM) based reconstruction method to compensate for over-smoothed image edges in few-view reconstructions [7]. In addition, there exist methods that utilize dictionary learning techniques, coupled with sparse representations or total variation optimization, for both few-view and low-dose CT reconstruction tasks [24, 34]. Despite the availability of such varied solutions, to the best of our knowledge, our work is the first that addresses the problem of limited-view CT reconstruction, by directly operating on the limited-view sinogram using viewing angles up to only $90°$.

3. Preliminaries

In this section we outline the basics of the CT scan, and briefly describe the state-of-practice algorithms for recovering the tomographic images from the sinograms.
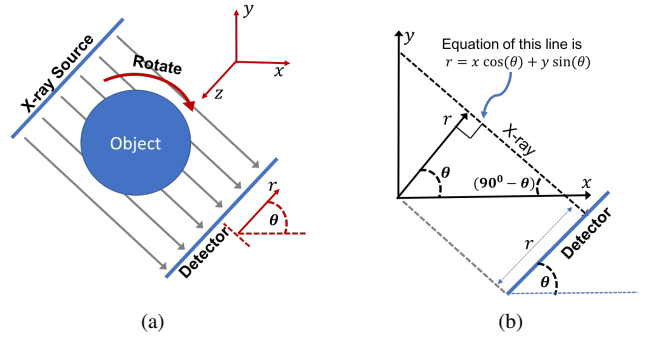

(a)   (b)

Figure 3. Experimental setup of x-ray computed tomography. An object is rotated along an axis and exposed to a parallel beam of x-rays. The intensity of attenuated x-rays exiting the object is measured by the detector at regular angular intervals. The projection at an angle of θ measured at a distance of r on the detector is the line integral of LAC values along the line perpendicular to the detector at r.

3.1. CT Reconstruction problem formulation

X-ray CT is a non-destructive imaging modality that is used to reconstruct the interior morphology of an object scanned using x-ray radiation. In our experiments, the object to be imaged is placed in between a source of parallel beam x-rays and a planar detector array. The x-rays get attenuated as they propagate through the object and the intensity of attenuated x-rays exiting the object is measured by the detector. To perform tomographic imaging, the object is rotated along an axis and repeatedly imaged at regular angular intervals of rotation. At each rotation angle of the object, the measurements at the detector can be expressed as the line integration of the linear attenuation coefficient (LAC) values along the propagation path. Assume that the object is stationary in the cartesian coordinate system described by the axes (x, y, z). Then, the projection at a distance of r on the detector is given by,

$$p_\theta(r, z) = \int \int \rho(x, y, z)\delta(x\cos(\theta) + y\sin(\theta) - r)dxdy.$$
(1)

where $\rho(x, y, z)$ is the LAC of the sample at the coordinates (x, y, z), $\delta(\ )$ is the standard Dirac delta function, and θ is the rotation angle. We refer to eq (1) as the x-ray transform in the remainder of the paper. Note that rotating the

object clockwise by an angle of θ is equivalent to rotating the source and detector pair counterclockwise by an angle of θ and vice versa. Notice that the equation (1) is separable in the z coordinate. Hence, the projection relation is essentially a 2D function in the $x - y$ plane that is repeatedly applied along the $z-$axis. Next, we describe two popular reconstruction algorithms.

3.2. Filtered Back Projection

Filtered back-projection (FBP) is an analytic algorithm for reconstructing the sample $\rho(x, y, z)$ from the projections $p_\theta(r, z)$ at all the rotation angles θ. FBP directly inverts the relation (1) to solve for the LAC values $\rho(x, y, z)$. In FBP, we first compute the Fourier transform $P_\theta(\omega)$ of the projection $p_\theta(r)$ as a function of r. The filtered back projection reconstruction Y is then given by [16],

$$Y(x, y) = \int_0^\pi Q_\theta \left(x \cos(\theta) + y \sin(\theta) \right) d\theta, \qquad (2)$$

where, $Q_\theta(r) = \int_{-\infty}^\infty P_\theta(\omega)|\omega| \exp(j2\pi\omega r)d\omega$. From equation (2), we can see that a filtered version of $p_\theta(r)$ is smeared back on the $x-y$ plane along the direction $(90° - \theta)$ (see figure 3). The FBP reconstruction thus consists of the cumulative sum of the smeared contributions from all the projections ranging from $0°$ to $180°$.

