Variational network learning for low-dose CT

Erich Kobler, Matthew J. Muckley, Baiyu Chen, Florian Knoll, Kerstin Hammernik,
Thomas Pock, Daniel K. Sodickson, and Ricardo Otazo

***Abstract*—The reconstruction of high quality images from low-dose X-ray CT scans data is a topic of significant technical and clinical relevance. In this paper, we develop learning-based variational networks (VNs) to reconstruct low-dose 3D helical CT data. We consider two dose reduction methods: (1) x-ray tube current reduction and (2) x-ray beam interruption also known as SparseCT. In the first case we train a VN to denoise the current-reduced reconstruction to account for the smaller signal-to-noise ratio, whereas, in the second case the VNs learn reconstruction schemes that suppress undersampling artifacts. We use 4 clinical abdominal 3D scans to train VNs for 4-fold dose reduction and compare against state-of-the-art model-based denoising and sparse reconstruction methods on a 5th clinical abdominal test scan. The proposed VNs improve performance over state-of-the-art iterative model-based denoising and sparse reconstruction techniques. VNs for SparseCT compare favorably to VNs for current reduction, particularly for reconstruction of small low-contrast features.**

***Index Terms*—CT image reconstruction, SparseCT, low-dose CT, compressed sensing, machine learning, variational networks**

I. Introduction

The increasing utilization of CT scanners in clinical imaging examinations, has triggered the need to reduce the radiation dose, particularly for recurrent studies. One of the most common approaches is to reduce the tube current, e.g., tube current modulation [1], or lower tube currents in conjunction with iterative model-based denoising methods [2]. These techniques have been successfully integrated in commercial scanners, but they only offer moderate radiation dose reductions of 30-40% in practice, due to compromises between denoising and smoothing.

The radiation dose can also be mitigated without reducing the tube current by decreasing the number of X-rays that penetrate a patient during a CT scan. The compressed sensing (CS) theory [3] supports this approach, since CT images are compressible in a transform domain and reducing the number of X-ray projections results in small additive incoherent streaking artifacts. A simple way to omit projections is to perform angular undersampling, i.e., just acquire projections for a fraction of the angular views, as proposed by [4]. The SparseCT method [5] extended this idea by blocking a subset of X-rays in an incoherent way across the angular and slice dimensions, which divides the overall undersampling along multiple dimensions and thus increases the perfromance of CS for reconstruction of the whole volume.

Recent low-dose CT reconstruction algorithms for low-current and/or undersampled data are typically model-based iterative methods that incorporate prior knowledge to increase image quality. These prior models are typically rather simple and model just a small subset of the CT image statistics, e.g., the popular total variation (TV) prior enforces sparsity in the image gradient domain. In addition, the balance between a regularizing prior term and a data fidelity term has to be empirically tuned to generate suitable reconstructions. In accelerated magnetic resonance imaging, deep learning was introduced to overcome this empirical tuning and to learn image models that are tailored towards medical imaging, demonstrating significant improvements over standard compressed sensing algorithms [6]. Likewise, recent work on deep learning for low-dose CT demonstrated improved performance compared to standard denoising and sparse reconstructions [7]–[9]. The U-net-like structures of [8] and [9] as well as the residual encoding network of [7] learn a mapping from low-dose filtered back-projection images to reference images that encodes and decodes the relevant information, in contrast to the step-wise refinement structure of [6].

In this work, we propose to learn variational networks for low-dose CT data acquired with tube current reduction and SparseCT. We train the VNs on four clinical abdominal data sets and evaluate the reconstruction quality of the proposed VNs on a test data set and compare it to state-of-the-art model-based reconstructions.

II. Model-based CT Reconstruction

The process of acquiring CT data of a volume $u \in \mathbb{R}^{M\times N\times D}$ can be formalized as

$$d = Au + n\,, \tag{1}$$

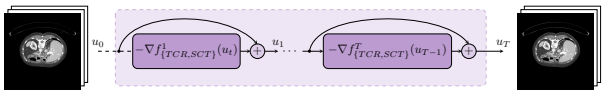
where $d \in \mathbb{R}^P$ is the post-log measured data of P X-ray projections. The random variable n models the effects of quantum and electronic noise and is assumed to be Gaussian due to preprocessing. The linear forward operator $A : \mathbb{R}^{M\times N\times D} \mapsto \mathbb{R}^P$ implements the mapping from the volume to the measurement data that is defined by the scanner geometry. For SparseCT A additionally implements the undersampling pattern.

