# VARIATIONAL DEEP LEARNING FOR LOW-DOSE COMPUTED TOMOGRAPHY

Erich Kobler<sup>1</sup>, Matthew Muckley<sup>2</sup>, Baiyu Chen<sup>2</sup>, Florian Knoll<sup>2</sup>, Kerstin Hammernik<sup>1,2</sup>, Thomas Pock<sup>1,3</sup>, Daniel Sodickson<sup>2</sup> and Ricardo Otazo<sup>2</sup>

<sup>1</sup> Institute of Computer Graphics and Vision, Graz University of Technology, Austria
<sup>2</sup> Department of Radiology, New York University School of Medicine, USA
<sup>3</sup> Digital Safety and Security Department, Austrian Institute of Technology, Austria

## ABSTRACT

In this work, we propose a learning-based variational network (VN) approach for reconstruction of low-dose 3D computed tomography data. We focus on two methods to decrease the radiation dose: (1) x-ray tube current reduction, which reduces the signal-to-noise ratio, and (2) x-ray beam interruption, which undersamples data and results in images with aliasing artifacts. While the learned VN denoises the currentreduced images in the first case, it reconstructs the undersampled data in the second case. Different VNs for denoising and reconstruction are trained on a single clinical 3D abdominal data set. The VNs are compared against state-of-the-art model-based denoising and sparse reconstruction techniques on a different clinical abdominal 3D data set with 4-fold dose reduction. Our results suggest that the proposed VNs enable higher radiation dose reductions and/or increase the image quality for a given dose.

*Index Terms*— computed tomography, medical imaging, compressed sensing, machine learning, variational networks

## 1. INTRODUCTION

Throughout the last decades, computed tomography (CT) has become a standard tool for diagnostic radiology. However, the radiation dose employed in clinical CT examinations has heightened concerns about potential risks for patients undergoing recurrent imaging [1]. To reduce these concerns, numerous radiation dose reduction techniques have been proposed, including tube current modulation [2], adaptive collimators [3], reduced tube current and iterative model-based denoising [4]. However, these x-ray tube current reduction methods are inherently limited by the minimal useful tube current and in practice only moderate dose reductions in the order of 30-40% are clinically accepted.

A different approach to reduce radiation dose is motivated by compressed sensing (CS) theory [5]. The idea of compressed sensing is to exploit the natural fact that images are compressible in a transform domain to significantly reduce the number of measurements without losing relevant information. CT images are naturally compressible and thus especially suitable for the application of CS as omitting projections results in additive incoherent low-value streaking artifacts. The application of CS to CT was already proposed in one of the original CS publications [5], and soon after that, angular undersampling schemes were introduced for 2D imaging [6]. The first practical CS method for clinical scanners (SparseCT) was recently proposed using the concept of a moving multi-slit collimator, which will block x-rays in an incoherent way along the angular and slice dimensions [7]. Instead of omitting entire angular views, which distributes aliasing artifacts only along slices, a different subset of projection slices is blocked for each angle such that the aliasing artifacts are distributed along different views and slices simultaneously. This increases the incoherence of the aliasing artifacts and thus performance of CS reconstruction.

Both denoising of low-current data and reconstruction of undersampled data use an iterative algorithm based on image models and regularization functions to exploit prior knowlege, for example sparsity in a transform domain. Image models are usually very simple (e.g. finite differences) and do not capture the complexity of medical images, and the procedure to tune regularization parameters is empirical. A promising alternative to these limitations in medical imaging is machine learning to learn transforms, regularization functions and parameters from training data sets [8]. Initial work on deep learning for low-dose CT has shown advantages over standard denoising and sparse reconstruction methods [9, 10, 11]. Those approaches applied a U-net-like structure to learn a mapping from an initial low-dose filtered back-projection reconstruction to a full dose reference. [12] extended this idea by learning a convolutional neural network that denoises contourlet transform coefficients of low-dose reconstructions.

In this work, we propose to apply deep learning, in particular variational networks (VNs) [13], to learn suitable de-

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.

noising and reconstruction schemes for low-dose 3D CT, such as tube current reduction and interrupted-beams (SparseCT). Different VNs for denosing and reconstruction are trained using one in vivo clinical 3D abdominal data set. The performance of the VNs for radiation dose reduction is compared against state-of-the-art denoising and sparse reconstruction techniques.

