

A residual U-Net network with image prior for 3D image denoising

J. F. P. J. Abascal
Univ Lyon, INSA-Lyon,
CNRS UMR 5220, U1206

Lyon, France
juan.abascal@creatis.insa-lyon.fr

S. Bussod
Univ Lyon, INSA-Lyon,
CNRS UMR 5220, U1206

Lyon, France
Suzanne.Bussod@creatis.insa-lyon.fr

N. Ducros
Univ Lyon, INSA-Lyon,
CNRS UMR 5220, U1206

Lyon, France
Nicolas.Ducros@creatis.insa-lyon.fr

S. Si-Mohamed
Univ Lyon, INSA-Lyon,
CNRS UMR 5220, U1206

Lyon, France
Salim.Si-Mohamed@creatis.insa-lyon.fr

P. Douek
Univ Lyon, INSA-Lyon,
CNRS UMR 5220, U1206

Lyon, France
philippe.douek@creatis.insa-lyon.fr

C. Chappard
B3OA, CNRS UMR 7052, U 1271
University Paris Diderot

Paris, France
christine.chappard@inserm.fr

F. Peyrin

Univ Lyon, INSA-Lyon,
CNRS UMR 5220, U1206,
Lyon, France
peyrin@esrf.fr

Abstract—Denoising algorithms via sparse representation are among the state-of-the art for image restoration. On previous work, we proposed SPADE - a sparse- and prior-based method for 3D-image denoising. In this work, we extend this idea to learning approaches and propose a novel residual-U-Net prior-based (ResPrU-Net) method that exploits a prior image. The proposed ResPrU-Net architecture has two inputs, the noisy image and the prior image, and a residual connection that connects the prior image to the output of the network. We compare ResPrU-Net to U-Net and SPADE on human knee data acquired on a spectral computerized tomography scanner. The prior image is built from the noisy image by combining information from neighbor slices and it is the same for both SPADE and ResPrU-Net. For deep learning approaches, we use four knee samples and data augmentation for training, one knee for validation and two for test. Results show that for high noise, U-Net leads to worst results, with images that are excessively blurred. Prior-based methods, SPADE and ResPrU-Net, outperformed U-Net, leading to restored images that present similar image quality than the target. ResPrU-Net provides slightly better results than SPADE. For low noise, methods present similar results.

Index Terms—Image denoising, deep learning, U-Net

I. INTRODUCTION

Sparse methods have been among the state-of-the art denoising methods for many years. Total variation (TV) preserves

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 701915. It was also performed within the framework of the LabEx PRIMES (ANR-11-LABX-0063) of University de Lyon and under the support of the ANR project SALTO (ANR-17-CE19-0011-01) . The project has also received funding from the European Union’s Horizon 2020 research and innovation programme under the grant agreement N 668142. This work was partly funded by France Life Imaging (grant ANR-11-INBS-0006) from the French Investissements d’Avenir.

edges and was termed as a “proper norm” for images [1]. However, TV leads to the so called “staircase artefacts” when images are highly corrupted by noise. To mitigate these effects generalizations of TV, such as total generalized variation [2] and nonlocal TV [3], and patch-based [4] and dictionary learning (DL) [5] methods were proposed.

Recently, a previous work presented a new paradigm for 3D image denoising based on adding a regularization functional that promoted sparsity with respect to a prior image that contained information from several slices in the 3D volume [6]. Authors showed that an implementation of this approach, named SPADE, which was a modification of TV with the additional prior term, led to restored images without the “staircase artefacts”. The prior image has two important benefits. First, it allows the restored image to preserve the image texture of the prior image. Second, the prior image guides the algorithm leading to a more stable convergence. This stability permits to highly iterate the algorithm in order to recover finer details in the image without introducing artefacts.

In the last years, deep convolutional neural networks (CNN) have shown outstanding results in several image processing tasks, including classification and segmentation [7]. CNNs have been also proposed for learning several inverse problems, including image denoising and image reconstruction [8]–[10]. Among the diverse CNN configurations, U-net has shown outstanding results for image segmentation [11] and has been applied to postprocessing CT images [10], [12].

The goal of this work is to propose a residual-U-Net prior-based (ResPrU-Net) architecture for 3D image denoising and to validate it on human knee data acquired on a spectral com-

puterized tomography (SCT) scanner. The proposed ResPrU-Net architecture has two inputs, the noisy image and the prior image, and a residual connection that connects the prior image to the output of the network. By adding the residual connection, the network learns the difference with respect to the prior image, and by including the prior in the input, the network is guided by the prior image in the restoration process. The idea is to extend to learning approaches the prior guidance seen in SPADE method [6]. ResPrU-Net is compared to U-Net and SPADE methods. For simplicity, denoising is done in 2D for all methods, but the prior image for each slice is computed as a running average of neighbor slices, which contains 3D information and has a high signal-to-noise ratio. The same prior image is used for all methods.