*Erich Kobler is with the Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria (e-mail: erich.kobler@icg.tugraz.at).

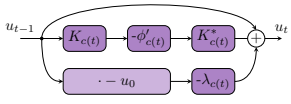
Matthew J. Muckley, Baiyu Chen, Florian Knoll, Daniel K. Sodickson and Ricardo Otazo are with the New York University School of Medicine, New York, NY, 10016 USA.

Kerstin Hammernik and Thomas Pock are with the Institute of Computer Graphics and Vision, Graz University of Technology, Graz, Austria.

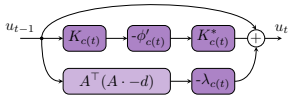
We acknowledge support from the Austrian Science Fund (FWF) under the START project BIVISION, No. Y729, the European Research Council under the Horizon 2020 program, ERC starting grant HOMOVIS, No. 640156. and the National Institutes of Health (NIH) grants U01-EB018760 and P41-EB017183.



(a) Variational Network (VN) structure for CT



(b) VU for CT denoising



(c) VU for CT reconstruction

Fig. 1. (a) Illustration of the VN for CT and the variational units (VU) for (b) CT denoising and (c) CT reconstruction.

For a given noisy and possibly undersampled CT scan data d, the inverse problem of recovering the volume u is usually defined by a variational minimization problem such as

$$\min_u F(u) := \beta\|\nabla u\|_1 + \frac{1}{2}\|Au - d\|_2^2 \,. \tag{2}$$

Here the scalar $\beta \geq 0$ is used to balance the solution between smoothness, which is enforced by the total variation (TV), i.e., ℓ_1-norm of the image gradients, and data fidelity. A suitable algorithm to solve (2) is the primal-dual approach with line search [11], since it requires just a view evaluations of the operator A that are computationally expensive.

III. Variational Networks for CT

Typical optimization schemes for variational imaging models, such as (2), can be implemented using convolutional networks. This observation, inspired [12] to train all parameters of a gradient descent scheme for variational image reconstruction models, i. e., analysis operators, potential functions, weighting and step sizes, from data. Variational networks (VNs) [13] connect this scheme, convolutional neural networks and variational minimization. To adapt VNs for CT, we apply fields-of-experts-type priors [14] of the form

$$R_c(u) = \langle 1, \phi_c(K_c u; W_c)\rangle \tag{3}$$

that are parameterized by a convolution operator $K_c : \mathbb{R}^{M\times N\times D} \mapsto \mathbb{R}^{M\times N\times D\times N_k}$, which stacks N_k 3D convolutions $K_c^i : \mathbb{R}^{M\times N\times D} \mapsto \mathbb{R}^{M\times N\times D}$, and corresponding potential functions $\phi_c^i(\cdot; w_c^i) : \mathbb{R} \mapsto \mathbb{R}$. These functions are point-wisely applied to the corresponding filter response and are parameterized by the weights $w_c^i \in \mathbb{R}^{N_w}$. For the sake of simplicity, we group all these functions into $\phi_c(\cdot, W_c)$ and their parameters $(w_c^i)_{i=1}^{N_k}$ into W_c.

We use this prior model to construct a variational energy that fits into the VN framework [13] and define it as

$$F_{\{TCR,SCT\}} := \sum_{c=1}^{C} f^c_{\{TCR,SCT\}}(u) \tag{4}$$

$$f^c_{\{TCR,SCT\}}(u) = R_c(u) + \frac{\lambda_c}{2} D_{\{TCR,SCT\}}(u) \,, \tag{5}$$

where the data term $D_{\{TCR,SCT\}}(u)$ is adapted according to the dose-reduction approach. In the case of tube current reduction (TCR) we learn to denoise initial low-dose reconstruction, hence we use a simple ℓ_2-norm denoising data term

$$D_{TCR}(u) = \|u - u_0\|_2^2 \,. \tag{6}$$

In the case of SparseCT (SCT) we use the forward operator A and the undersampled data d to enforce data consistency to the undersampled data and facilitate the reconstruction scheme

