

Continuous Domain Reconstruction in CT Imaging with Coordinate-based Neural Networks

Gregory Ongie

Abstract—The majority of iterative algorithms for CT reconstruction rely on discrete-to-discrete modeling, where both the sinogram measurements and image to be estimated are discrete arrays. However, tomographic projections are ideally modeled as line integrals of a continuous attenuation function, i.e., the true inverse problem is discrete-to-continuous in nature. Recently, coordinate-based neural networks (CBNNs), also known as implicit neural representations, have gained traction as a flexible type of continuous domain image representation in a variety of inverse problems arising in computer vision and computational imaging. Using standard neural network training techniques, a CBNN can be fit to measurements to give a continuous domain estimate of the image. In this study, we empirically investigate the potential of CBNNs to solve the continuous domain inverse problems in CT imaging. In particular, we experiment with reconstructing an analytical phantom from its ideal sparse-view sinogram measurements. Our results illustrate that reconstruction with a CBNN are more accurate than filtered back projection and algebraic reconstruction techniques at a variety of resolutions, and competitive with total variation regularized iterative reconstruction.

Index Terms—Coordinate based neural networks, implicit neural representations, CT reconstruction, continuous domain inverse problems.

I. INTRODUCTION

Discrete image representations, which encode images as fixed-resolution arrays of pixels or voxels, form the backbone of modern computational imaging. However, estimating discrete images with high spatial resolution from indirect measurements is computationally demanding and often ill-posed because the number of unknown parameters (i.e., pixels or voxels) far exceeds the available measurements. This often necessitates introducing complicated regularization functionals to stabilize the recovery, which then require custom iterative solvers to implement.

Recently, a non-linear method of representing images as neural networks has gained traction in solving a variety of inverse problems in computer vision and computational imaging [1], [2]. These so-called *coordinate-based neural networks* (CBNNs)—also known as *implicit neural representations*—parameterize an image as a small-scale multi-layer perceptron taking spatial coordinates as inputs. Compared to traditional pixel/voxel representations, CBNNs are typically far more parameter-efficient, and scale well to multi-dimensional imaging problems. Furthermore, they can be fit to measurements using standard neural network algorithms, bypassing the need for custom iterative solvers.

G. Ongie is with the Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA e-mail: gregory.ongie@marquette.edu

In the context of computed tomography (CT) reconstruction, one advantage of CBNNs is that they allow for more accurate modelling of the ideal continuous-to-discrete measurement process. In particular, since the CBNN is a continuous function, samples of its continuous domain Radon transform can be approximated to an arbitrary degree of accuracy.

Several studies have proposed specialized CBNN approaches for CT reconstruction, e.g., [3], [4]. Yet, to the best of our knowledge, no work has investigated whether CBNNs enable the solution of the ideal continuous domain CT reconstruction problem. In this study, we explore this question empirically. For simplicity, we focus on a 2D imaging scenario and analytical phantoms where computing exact samples of Radon transform is possible. We investigate the recovery of sparse-view sinogram data with CBNNs using a close-to-exact continuous domain forward model versus recovery in a pixel basis with model-based iterative methods using a discretized forward model.

II. PROBLEM FORMULATION

Recovery of a continuous-domain image $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ from discrete sinogram measurements $\mathbf{y} \in \mathbb{R}^M$ can be posed as an optimization over a function space:

$$\min_{f \in \mathcal{F}} \|\mathcal{A}f - \mathbf{y}\|^2 \quad (1)$$

where $\mathcal{F}$ is some class of functions, and the continuous-to-discrete linear operator $\mathcal{A} : \mathcal{F} \rightarrow \mathbb{R}^M$ returns a vector of samples of the continuous Radon transform of f , i.e., if $\mathbf{y} = \mathcal{A}f$ then its i th entry is $y_i = \mathcal{R}f(\theta_i, t_i)$ for some angle/offset pair $(\theta_i, t_i) \in [0, 2\pi) \times \mathbb{R}$ where

$$\mathcal{R}f(\theta, t) = \int_{-\infty}^{\infty} f(s \sin(\theta) + t \cos(\theta), -s \sin(\theta) + t \cos(\theta)) ds.$$

Here, for simplicity, we focus on the least-squares loss in (1), though other loss functions may be more appropriate in settings with a statistical noise model, e.g., the negative log likelihood.