II. METHODS

A. SPADE denoising

Let \tilde{u}_i be a 2D noisy slice in the 3D image, u_i^p its corresponding prior image that is assumed to be known and u_i the denoised image, SPADE imposes both spatial regularization and sparsity with respect to a prior image by solving the following problem:

$$\min_u (1 - \alpha) \|\Psi u_i\|_1 + \alpha \|\Phi(u_i - u_i^p)\|_1$$

such that $\|u_i - \tilde{u}_i\|_2^2 \leq \sigma^2$, for $i = 1, \dots, I$, (1)

where I is the number of slices in the image volume. Operators Ψ and Φ are chosen to be equal to the gradient as in [6] ($\nabla_{x,y}$ here), so SPADE minimizes a combination of $TV(u_i)$ and $TV(u_i - u_i^p)$. The last term in (1) constrains the denoised image to lie close to the noisy image, assuming that noise follows a Gaussian distribution.

To choose the prior image, we follow the method used in previous work [6]. The prior image is obtained by combining information from neighbor slices by using a running average to provide a specific prior image for each slice. Each prior slice u_i^p is built from the 3D noisy image as follows

$$u_i^p = \sum_{j=1}^M \theta_{ij} \tilde{u}_j = \sum_{j=1}^M \exp(-\gamma|j-i|) \tilde{u}_j, \quad (2)$$

where $\gamma = 0.04$ as set in previous work, $j = 1, \dots, M$ is the index of the slice and M is the number of slices along the z-dimension in the stack used for computing the prior ($M = 10$ here). The prior image, which is a low-pass filtered version of the noisy image, has low noise but it is obviously blurred.

B. Deep learning denoising with ResPrU-Net

Deep learning has been recently proposed for learning inverse problems [13]–[15]. In this work, we aim to learn the mapping from a noisy image \tilde{u}_i to a restored image u_i , knowing the prior image u_i^p , given by

$$h(\tilde{u}_i; \beta | u_i^p) = u_i, \quad (3)$$

where β indicates the parameters of the neural network. With this framework, we learn the mapping conditioned on the prior image.

To achieve this, we propose a residual-U-Net prior-based (ResPrU-Net) architecture that exploits the knowledge of a prior image. To build ResPrU-Net, we start with a U-Net [11] architecture with two inputs, the noisy image and the prior image, and modify it by adding a residual connection that connects the prior image directly to the output. With this design we gain from both the multi-scale nature of U-Net and the prior image. The starting U-net consists of 14 hidden layers. The contracting path comprises 3×3 convolutions ('same' padding convolutions) each followed by a rectified linear unit (ReLU) and alternated with a 2×2 max pooling operation with stride 2 for downsampling by half every two convolution operations. At each downsampling the number of feature channels is doubled, with 32 channels in the first level and 128 channels in the last level. The expansive path comprises same convolution operations as in the contracting path but alternated with upsampling by two every two convolution operations. The final layer is a 1×1 convolution used to map the 32 channel layer to the output, which is a 2D slice. A sketch of ResPrU-Net architecture is given in Fig. 1.

Learning means finding the parameters β that minimizes the following loss function

$$L(\beta) = \sum_{n=1}^N \|h(\tilde{u}_i^{(n)}; \beta | u_i^{p,(n)}) - u_i^{(n)}\|^2, \quad (4)$$

where $(\tilde{u}^{(n)}, u^{p,(n)}; u^{(n)})$ are input-output image pairs of size $(P \times P)$ and N is the number of image pairs in the training set. Minimization of (4) was done with Adam [16] under TensorFlow [17], with learning rate 10^{-5} and batch size of 32. Training and test losses were computed during training and early stopping was adopted to avoid overfitting. We selected 50 epochs for training on both U-Net and ResPrU-Net.

C. Data sets and image analysis

Methods were tested on a numerical phantom of the human knee and on experimental knee data. Experimental data consisted of excised human knee ($n = 7$) images acquired at Philips prototype SCT scanner, at CERMEP, Lyon [18]. Knee samples were taken from the Institut d'Anatomie Paris Descartes and provided by B2OA (Bioingénierie et Bioimagerie ostéo-articulaire) CNRS 7052, Paris Diderot University. The study was approved by the ethics committee of Descartes University, Paris. To get a ground truth, we applied SPADE to the acquired SCT image at 60 keV, using 40 iterations and $\mu = \lambda = 1$ as described in [6]. Spectral CT images were of size $640 \times 640 \times 400$. Realistic numerical phantoms were built from the experimental data by adding additive Gaussian noise to the original data sets. We considered two levels of noise: 2 % and 5% of the maximum of the image. In the case of experimental data, no additional noise was added to the images provided by the scanner. We used 4 data sets for training, 1 for validation and 2 for test. Training data consisted of 1600 images (from four phantoms) that have been randomly shuffled and normalized to its maximum value. Data augmentation was done using several strategies. During learning, we apply