$$D_{SCT}(u) = \|Au - d\|_2^2 \,. \tag{7}$$

For both low-dose VNs we use a cyclic component selection function, i.e., $c(t) = \mod(t, C)$, and follow [13] to define a variational unit (VU) as

$$u_t = u_{t-1} - \nabla f^{c(t)}_{\{TCR,SCT\}}(u_{t-1}) \,, \tag{8}$$

where the gradients of the energy components are given by

$$\nabla f^c_{TCR}(u) = K_c^* \phi_c'(K_c u; W_c) + \lambda_c(u - u_0) \tag{9}$$

$$\nabla f^c_{SCT}(u) = K_c^* \phi_c'(K_c u; W_c) + \lambda_c A^\top(Au - d) \,. \tag{10}$$

The adjoint operator of K_c is denoted as K_c^* and it is defined as a convolution with all 180 rotated filter kernels followed by a point-wise summation. Figure 1 illustrates the computation outline of a VN for low-dose CT. The input u_0 is transformed into the output u_T by applying T steps of the form (8).

A. Training of VNs for CT

To train a VN for a set of training samples $(u_0^s, u_{tar}^s)_{s=1}^S$, we minimize the problem

$$\min_{\theta\in\mathcal{T}} \frac{1}{2}\sum_{s=1}^{S} \|b^s \odot (u_T^s - u_{tar}^s)\|_2^2 \,, \tag{11}$$

where $\theta = \{W_c, K_c, \lambda_c,\, c = 1 \dots C\}$ holds all the parameters of the VN. As [13], we constrain the parameters to an admissible set $\mathcal{T}$ that enforces $\lambda_c \geq 0$ and that each convolution filter has zero-mean and its ℓ_2-norm lies on the unit ball. We are only interested in reconstructing the central scan regions because of the missing ray density at border regions. Thus we apply a binary mask $b^s \in \{0,1\}^{M\times N\times D}$ that selects the 9 central slices where $u_{tar} \in [0,1]$ and $\odot$ indicates a point-wise multiplication. Note that we rescaled the images such that the HU interval $[-200, 280]$ is mapped to $[0,1]$ to ease training and account for the desired HU range. We solve the constrained training problem (11) by using the Adam optimizer [15] extended by an additional back projection step onto $\mathcal{T}$ after each gradient step. We perform 1000 gradient steps using the default moments of the Adam optimizer and a step size of 1×10^{-2}.

B. Experimental Setup

For the reconstruction of low-dose CT data we apply $T = C = 10$ variational units and use $N_k = 32$ convolution filters of size $11 \times 11 \times 3$ and their corresponding activation functions are parameterized by $N_w = 31$ Gaussian radial basis functions. We scaled the volumes for both training and test data such that the interesting Hounsfield unit interval $[-200, 280]$ is mapped onto $[0,1]$ to ease the training of the parameters. We use 8 filter-function-pairs that are defined

TABLE I
TRAINING AND TEST DATA SETS

	reference tube current mAs	tube voltage kV	radiation dose CTDIvol	gantry rotations -
train	240	120	21.19	16
	240	120	19.01	26
	240	120	22.26	19
	350	120	29.63	20
test	320	100	12.90	17

on the interval $[-4, 4]$ to regularized the entire HU range, whereas, the remaining 24 filter-function-pairs are defined on $[-1, 1]$ to account for the details in the desired tissue interval. In total 126,090 parameters were trained for each VN.

We used four clinical 3D in vivo abdominal CT scans of different patients of a Siemens Definition AS scanner. Table I shows acquisition properties of the train and test scans. In order to fit the CT data reconstruction onto a single GPU, we split the data of each CT scan after a full gantry rotation and ended up with 81 batches for training and 17 test samples. For every sample we reconstructed an imaged volume of size $384 \times 384 \times 30$. The target volumes u^s_{tar} were computed by solving (2) with $\beta = 1$ using [11] on the full-dose CT data. Likewise, the initial reconstructions u^s_0 were generated with $\beta = 1 \times 10^{-9}$ using either simulated fully-sampled low-dose data [16] or binary subsampled full-dose data for SCT. We apply the same W1S4 undersampling pattern as in [17] for a 4-fold dose reduction.