If the projections are acquired over a limited angular range, then the integration in (2) will be incomplete in the angular space. Since each projection $p_\theta(r)$ contains the cumulative sum of the LAC values at a rotation angle of θ, it also contains information about the edges that are oriented along the angular direction $(90° - \theta)$ in Figure 3. Now, suppose data acquisition starts at $\theta = 0°$ and stops at an angle of $\theta = \theta^{max} < \pi$. Then, the edge information contained in the projections at the angles $\theta \in [\theta_{max}, \pi]$ will be missing in the final reconstruction. This is the reason behind the edge blur in the reconstructions shown in figure 1.

3.3. Weighted Least Squares

Weighted least squares (WLS) is an iterative method for reconstructing the sample by formulating the reconstruction as the solution to a cost minimization problem. It belongs to the class of model-based reconstruction algorithms [3, 26], albeit without any form of regularization.

Let Y represent the reconstruction, that contains all the sampled values of the LAC $\rho(x, y, z)$ in 3D space, and S represent the sinogram. Then, the WLS reconstruction is given by solving the following optimization problem,

$$\hat{Y} = \arg\min_Y \left\{ (S - AY)^T W (S - AY) \right\} \qquad (3)$$

where A is the forward projection matrix that implements the line integral of (1) in discrete space and W is an estimate of the inverse covariance matrix of S [3], computed as a diagonal matrix with $W_{ii} = \exp(-S_i)$.

4. Proposed Approach

An overview of the proposed approach is described in Figure 2. In this section, we describe the details of our approach, and outline the training and testing strategies. A limited angle sinogram is a collection of measurements of a given object, stored in matrix form, from a range of views spanning from $0°$ to $\theta < 180°$ (fixed at $90°$ in this paper), where each row corresponds to a single view. Completing the sinogram directly is challenging, since there are consistency conditions in the sinogram space [33, 25] that cannot be easily enforced during the training process. Therefore, we resort to an implicit sinogram completion process, that converts a limited-angle sinogram to a full view one as described next.

4.1. From Half-View to Full-View: Implicit Sinogram Completion with CTNet

CTNet first embeds the limited angle sinogram into a latent space using a fully convolutional 1D CNN. The 1D convolutions are meaningful in this context, since they allow the use of a simple attention model to study correlations across neighboring views. We interpret the sinograms as a sequence of projections, corresponding to different viewing angles, similar to the sentence modeling in the NLP literature [19]. In the 1D CNN architecture, we use multiple filters with varying window sizes, in order to capture information across different sized neighborhoods. Each filter produces an embedding corresponding to a window size, resulting in the final embedding with dimensions $N_{filters} \times N_h$, where N_h denotes the different number of window sizes considered. See [19] for more details on the implementation. In our case we have $N_{filters} = 256$ filters, with $N_h = 5$ window sizes, resulting in a 1280 dimensional embedding. By design, this formulation supports varying number of rows (views) in the input sinogram. This latent representation is decoded into its corresponding CT image using a 2D CNN to predict the desired CT image. Our decoder is fully convolutional but for a projection layer in the beginning, and consists of residual units [12].

Training Losses: We trained CTNet with the standard mean squared error (MSE) as the loss: $\mathcal{L}_{mse} = ||\hat{Y} - Y||_2^2$, where $\hat{Y}$ and Y denote the predicted and ground truth images respectively. Training with MSE naturally results in the highest PSNR and SSIM metrics, as MSE optimizes specifically for them, however they result in highly smoothed images as the resulting solution is obtained as the average of many possible solutions.

We also use an adversarial loss [10], that uses a discriminator to guide CTNet to generate more realistic looking reconstructions. In practice, this results in sharper edges and visibly more high frequency content. Similar observations have been reported in the case of super resolution

[23], where it was found that PSNR is a weak surrogate for visual quality and using adversarial loss produces a more sharper rendering. The adversarial loss is measured as $\mathcal{L}_{adv} = -\log(D(\hat{Y}))$ where $D(.)$ represents the discriminator whose role is to distinguish between a generated image (fake) and an actual slice from the training dataset. The loss at the discriminator is $\mathcal{L}_D = -\log(D(Y)) - \log(1 - D(\hat{Y}))$. The final loss for the generator is hence obtained as $\mathcal{L} = \mathcal{L}_{mse} + \lambda \mathcal{L}_{adv}$. We found $\lambda = 0.05$ to be a suitable choice, resulting in the best reconstructions. Further details of all the networks inside CTNet can be found in the supplementary material.