A. Reconstruction in a pixel basis

When reconstructing in a pixel basis, the function class $\mathcal{F}$ in (1) is chosen to be the space of images $f(x, y) = \sum_{i,j} f_{i,j} \chi_{P_{i,j}}(x, y)$, where $\chi_{P_{i,j}}$ is the indicator function for a square region $P_{i,j}$ representing the i, j th pixel. In this case, (1) reduces to the least squares problem

$$\min_{\mathbf{f} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{f} - \mathbf{y}\|_2^2 \quad (2)$$

where $\mathbf{f} \in \mathbb{R}^N$ is the vector of all pixel intensities $f_{i,j}$, and the columns of $\mathbf{A} \in \mathbb{R}^{M \times N}$ are given by $\mathcal{A}\chi_{P_{i,j}}$ for all i, j .

One drawback to this approach is that when discretizing at a high resolution the number of pixels N can easily exceed the number of sinogram measurements M . In this case, since $\mathbf{A}$ is a $M \times N$ matrix, there are infinitely many solutions to (2), i.e., reconstruction is ill-posed. To overcome this issue, a common strategy is to solve the regularized least squares problem instead:

$$\min_{\mathbf{f} \in \mathbb{R}^N} \|\mathbf{A}\mathbf{f} - \mathbf{y}\|^2 + \lambda R(\mathbf{f}), \quad (3)$$

where $R(\mathbf{f})$ is a regularizer designed to promote desirable image properties and $\lambda > 0$ is a tunable regularization strength parameter. One common choice is total variation regularization

$$R_{TV}(\mathbf{f}) = \sum_{i,j} \sqrt{|f_{i+1,j} - f_{i,j}|^2 + |f_{i,j+1} - f_{i,j}|^2},$$

which promotes reconstructing images that are piecewise constant.

B. Reconstruction with CBNNs

The philosophy behind the CBNN approach is to choose $\mathcal{F}$ in (1) to be a parametric class of functions $f_{\theta} : \mathbb{R}^2 \rightarrow \mathbb{R}$ described by a fixed neural network architecture with parameters $\theta \in \mathbb{R}^p$. Image recovery in this class is then posed as an optimization problem over the network parameters:

$$\min_{\theta \in \mathbb{R}^p} \|\mathcal{A}f_{\theta} - \mathbf{y}\|_2^2. \quad (4)$$

Typically, the number of trainable parameters P can be chosen to be far less than the number of sinogram measurements M . In this case, problem (4) is not necessarily ill-posed, in the sense that it may have a unique global minimizer.

However, finding a global minimizer of (4) is challenging in general. This is because, in contrast to the pixel-basis approach, the mapping from parameters θ to an image f_{θ} is no longer linear, which means that (4) is a non-linear least squares problem whose objective function is typically non-convex. Nevertheless, standard neural network training methods, such as stochastic gradient descent, can be used to find a local minimizer that fits the data well [5]. Following this fitting procedure, a rasterized image can be generated by evaluating the CBNN f_{θ} on a pixel grid of any desired resolution.

III. SIMULATION DETAILS

Our goal is to compare the performance of pixel-basis formulations (2), (3) based on a discrete-to-discrete model, and the CBNN formulation (4) based on a discrete-to-continuous model, for the recovery of ideal CT data. In particular, we focus on recovery of an analytical phantom for which exact Radon transform samples can be computed. The phantom we use, shown in Figure 1, is a modified version of the Shepp-Logan phantom that includes a detail insert consisting of an 8×8 grid of discs of varying radii and contrasts. In addition, we include two high contrast ovals near the center of the phantom. These ovals induce streaking artifacts that intersect with the detail insert when reconstructing with filtered back