IV. RESULTS

We used the test data set to evaluate the reconstruction quality of the learned VNs for both TCR and SCT for 4-fold radiation dose reduction. Table II depicts a quantitative evaluation of the root mean squared error (RSME) of the proposed VNs and state-of-the-art model-based denoising and reconstruction approaches. In Fig. 2, we qualitatively compare representative abdominal slices reconstructed by the proposed VNs to the full-dose reference, SAFIRE [2] and TV reconstruction.

In the case of tube current reduction, the proposed VN for TCR outperforms SAFIRE [2] in terms of RMSE and also in reconstruction quality. The VN presents a higher noise reduction of the imaged volume, while keeping the fine structure of the vessels in the liver. The resulting images are slightly smoothed though. Since SAFIRE applies an edge-enhancing kernel to highlight edges in the reconstructions, we removed the skin region from the binary mask b in the evaluation process to perform a fair comparison. Fig. 3 depicts the difference to a corresponding reference slice for the considered methods. Clearly, SAFIRE yields higher differences at edge regions but also the remaining regions are rather noisy.

In the case of SparseCT, the trained VN yields a lower RMSE than the TV model-based reconstruction using 4-fold undersampled test data. The VN for SCT removes the aliasing artifacts better than the TV reconstruction, while maintaining the fine vessels in the liver. Moreover, the reconstructions of the VN for SCT present more details than those of the VN for TCR and are also sharper, highlighting the advantages of SparseCT over tube current reduction for the same dose reduction factor. In addition, the reconstructions of a VN for SCT using 6-fold undersampling are shown on the right in Fig. 2. Despite the increased dose reduction, the VN for SCT is able to reconstruct the fine details and remove aliasing artifacts and yield reconstructions with a similar quality.

TABLE II
QUANTITATIVE COMPARISON OF THE DIFFERENT 1/4-DOSE CT METHODS BY MEANS OF RMSE TO THE TARGET u_{tar}, MEASURED IN HU.

SAFIRE [2]	TV	VN TCR	VN SCT
17.75 ± 2.11	8.84 ± 1.20	7.91 ± 0.90	7.72 ± 0.82

V. CONCLUSION

In this work, we extended variational networks to reconstruct CT volumes from low-dose data. We learned VNs for two popular radiation dose reduction methods, namely tube current reduction and SparseCT. The proposed VNs yield reconstructions that outperform state-of-the-art denoising and sparse reconstruction methods for low-dose CT. The VNs present a higher noise and artifact reduction, while fine details such as vessels are properly reconstructed. The learned reconstructions for undersampled data (SCT) show more details and are sharper than the learned denoising scheme for reduced-current data (TCR). Our experiments suggest that the proposed VNs increase the image quality for a given radiation dose and would enable higher radiation dose reductions. Future work includes the extension of the binary undersampling masks of SparseCT to more realistic undersamling masks as in [18]. Additionally, we work on speeding up the training and reconstruction process by means of ordered-subsets.

REFERENCES

[1] Cynthia H. McCollough, Michael R. Bruesewitz, and James JM. Kofler, "CT Dose Reduction and Dose Management Tools: Overview of Available Options," *RadioGraphics*, vol. 26, no. 2, pp. 503–512, 2006.

[2] Katharine Grant and Rainer Raupach, "SAFIRE: Sinogram Affirmed Iterative Reconstruction," *Siemens Medical Solutions Whitepaper*, 2012.

[3] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] Guang-Hong Chen, Jie Tang, and Shuai Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical Physics*, vol. 35, no. 2, pp. 660–663, 2008.