Sinogram Completion: Once we have the prediction from CT-Net, $\hat{Y}$, from a $90°$ sinogram, $S_{(0,90)}$, we obtain a full-view sinogram from the $\hat{Y}$ using $\hat{S}_{(0,180)} = \mathcal{F}(\hat{Y}, 180°)$, where $\mathcal{F}$ corresponds to the x-ray transform (specified in eq (1)). This computes the sinogram, as if views from $180°$ were available for the current image slice, from which we use rows corresponding to $(90°, 180°)$ to obtain the completed sinogram – $S_{complete} = [S_{(0,90)}, \hat{S}_{(90,180)}]$. We obtain the final reconstruction with $S_{complete}$ using FBP, or WLS using the equations (2) or (3) respectively.

4.2. A New Confidence Score for Reconstructions from Limited Angle Sinograms

CT reconstruction is often used in critical applications such as healthcare or security, where an incorrect or misleading reconstruction can have negative consequences. This fact is even more important when we perform reconstruction from incomplete data, as we are operating in a highly under-constrained setting. In order to address this, we propose a confidence score, which measures the reliability of the reconstruction, for a given limited angle sinogram. This score is evaluated only at the test time and does not require the ground truth for its estimation.

At test time, the 1D-CNN embeds the limited-angle sinogram into a latent space, where we randomly perturb the vector using a dropout strategy and reconstruct the CT image from the perturbed latent representations. In all our experiments, the dropout probability was set at 0.05. For a given sinogram, we repeat this multiple times, and measure the per-pixel variance in the resulting reconstructions. The intuition here is that if the network has sufficient information to recover the slice, small perturbations in the latent space should not affect the final output significantly. However, if the sinogram does not reliably capture the information in the scene, its corresponding latent representation tends to be unstable, thus leading to significant changes in the reconstruction for small perturbations to its latent vector. Though a simple heuristic such as the ℓ_1 norm of the per-pixel variances can be directly used as the confidence metric, it can be highly sensitive to the number of objects
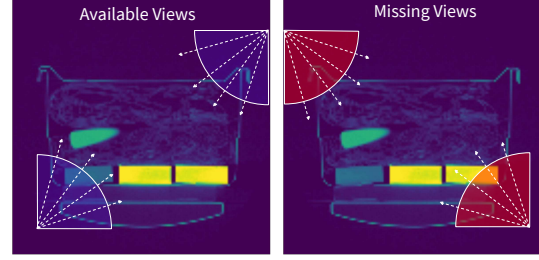

Figure 4. **Setup:** The views used in training, shown with reference to the physical arrangement of the training data.

present in the scan. Hence, we propose the following metric as a confidence score for the reconstruction: For the given per-pixel variance matrix V, and the actual predicted image $\hat{Y}_k$ obtained with no latent space perturbation, we define

$$ r_k = \exp\left(-\frac{\sum_i \sum_j V_{ij}}{||\hat{Y}_k||_2} \right), \qquad (4) $$

where the variances are normalized by the total ℓ_2 norm of the reconstructed image. We find that this metric acts as a strong indicator of the actual reconstruction quality, measured as the PSNR with respect to the ground truth image. Since evaluation of this metric does not require ground truth, this can be used by the practitioner to evaluate the reconstruction at test time, without actually generating the ground truth.

5. Experiments

In this section, we compare the sinogram completion approach with several baselines, and demonstrate its effectiveness in industrial CT reconstruction and segmentation.