## 2. METHODOLOGY

## 2.1. Data acquisition model

In CT, the post-log data  $d \in \mathbb{R}^P$  consisting of P x-ray projections, can be represented as

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

where  $u \in \mathbb{R}^{M \times N \times D}$  is the imaged volume, which is typically measured in Hounsfield units (HU), and *n* represents the noise of the measurement, which depends on the x-ray tube current and the thermal noise in the detectors and due to preprocessing steps and for the sake of simplicity is assumed to be Gaussian.  $A : \mathbb{R}^{M \times N \times D} \mapsto \mathbb{R}^{P}$  is the forward acquisition operator that incorporates the scanner geometry, and in the case of interrupted-beam (SparseCT), the undersampling pattern.

## 2.2. Variational reconstruction for CT

The inverse problem of reconstructing the image u from the noisy and possibly undersampled measured data d can be solved using the following variational optimization problem

$$\min_{u} F(u) := \beta \|\nabla u\|_{1} + \frac{1}{2} \|Au - d\|_{2}^{2}, \qquad (2)$$

where the  $\ell_1$ -norm of the gradients enforces sparsity of the volume gradients and  $\beta \ge 0$  balances regularization against data consistency. Since the operator A involves the computation of  $P > 10^8$  detector responses, i.e., the line integral along each x-ray, each operator evaluation during the course of optimization is costly. A suitable choice to optimize (2) is the primal-dual algorithm with linesearch [14].

## 2.3. Learning VNs for CT

To avoid empirical tuning of the weighting parameter  $\beta$  and describe higher-order features of CT images, we apply trainable fields of experts [15] type priors  $R_c(u) : \mathbb{R}^{M \times N \times D} \mapsto \mathbb{R}$  of the form

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

where the operator  $K_c : \mathbb{R}^{M \times N \times D} \mapsto \mathbb{R}^{M \times N \times D \times N_k}$  stacks  $N_k$  3D convolution operators  $K_c^i : \mathbb{R}^{M \times N \times D} \mapsto \mathbb{R}^{M \times N \times D}$ . Every convolution filter has a corresponding potential function  $\phi_c^i(\cdot; w_c^i) : \mathbb{R} \mapsto \mathbb{R}$  that is point-wise applied to the filter



(a) Variational Network (VN) structure for CT



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

response and parameterized by  $w_c^i \in \mathbb{R}^{N_w}$ . All these functions are grouped into  $\phi_c(\cdot, W_c)$  and their parameters  $(w_c^i)_{i=1}^{N_k}$  into  $W_c$ .

VNs [13] are applied based on the above regularization for two cases of radiation dose reduction: (a) tube current reduction (TCR) and (b) interrupted-beam or SparseCT (SCT). In the TCR case, we learn a VN for "denoising" an initially reconstructed volume  $u_0$ . The corresponding energy of this VN is defined as

$$F_{TCR} := \sum_{c=1}^{C} f_D^c(u) = R_c(u) + \frac{\lambda_c}{2} \|u - u_0\|_2^2.$$
(4)

In the SCT case, we include the forward operator and the measured data in the reconstruction process. The forward operator adds additional knowledge about the known undersampling pattern to facilitate reconstruction from undersampled data. Consequently, the related energy of this reconstruction VN reads as

$$F_{SCT} := \sum_{c=1}^{C} f_R^c(u) = R_c(u) + \frac{\lambda_c}{2} \|Au - d\|_2^2.$$
 (5)

The parameters of each corresponding VN are grouped into an individual parameter set  $\theta_{\{TCR,SCT\}} = \{W_c, K_c, \lambda_c, c = 1...C\}$ . For both VNs we define the variational unit in analogy to [13] as

$$u_t = u_{t-1} - \nabla f^{c(t)}_{\{TCR,SCT\}}(u_{t-1}), \qquad (6)$$

where the component selection function is c(t) = mod(t, C) + 1 and the gradients of the energy components are defined as

$$\nabla f_{TCR}^c(u) = K_c^\top \phi_c'(K_c u; W_c) + \lambda_c(u - u_0) \tag{7}$$

$$\nabla f_{SCT}^c(u) = K_c^\top \phi_c'(K_c u; W_c) + \lambda_c A^\top (Au - d) \,. \tag{8}$$

The output  $u_T$  of the VNs is generated by applying T steps of the form (6) starting with  $u_0$ , see Fig. 1.