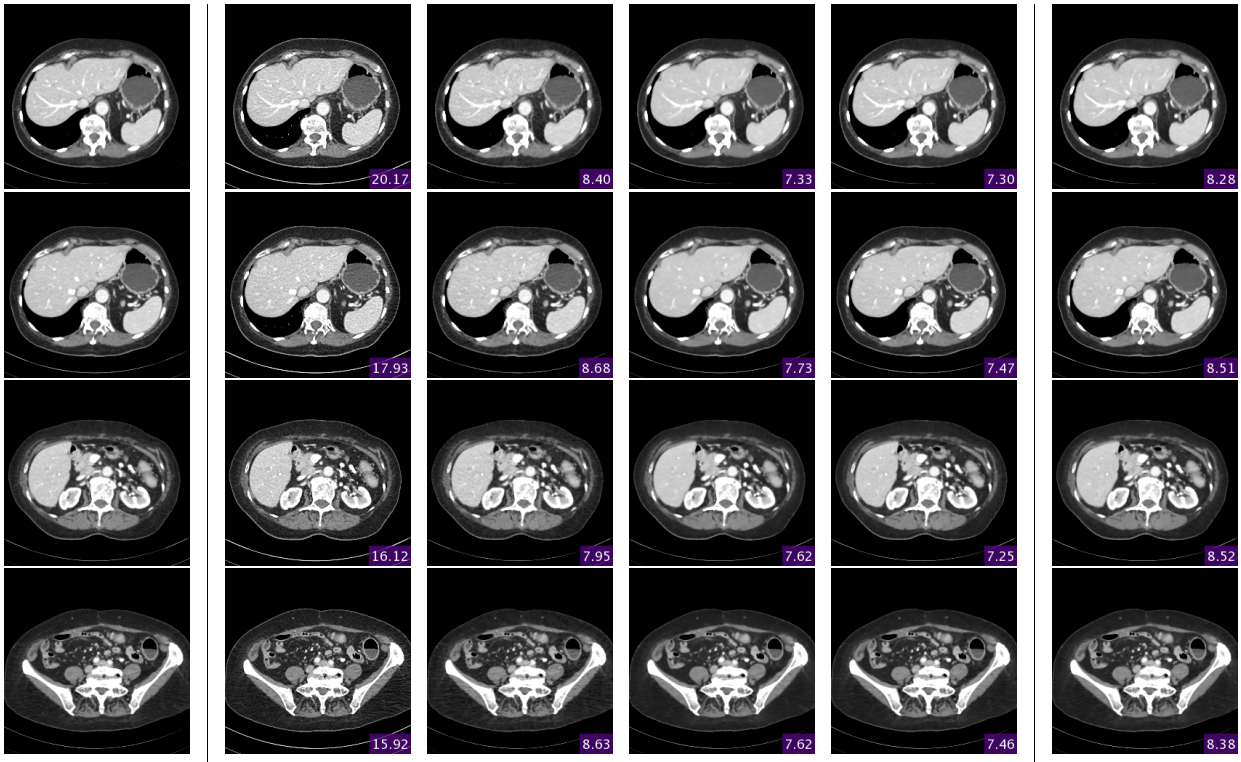
[5] Thomas Koesters, Florian Knoll, Aaron Sodickson, Daniel K. Sodickson, and Ricardo Otazo, "SparseCT: Interrupted-beam acquisition and sparse reconstruction for radiation dose reduction," *Proc.SPIE*, vol. 10132, pp. 10132 – 10132 – 7, 2017.

[6] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll, "Learning a Variational Network for Reconstruction of Accelerated MRI Data," *Magnetic Resonance in Medicine*, 2017, in press.

[7] Hu Chen, Yi Zhang, Jiliu Zhou, and Ge Wang, "Deep learning for low-dose ct," in *Developments in X-Ray Tomography XI*. International Society for Optics and Photonics, 2017, vol. 10391, p. 103910I.