Dataset We evaluate the our methods on a dataset of CT scans of common checked-in luggage collected using an Imatron electron-beam medical scanner – a device similar to those found in transportation security systems. The dataset is provided by the DHS ALERT Center of Excellence at Northeastern University [1] for the development and testing of Automatic Threat Recognition (ATR) systems. We repurpose this dataset for generating CT reconstructions from sinograms. The dataset is comprised of 188 bags, with roughly 250 slices per bag on an average. In total, the dataset consists of 50K full view sinograms along with their corresponding FBP reconstructions. The original slices are 1024×1024, but we perform experiments on their downsampled versions of size 128×128, and correspondingly the sinograms are subsampled to be of size 720×128. This corresponds to views obtained at every $0.25°$ sampled uniformly from $180°$. We split the bags into a training set of 120 bags and a test set with the rest. This split resulted in about $35K$ image slices for training and around $15K$ image slices for testing. The bags contain a variety of everyday
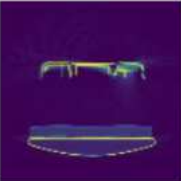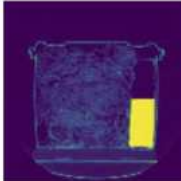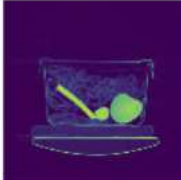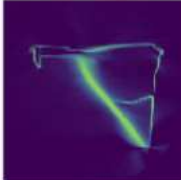
Figure 5. **Reconstruction Results:** PSNR (dB) is shown for each reconstruction, measured against the ground truth. It should be noted that sinogram completion is much better at preserving geometric shape (such as round objects) than baselines, apart from being superior in terms of PSNR. For sinogram completion, CTNet -mse and CTNet -adv performed nearly the same when followed up by WLS. We also show the nearest reconstruction from the training set, based on distance in the sinogram latent space.

objects such as clothes, food, electronics etc. that are arranged in random configurations. In all our experiments, we assume access to only the top half of the sinogram, which results in observing half the views $(0°, 90°)$ (Figure 4). We use this partial sinogram to train CTNet and the different baselines. As an initial processing step, we perform basic filtering to remove low intensity noise in the ground truth

reconstructions, obtained using FBP with all the views. We utilize LLNL's Livermore Tomographic Tool (LTT) [4] to compute the WLS, FBP, and xray transforms.

**Setup:** From the training data, we drop rows $(360, 720]$ in the sinogram matrix, corresponding to views obtained from $(90°, 180°]$. Physically, they represent the viewing angles going 0-90°clockwise as shown in figure 4. We consistently

6348
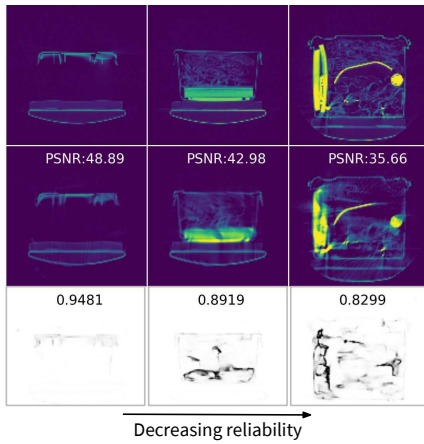
(a) top row: ground truth, middle: proposed reconstruction, and bottom row: pixel-wise confidence map. Light indicates more confident.
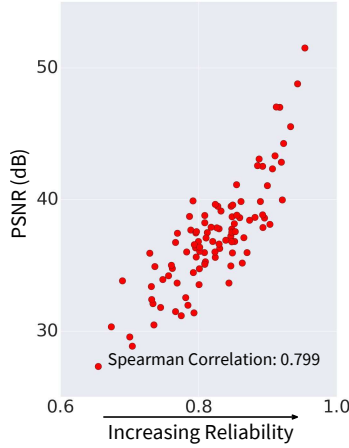
(b) PSNR v Reliability
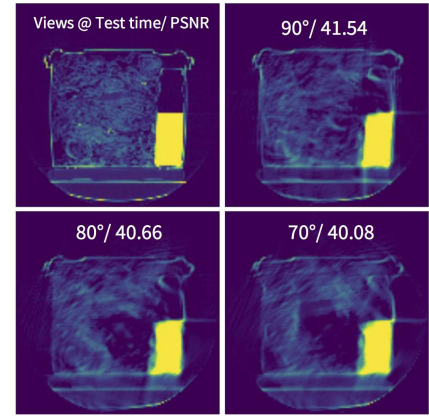
(c) Robustness to different views at test time: top left is ground truth, and others are reconstructions using CTNet-adv+WLS.

Figure 6. **Properties of CTNet**: (a) Pixel-wise confidence measures, values are shown on the same scale with values between $(0, 3 \times 10^{-4})$. (b) Proposed measure of confidence acts as a strong indicator of the quality of reconstruction, (c) Demonstration of the behavior of the proposed approach for varying number of views during testing.

drop the same views across all training images and testing images.