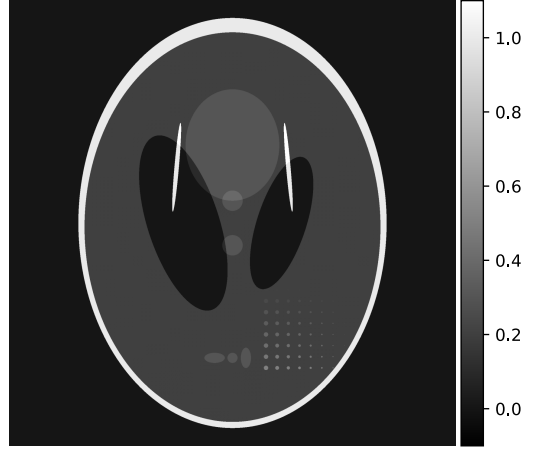


Fig. 1. Modified Shepp-Logan phantom used in our experiments.

projection, which generally makes the faithful recovery of the detail insert features more challenging.

For sinogram measurements, we simulate a circular fan-beam acquisition of the phantom. We assume the object lies in a circular region-of-interest (ROI) with radius 9 cm, and the x-ray source follows a circular trajectory with a radius of 50 cm centered in the middle of the ROI. The source-to-detector-center distance is 100 cm, and the detector is modeled as a linear array. We simulate 128 equally distributed views over a 360 degree scan and the projections are sampled on a 1024-pixel detector, resulting in a total of $M = 128 \times 1024 = 131,072$ sinogram measurements. Analytic line integrals are calculated for the phantom using exact formulas for the Radon transform of the indicator function of an ellipse.

As baselines, we compare against analytic reconstruction using filtered back projection (FBP) and two model-based iterative methods: least squares (LS), and regularized least squares with total variation regularization (TV-RLS). FBP is implemented using a weighted pixel-driven back-projection and a ramp filter [6]. For LS and TV-RLS, we use a standard line-intersection model for the x-ray transform of a discrete image to define the discrete ray transform matrix $\mathbf{A}$. The images are pixelated on a square $n \times n$ pixel grid, where we vary $n \in \{128, 256, 512\}$. For LS we find an approximate solution of (2) by applying the conjugate gradients algorithm to the normal equations with zero initialization and early stopping at 100 iterations, which approximates a pseudoinverse solution. For PLS-TV, we solve (3) using the Chambolle-Pock primal-dual algorithm [7]. To set the regularization strengths λ , we performed a grid search over $\lambda \in \{1, 0.1, 0.01, 0.001\}$ and chose the value that gave the lowest root mean-squared error in an ROI containing the detail insert.

For the CBNN approach, we approximate the continuous domain Radon transform using a Riemann approximation by densely sampling each ray uniformly along a trajectory through the circular ROI with 100 evaluation points per linear cm, such that the longest ray is approximated with 1800 evaluations. As our CBNN architecture, we use a multi-layer perceptron (MLP) with a Fourier features layers as proposed