We parameterize the activation functions  $\phi_c^i(\cdot)$  by linearly interpolating between  $N_w = 31$  equally distributed bins in the interval [-1.2, 1.2]. In each of the T = C = 10 steps, we use  $N_k = 48$  convolution filter of size  $15 \times 15 \times 3$ , which results in a total number of 338,890 parameters. Note that the VNs in this setup can be interpreted as trainable reaction diffusions [16].

Given a set of training samples  $(u_0^s, u_{tar}^s)_{s=1}^S$ , we define the training problem of the VNs as

$$\min_{\theta \in \mathcal{T}} \sum_{s=1}^{S} \sum_{t=1}^{T} \frac{\alpha_t}{2} \| b^s \odot (u_t^s - u_{tar}^s) \|_2^2,$$
(9)

where  $u_t^s$  is the output of a VN step and  $\alpha_t$  is 1 for t = Tand 0.1 else. We use the binary mask  $b^s \in \{0,1\}^{M \times N \times D}$  to train on the Z central slices where  $u_{tar}^s \in [0,1]$  such that the background and the sparsely sampled volume edges along the z-axis are masked out. The symbol  $\odot$  indicates a point-wise multiplication. In analogy to [13] the set  $\mathcal{T}$  enforces constraints on the parameters such as  $\lambda_c \ge 0$  and that all convolution filters have zero-mean and their 2-norm is bounded by 1. We minimize (9) using the Adam optimizer [17] and set the step size to 4e-4 and the moments to  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ . Note that after each gradient step of Adam we perform a back projection of the parameters  $\theta_{\{TCR,SCT\}}$  onto  $\mathcal{T}$ .

#### 2.4. Experimental setup

To train and test the models, we used two 3D in vivo abdominal CT data sets acquired on different patients using a Siemens Definition AS scanner. The training scan was acquired with tube current modulation turned on, using a reference mAs of 320 and tube voltage 100 kV (CTDIvol = 12.9 mGy), whereas, the test scan used a reference mAs of 240 and tube voltage of 120 kV (CTDIvol = 21.19mGy). We split the CT data of each scan into batches such that each batch contained all projections of a full gantry rotation P =108, 527, 616, resulting in 17 training samples and 16 test samples. For each batch we reconstructed a volume of size  $384 \times 384 \times 30$  and used the Z = 9 central slices in the computation of the loss (9). We computed the target volumes  $u_{tar}^s$  by solving(2) with  $\beta = 1$  using [14] on the full dose and fully-sampled CT data. In the same fashion we computed the initial estimates  $u_0^s$  with  $\beta = 1e-9$  using the fully-sampled low-dose data, simulated as in [18], in the TCR case and the subsampled full dose data in the SCT case. In addition, we scaled the volumes for both training and test set such that the interesting Hounsfield unit interval [-200, 280] is mapped onto [0, 1] to ease the training of the parameters.

#### 3. RESULTS

We evaluated the reconstruction quality of the proposed VNs for TCR and SCT on the test scan. In Table 1, we quantitatively compared our results to state-of-the-art reconstruction and denoising CT methods by means of root mean squared error (RMSE) to the target reconstruction. Additionally, Fig. 2

	VN	TV $\beta = 2$	SAFIRE [4]
TCR	$14.49 \pm 2.55$	-	$24.13 \pm 2.98$
SCT	$12.27 \pm 1.34$	$13.46 \pm 1.73$	-

**Table 1.** Quantitative comparison of the different CT methods by means of RMSE to the target  $u_{tar}$ , measured in HU. The results in the first row were computed using a 75% dose reduction by means of TCR and the VN for CT denoising  $F_{TCR}(u)$ . In the second row the results were computed using the same dose reduction with SCT and the VN for CT reconstruction  $F_{SCT}(u)$ .

and Fig. 3 depict abdominal slices for both considered dose reduction methods for qualitative comparison.