Training details and parameters: We trained our networks using Tensorflow [2] on a NVIDIA Tesla K40m GPU. We do not perform any scaling on the sinograms or reconstructions, since the value ranges are globally consistent across the entire dataset. Further, we use the same decoder in all our networks, regardless of the loss function. For the 1D CNN, we used filter sizes $[1, 2, 3, 4, 5]$ and employed the Adam optimizer[20] with learning rate 1×10^{-3} when using the MSE loss (CTNet-mse). For the mixture of adversarial and MSE losses (denoted CTNet-adv), we set the learning rate at 2×10^{-4} and the exponential decay rate for the first moment estimates at $\beta_1 = 0.5$.

5.1. Evaluating Quality of Reconstruction

The reconstruction results obtained using different baseline solutions and the proposed approach are shown in Figure 5. The baseline techniques include the state-of-practice methods, namely FBP and WLS, and variants of CTNet with both MSE and adversarial losses. We observe that, in general, implicit sinogram completion followed by an analytical reconstruction such as FBP or WLS produces more accurate reconstructions compared to methods that directly predict the CT image (CTNet-mse and CTNet-adv). In particular, CTNet-mse and CTNet-adv performed nearly the same when followed up by WLS, therefore we only show CTNet-adv+WLS in figure 5 for brevity. For each reconstruction, we compute the PSNR value (dB) and Structural Similarity (SSIM) measures, with respect to the ground truth. From Figure 5, it is evident that the proposed solution, in particular with WLS, is significantly better than existing

approaches and the baseline architectures. Furthermore, in Table 1, we show the mean PSNR and SSIM for 100 randomly chosen slices from the test set. Even though our performance is better as measured by PSNR and SSIM in the image space, they are not reflective of the large improvements in the reconstruction quality. Hence, we also measure the PSNR in the sinogram space by forward projecting the images using the X-ray transform, and comparing them to the ground truth $180°$ sinogram. Denoted by S-PSNR in Table 1, the sinogram space PSNR shows that CTNet is significantly better than existing baseline approaches, and overall the proposed solution of sinogram completion with WLS performs the best in terms of all metrics.

Confidence Score for Reconstruction: As described in Section 4.2, our proposed confidence score can provide guidance to qualitatively evaluate the reconstructions, without actually generating the ground truth for a test sample. We illustrate the pixel-wise confidence measures in Figure 6 for 3 different images, with decreasing levels of reliability obtained using CTNet-adv. Notice that in cases with noisy reconstructions, i.e., cases where the partial views considered do not sufficiently capture the properties of the scene, directly correspond to a lower confidence (or higher variance) as shown by the measure displayed on top of the variance maps. In Figure 6(b), we test the hypothesis that the proposed confidence measure can act as a strong indicator of the actual reconstruction quality, as measured by the PSNR. The strong correlation between PSNR and the proposed metric on the set of test images validates our hypothesis - an overwhelming evidence against the null hypothesis that they are not related, p-value=1.808×10^{-23}.

Table 1. PSNR and SSIM measures comparing 100 randomly sampled test slices with the ground truth. Completing the sinogram with CTNet followed by WLS is superior to all baseline methods. While PSNR and SSIM are metrics in the image space, S-PSNR is in the sinogram space.

Method	PSNR (dB)	SSIM	S-PSNR (dB)
FBP	35.65	0.846	26.56
WLS	37.3	0.939	32.08
CTNet -mse	37.70	0.940	33.70
CTNet -adv	37.19	0.935	32.82
CTNet -mse + FBP	37.42	0.932	33.22
CTNet -adv + FBP	37.35	0.930	33.46
CTNet -mse + WLS	**38.13**	**0.952**	34.11
CTNet -adv + WLS	38.08	0.950	**34.14**

Testing with fewer views than training: By design, our formulation allows the use of variable number of views in the sinogram. In practice, this translates to having reasonably stable reconstructions for further reduction in the number of views at test time. Note that, the views that are dropped can be random or can be consecutive in the sequence. Retraining a network for every unique set of views can be an arduous task. Our network can handle these cases well, by producing reconstructions whose quality degrades gracefully with incremental loss in the number of views. An example is shown in figure 6(c), where we show the reconstructions obtained using views ranging from $70°$ to $90°$ during test time.
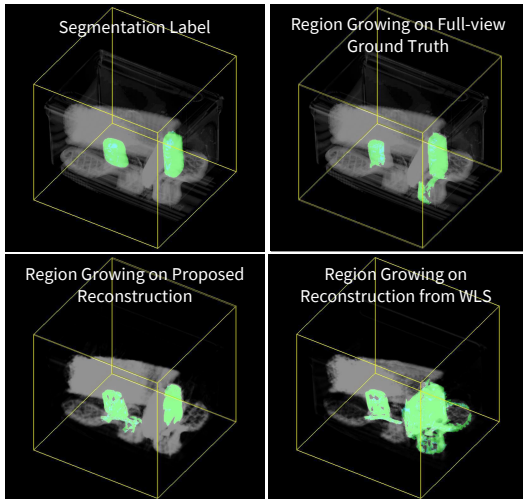