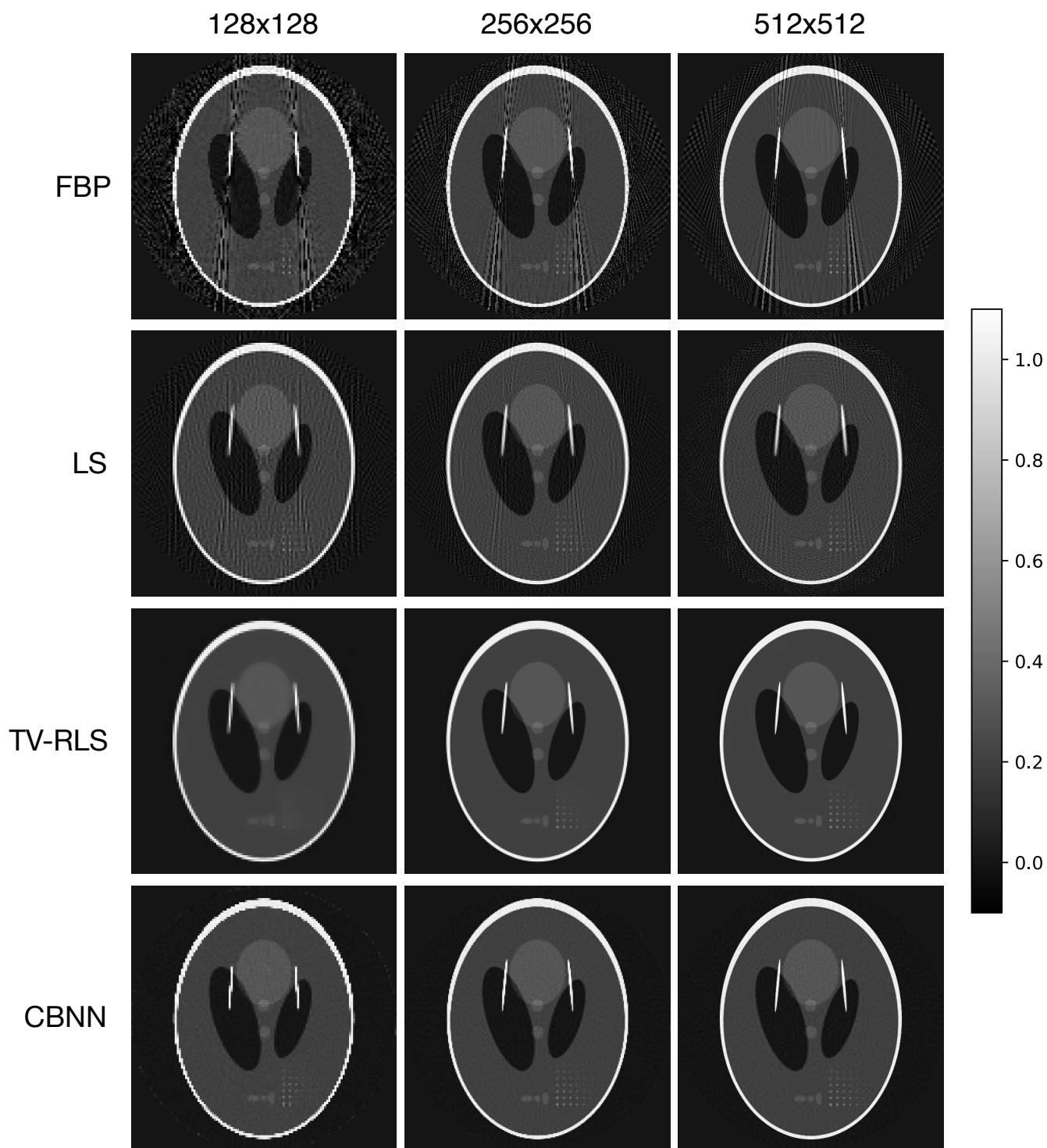


Fig. 2. Reconstruction of analytical phantom from a simulated sparse-view acquisition at different pixel resolutions. Sinogram data is simulated under a circular fan-beam geometry with 128 equi-spaced views and 1024 detector bins using an exact line integral model. Key: **FBP**=Filtered Back Projection, **LS**=Least Squares, **TV-RLS**=Total Variation Regularized Least Squares, **CBNN**=Coordinate-based neural network. Note that the CBNN reconstructions are rasterizations of the same continuous domain function onto different grid sizes.

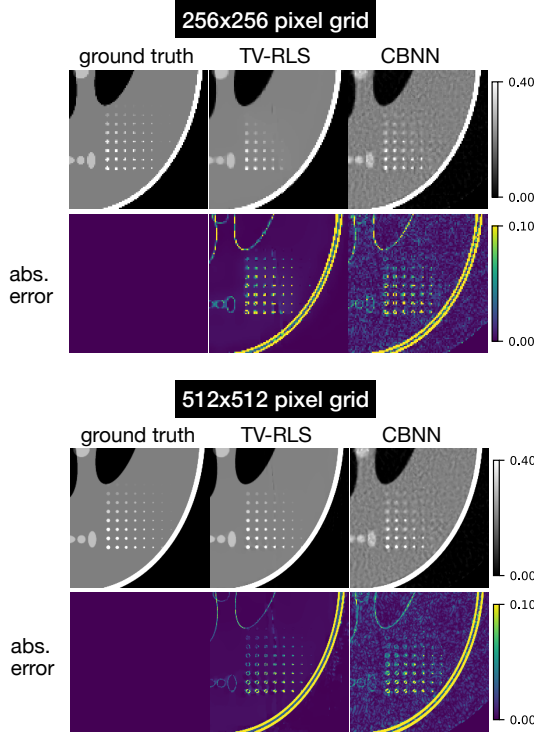


Fig. 3. Blow-up an ROI containing the contrast detail insert.