The proposed VN for TCR outperforms SAFIRE [4], which is a commercial state-of-the-art denoising technique from Siemens, in terms of noise reduction and RMSE, at the expense of slight smoothing. Note that in the TCR case neither SAFIRE nor the learned VN were capable of re-constructing the fine vessels in the central slice in Fig. 2, indicated by the red ellipses.

In the case of SparseCT, the learned VN reconstruction presents higher removal of aliasing artifacts and less smoothing in conjunction with lower RMSE than the TV reconstruction of the 4-fold undersampled data. Moreover, the fine vessels in the liver are captured by both VN and TV reconstruction, as highlighted by the red ellipses. The improved performance of the VN for SCT over TV is likely due to better representation of complex image texture and the selection of regularization parameters. The VN for SCT also presented lower RMSE than the VN for TCR (Table 1), which highlights the advantages of beam-interruption over tube current reduction for the same dose reduction factor.

### 4. CONCLUSION

The learned VNs outperform state-of-the-art denoising and sparse reconstruction techniques for low-dose CT, which would enable higher radiation dose reductions and/or increases the image quality for a given dose. Compared to other deep learning approaches, our proposed VN architecture does not require several data sets due to the small model size and the large dimensionality of 3D CT data. Our learned reconstruction for undersampled data (SCT) presents improved performance compared to the learned denoising of reduced-current data (TCR), which is in concordance with previous results and in part due to the reduced sensitivity of undersampling to electronic noise compared to current reduction. Future work will study the use of ordered-subsets to speed up the training process and evaluate the performance of the variational deep learning for higher radiation-dose reductions and realistic undersampling cases based on the multi-slit collimator proposed in SparseCT.



**Fig. 2**. Representative slices for denoising of in vivo abdominal test data using 75% dose reduction by TCR. The purple boxes report RMSE values. (a) Target: TV ( $\beta = 1$ ) reconstruction of the fully-sampled high dose data, (b) VN reconstruction using T = 10 steps and (c) SAFIRE [4] reconstruction. The VN presents higher noise reduction than SAFIRE without compromising image resolution. However, low-contrast features, such as the vessels in the red ellipse, were not adequately reconstructed.



**Fig. 3**. Representative slices for reconstruction of in vivo abdominal test data using 4-fold undersampling (SCT). The purple boxes report RMSE values. (a) Target: TV ( $\beta = 1$ ) reconstruction of the fully-sampled high dose data, (b) VN reconstruction using T = 10 steps and (c) TV reconstruction with  $\beta = 2$ . The VN outperforms TV in terms of residual aliasing suppression. It is also able to adequately reconstruct the vessels in the liver, which shows the advantages of SCT over TCR for CT.

## 5. REFERENCES

- David J. Brenner and Eric J. Hall, "Computed Tomography An Increasing Source of Radiation Exposure," *New England Journal of Medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.
- [2] Cynthia H. McCollough, Michael R. Bruesewitz, and James JM. Kofler, "CT Dose Reduction and Dose Management Tools: Overview of Available Options," *Radio-Graphics*, vol. 26, no. 2, pp. 503–512, 2006.
- [3] Paul D. Deak, Oliver Langner, Michael Lell, and Willi A. Kalender, "Effects of Adaptive Section Collimation on Patient Radiation Dose in Multisection Spiral CT," *Radiology*, vol. 252, no. 1, pp. 140–147, 2009.
- [4] Katharine Grant and Rainer Raupach, "SAFIRE: Sinogram Affirmed Iterative Reconstruction," *Siemens Medical Solutions Whitepaper*, 2012.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.

- [11] Hu Chen, Yi Zhang, Mannudeep K Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang, "Lowdose ct with a residual encoder-decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.
- [12] Eunhee Kang, Junhong Min, and Jong Chul Ye, "A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction," *Medical physics*, vol. 44, no. 10, 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] Yura Malitsky and Thomas Pock, "A first-order primal-dual algorithm with linesearch," *arXiv preprint arXiv:1608.08883*, 2016.
- [15] Stefan Roth and Michael J Black, "Fields of Experts," *International Journal of Computer Vision*, vol. 82, no. 2, pp. 205–229, 2009.
- [16] 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.
- [17] Diederik P. Kingma and Jimmy Lei Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations*, 2015.
- [18] 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.