Figure 7. **3D Segmentation:** We employ a region growing 3D segmentation in all cases and the resulting segmentations are shown in color, against a 3D rendering of the reconstructed 2D images underneath. It is clearly evident that our method performs very similar to ground truth in determining the object boundaries compared to WLS.

3D Segmentation from CT Reconstructions: CT images are primarily used to study 3D objects, and hence evaluating the quality of the reconstructions in 3D segmentation can clearly demonstrate their usefulness in practice. We consider the 3D segmentation process, since it is often a critical step prior to performing complex inference tasks

such as threat detection [18]. To this end, we use the popular region-growing based segmentation proposed in [32] to identify high intensity objects in the bags from their reconstructions with partial views. We show an example for a bag with 260 image slices, that has been rendered in 3D using the 2D slices reconstructed with the proposed CTNet -adv+WLS and WLS alone respectively, in Figure 7. We compare the segmentations obtained using our method to the segmentation labels, and those obtained using just WLS. It can be seen in this example, WLS preserves 3D edges poorly resulting in spurious segments, whereas the proposed reconstruction is significantly better, resembling the ground truth. Additional segmentation results can be found in the supplementary material.

6. Discussion

In this paper we proposed to accurately recover CT images when the viewing angle is limited to only $90°$. We pose this problem as sinogram completion, but solve it in the image domain. Our empirical studies demonstrate the effectiveness of our three-stage approach – first computes a neural network based reconstruction, obtains a full-view sinogram using the x-ray transform based on the reconstruction, and then obtains the final reconstruction through WLS on the completed sinogram. We also proposed a confidence score to gauge the reliability of the recovery process, while being a reasonable surrogate for image quality.

Failure Cases: Our method works best on scenes with large objects that appear as low frequency content in images. We observe no significant gain over WLS in more complicated scenes that contain multiple small objects or intricate designs, since they manifest as high frequency image content, which are very hard to recover in this ill-posed problem setup. However, even with this limitation, we are able to recover 3D structure very well as shown in Figure 7, on realistic data that can be of practical use. Finally, since this dataset was not intended for limited angle reconstruction, there are some examples when the objects are just not in view (i.e. completely invisible within 0-90°), and our network has no information to recover them.

Future Work: There are several important directions forward for this work – (i) include the forward projection step (using x-ray transform) as a final layer inside CTNet , so that we can optimize the reconstruction setup end-to-end; (ii) Along with the reconstruction, jointly infer the segmentations from the sinograms in a multi-task learning setting.

Acknowledgement

References

[1] D. Center of Excellence at Northeastern University. ALERT TO4 datasets for automated threat recognition. available at http://www.northeastern.edu/alert. 2014.

[2] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[3] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Transactions on Image Processing*, 5(3):480–492, Mar 1996.

[4] K. Champley. Livermore Tomography Tools (LTT) User's Manual, Version 1.4. *LLNL-SM-744257*, pages 1–206, 2017.

[5] J. Chang, C.-L. Li, B. Póczos, B. Kumar, and A. C. Sankaranarayanan. One network to solve them all—solving linear inverse problems using deep projection models. *arXiv preprint arXiv:1703.09912*, 2017.

[6] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang. Low-dose ct via convolutional neural network. *Biomed. Opt. Express*, 8(2):679–694, 2017.

[7] Z. Chen, H. Qi, S. Wu, Y. Xu, and L. Zhou. Few-view ct reconstruction via a novel non-local means algorithm. *Phys Med*, 32(10), 2016.

[8] J. H. Cho and J. A. Fessler. Motion-compensated image reconstruction for cardiac ct with sinogram-based motion estimation. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013 IEEE*, pages 1–5. IEEE, 2013.

[9] J. Frikel and E. T. Quinto. Characterization and reduction of artifacts in limited angle tomography. *Inverse Problems*, 29(12):125007, 2013.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems (NIPS)*, pages 2672–2680, 2014.