in [5]. We use an MLP with three hidden-layers with a ReLU activation and width 50 in each layer, resulting in 30801 trainable parameters. For the random Fourier features layer, we sample 256 random frequencies from a standard Gaussian with bandwidth parameter $\sigma = 1.0 \text{ cm}^{-1}$. To train the CBNN, we use a stochastic gradient descent approach where 8 random views are sampled at each iteration and used to approximate the least squares loss (4). The network is trained for a total of 30000 SGD iterations using the Adam optimizer with an initial learning rate of 0.001, which is decreased by a power of 10 every 10000 iterations. Finally, to compare with the discrete reconstruction methods, we rasterize the CBNN on a variety of pixel grid sizes.

IV. RESULTS

Figure 2 shows the reconstructions obtained by all methods at three pixel resolutions. Additionally, Figure 3 shows a blow-up of an ROI containing the detail insert comparing TV-RLS and CBNN reconstructions at 256×256 and 512×512 pixel resolutions. First, we observe that the FBP and LS reconstructions contain significant streaking artifacts at all pixel resolutions, as expected from the sparse-view sampling. Additionally, visible aliasing-like artifacts occur in the FBP and LS reconstructions at the 128×128 resolution. Relative to FBP and LS, both TV-RLS and CBNN reconstructions show essentially no visible streaking or aliasing artifacts. At the two lowest resolutions, there is some loss of fine detail in the TV-RLS reconstructions. In particular, at a 256×256 resolution, some detail insert features are wiped out in the TV-RLS reconstruction due to over-regularization (see Figure 3). However, compared to TV-RLS at a 512×512 resolution, the

CBNN reconstruction has less sharp boundaries and preserves fewer of the discs in the detail insert. Also, the CBNN reconstruction has a uniform noise-like texture that is absent from the TV-RLS reconstruction.

V. DISCUSSION AND CONCLUSION

We investigate the possibility of recovering a continuous domain solution to a sparse-view CT inverse problem using CBNNs and an analytical phantom. Our experiments illustrate that reconstruction with a CBNN can be more accurate than filtered back projection and standard iterative reconstruction techniques, especially when performed on a low resolution pixel grid. However, we observe the CBNN recovery is less accurate than a total variation regularized reconstruction on a sufficiently high resolution grid. This may be in part because the phantom under consideration is a piecewise constant image, which is well-aligned with the total variation prior. An open question is whether CBNNs are better suited to alternative object models, e.g., piecewise smooth images.

Additionally, in our experiments we focused on only one type of CBNN architecture based on [5]. Several alternative architectures have been proposed (e.g., [8], [9]), which primarily differ in their activation functions. Exploring the full impact of changes to the CBNN width, depth, activation function, and training protocol on image estimation is an interesting direction for further study.

ACKNOWLEDGEMENTS

This work was supported by NSF CRII award CCF-2153371.

REFERENCES

- [1] V. Sitzmann, M. Zollhoefer, and G. Wetzstein, "Scene representation networks: Continuous 3d-structure-aware neural scene representations," 2019, pp. 1121–1132.
- [2] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 405–421.
- [3] Y. Sun, J. Liu, M. Xie, B. Wohlberg, and U. S. Kamilov, "CoIL: Coordinate-based internal learning for tomographic imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1400–1412, 2021.
- [4] B. Song, L. Shen, and L. Xing, "PINER: Prior-informed implicit neural representation learning for test-time adaptation in sparse-view ct reconstruction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 1928–1938.
- [5] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.
- [6] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. SIAM, 2001.
- [7] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle-pock algorithm," *Physics in Medicine & Biology*, vol. 57, no. 10, p. 3065, 2012.
- [8] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.
- [9] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, "WIRE: Wavelet implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 507–18 516.