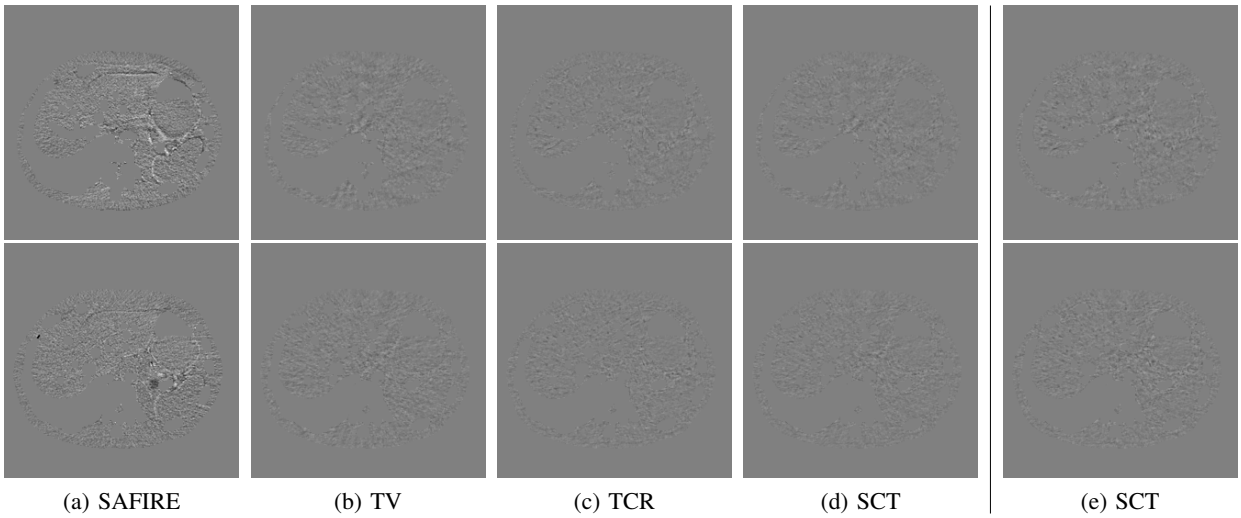
[8] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[9] Han Yo Seob and Jong Chul Ye, "Deep Residual Learning Approach for Sparse-view CT Reconstruction," in *Proceedings of the International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.



(a) full-dose (b) SAFIRE (c) TV (d) TCR (e) SCT (f) SCT

Fig. 2. Representative slices for reconstruction of in vivo abdominal test data for low-dose CT. The purple boxes report RMSE values. (a) Target: TV ($\beta = 1$) reconstruction of the fully-sampled high dose data, (b) SAFIRE [2] using $1/4$ dose, (c) TV ($\beta = 1.75$) reconstruction using 4-fold undersampling, (d) VN for TCR reconstruction using $T = 10$ steps and $1/4$ dose, (e) VN for SCT reconstruction using $T = 10$ steps and 4-fold undersampling, and (f) VN for SCT reconstruction using $T = 10$ steps and 6-fold undersampling.



(a) SAFIRE (b) TV (c) TCR (d) SCT (e) SCT

Fig. 3. Error to the reference reconstruction u_{tar} for the first two slices presented in Fig. 2. (a) SAFIRE [2] using $1/4$ dose, (b) TV ($\beta = 1.75$) reconstruction using 4-fold undersampling, (c) VN for TCR reconstruction using $T = 10$ steps and $1/4$ dose, (d) VN for SCT reconstruction using $T = 10$ steps and 4-fold undersampling, and (e) VN for SCT reconstruction using $T = 10$ steps and 6-fold undersampling. Note that we mapped the HU interval [-150, 150] to [0, 1] to ease visualization.

[10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.
[11] Yura Malitsky and Thomas Pock, "A first-order primal-dual algorithm with linesearch," *arXiv preprint arXiv:1608.08883*, 2016.
[12] Yunjin Chen and Thomas Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1256–1272, 2017.
[13] Erich Kobler, Teresa Klatzer, Kerstin Hammernik, and Thomas Pock, "Variational Networks: Connecting Variational Methods and Deep Learning," in *German Conference on Pattern Recognition*, 2017.
[14] Stefan Roth and Michael J Black, "Fields of Experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.
[15] Diederik P. Kingma and Jimmy Lei Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.
[16] Lifeng Yu, Maria Shiung, Dayna Jondal, and Cynthia H McCollough, "Development and validation of a practical lower-dose-simulation tool for optimizing computed tomography scan protocols," *Journal of Computer assisted Tomography*, vol. 36, no. 4, pp. 477–487, 2012.
[17] Matthew Muckley, Baiyu Chen, Thomas Vahle, Aaron Sodickson, Florian Knoll, Danieln Sodickso, and Ricardo Otazo, "Regularizer performance for sparsect image reconstruction with practical subsampling," in *The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017, vol. 14.
[18] Baiyu Chen, M. Muckley, T. ODonnell, A. Sodickson, T. Flohr, K. Stierstorfer, B. Schmidt, F. Knoll, A. Primak, D. Faul, D. Sodickson, and R. Otazo, "Realistic undersampling model for compressed sensing using a multi-slit collimator," in *International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2017.