[11] Y. Han and J. C. Ye. Framing u-net via deep convolutional framelets: Application to sparse-view ct. *arXiv preprint arXiv:1708.08333*, 2017.

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

[13] Y. Huang, X. Huang, O. Taubmann, Y. Xia, V. Haase, J. Hornegger, G. Lauritsch, and A. Maier. Restoration of missing data in limited angle tomography based on helgasonludwig consistency conditions. *Biomedical Physics & Engineering Express*, 2017.

[14] N. Hyvönen, M. Kalke, M. Lassas, H. Setälä, and S. Siltanen. Three-dimensional dental x-ray imaging by combination of panoramic and projection data. *Inverse Problems and Imaging*, 4(2):257–271, 2010.

[15] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[16] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[17] E. Kang, J. Min, and J. C. Ye. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *arXiv preprint arXiv:1610.09736*, 2016.

[18] H. Kim, J. J. Thiagarajan, and P.-T. Bremer. A randomized ensemble approach to industrial ct segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1707–1715, 2015.

[19] Y. Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.

[20] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations ICLR*, 2014.

[21] A. B. Konovalov and V. V. Vlasov. Spatial resolution analysis for few-views discrete tomography based on mart-ap algorithm. *ISRN Signal Processing*, 356291, 2013.

[22] K. Kulkarni, S. Lohit, P. Turaga, R. Kerviche, and A. Ashok. Reconnet: Non-iterative reconstruction of images from compressively sensed measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 449–458, 2016.

[23] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

[24] Y. Lu, J. Zhao, and G. Wang. Few-view image reconstruction with dual dictionaries. *Phys Med Biol*, 57(1):173–189, 2012.

[25] S. R. Mazin and N. J. Pelc. Fourier properties of the fan-beam sinogram. *Medical physics*, 37(4):1674–1680, 2010.

[26] K. A. Mohan, S. Venkatakrishnan, J. W. Gibbs, E. B. Gulsoy, X. Xiao, M. De Graef, P. W. Voorhees, and C. A. Bouman. Timbir: A method for time-space reconstruction from interlaced views. *IEEE Transactions on Computational Imaging*, 1(2):96–111, 2015.

[27] L. T. Niklason, B. T. Christian, L. E. Niklason, D. B. Kopans, D. E. Castleberry, B. Opsahl-Ong, C. E. Landberg, P. J. Slanetz, A. A. Giardino, R. Moore, et al. Digital tomosynthesis in breast imaging. *Radiology*, 205(2):399–406, 1997.

[28] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2536–2544, 2016.

[29] E. T. Quinto and O. Öktem. Local tomography in electron microscopy. *SIAM Journal on Applied Mathematics*, 68(5):1282–1303, 2008.

[30] S. V. Venkatakrishnan, L. F. Drummy, M. A. Jackson, M. De Graef, J. Simmons, and C. A. Bouman. A model based iterative reconstruction algorithm for high angle annular dark field-scanning transmission electron microscope (haadf-stem) tomography. *IEEE Transactions on Image Processing*, 22(11):4532–4544, 2013.

[31] V. V. Vlasov, A. B. Konovalov, and A. S. Uglov. *Few-Views Image Reconstruction with SMART and an Allowance for Contrast Structure Shadows*, pages 667–677. Springer International Publishing, Cham, 2015.

[32] D. F. Wiley, D. Ghosh, and C. Woodhouse. Automatic segmentation of ct scans of checked baggage. In *Proceedings of*

the 2nd International Meeting on Image Formation in X-ray CT, pages 310–313, 2012.

[33] P. Willis and Y. Bresler. Optimal scan for time-varying tomography. i. theoretical analysis and fundamental limitations. *IEEE transactions on image processing*, 4(5):642–653, 1995.

[34] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang. Low-dose x-ray ct reconstruction via dictionary learning. *IEEE Transaction on Medical Imaging*, 31(9), 2012.

[35] R. Yeh, C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with perceptual and contextual losses. *arXiv preprint arXiv:1607.07539*, 2016.

[36] H. Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, and G. Hu. Image prediction for limited-angle tomography via deep learning with convolutional neural network. *arXiv preprint arXiv:1607.08707*, 2016.

[37] J. Zhao, Z. Chen, L. Zhang, and X. Jin. Few-view ct reconstruction method based on deep learning. In *Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop*, 2016.