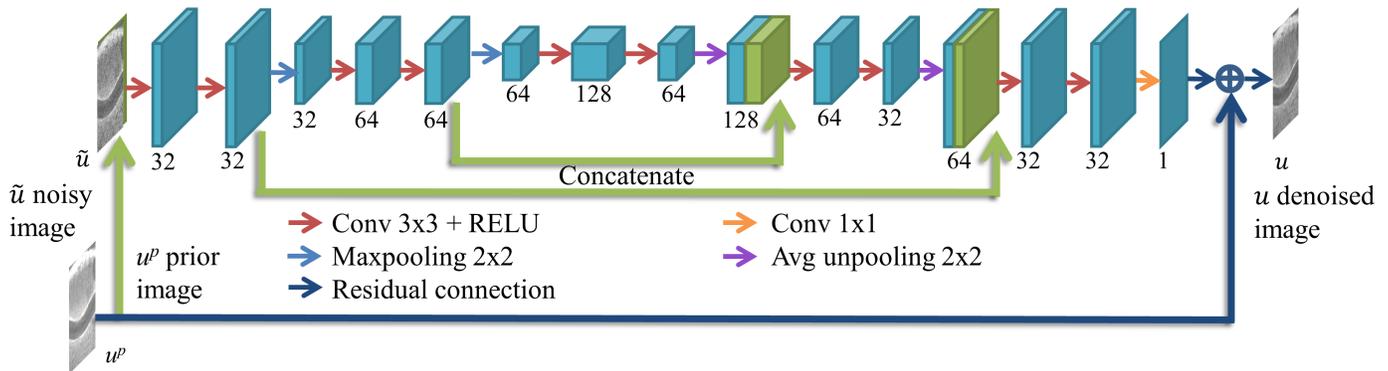


Fig. 1. ResPrU-Net architecture for image denoising. Input is the concatenation of the noisy image and the prior image, the output is the restored image, and the prior image is directly connected to the output for residual learning.

random crop to a final size of 256×256 , mirroring in two dimensions and added noise during training.

We remark that for experimental data (noiseless case), deep learning strategies are learning the output of SPADE but for the noisier case, they have the potential to improve SPADE result.



Fig. 2. Two experimental SCT images at 60 keV of human knee. Left: control. Right: patient with advanced osteoarthritis.

We compare ResPrU-Net, U-Net and SPADE methods. Restored images are assessed in terms of structural similarity (SSIM) with respect to the ground truth image, streak artefact measure (SAM), and noise. SAM is the total variation of the denoised image subtracted to the ground truth image, so it provides an estimate of how well edges are recovered [19]. SAM accounts for noise, artefacts and texture recovery. Noise is computed as the standard deviation (SD) on a small homogeneous region.

III. RESULTS

A. Numerical knee phantom data

Fig. 3 shows denoising results for a knee phantom for 2 % and 5 % additive Gaussian noise. For low noise, all methods provide satisfactory results, removing noise while maintaining image details and borders. For high noise, U-Net leads to worse results, providing excessively blurred images. On the other hand, prior-based methods (SPADE and ResPrU-Net) are able to remove noise while maintaining image quality, as they are guided by the prior image. This can be clearly appreciated in Fig. 4. Between SPADE and ResPrU-Net, SPADE presents a noisier texture. In all cases, ResPrU-Net not only recovers

similar image quality to the target but even seems to remove the noise pattern present in the target image. With regard to the noise pattern in SPADE, it is due to the fact that a large number of iterations have been used to recover well image details (20 iterations for high noise and 40 for low noise), as discussed in [6].

Metrics for the different methods are shown in Table I. In terms of image noise, U-net leads to the lowest noise level, followed by ResPrU-Net. In terms of SSIM and SAM, ResPrU-Net and SPADE provide best results.

TABLE I
METRICS FOR KNEE PHANTOM. LEFT: LOW NOISE. RIGHT: HIGH NOISE

	SSIM	SAM	SD	SSIM	SAM	SD
SPADE	0.98	309	0.018	0.91	599	0.029
U-Net	0.96	404	0.010	0.87	645	0.006
ResPrU-Net	0.98	325	0.015	0.91	541	0.015
Prior	0.91	491	0.014	0.88	620	0.022

B. Experimental data

Fig. 5 shows the denoising results by all methods applied to the experimental SCT image without extra additive noise added to it. All methods provide similar results in terms of noise removal, as expected, as the target for deep learning methods are the results provided by SPADE.

C. Computation time

The code was run on a Windows computer with a 64-bit operating system, Intel Xeon(R) E5-1650 v4 3.60 GHz CPU and 128 GB RAM. Training for deep learning methods was performed using GeForce NVIDIA GTX 1080 Ti graphics card and required 4.2 min per epoch and 6 hours to complete 50 epochs (evaluation error increased after 50 epochs). Image prediction to restore one data set of size $640 \times 640 \times 400$ took 8.6 min on CPU and 13 s on GPU, for deep learning approaches, and 5.2 min on CPU for SPADE using straight-forward parallelization with four cores.

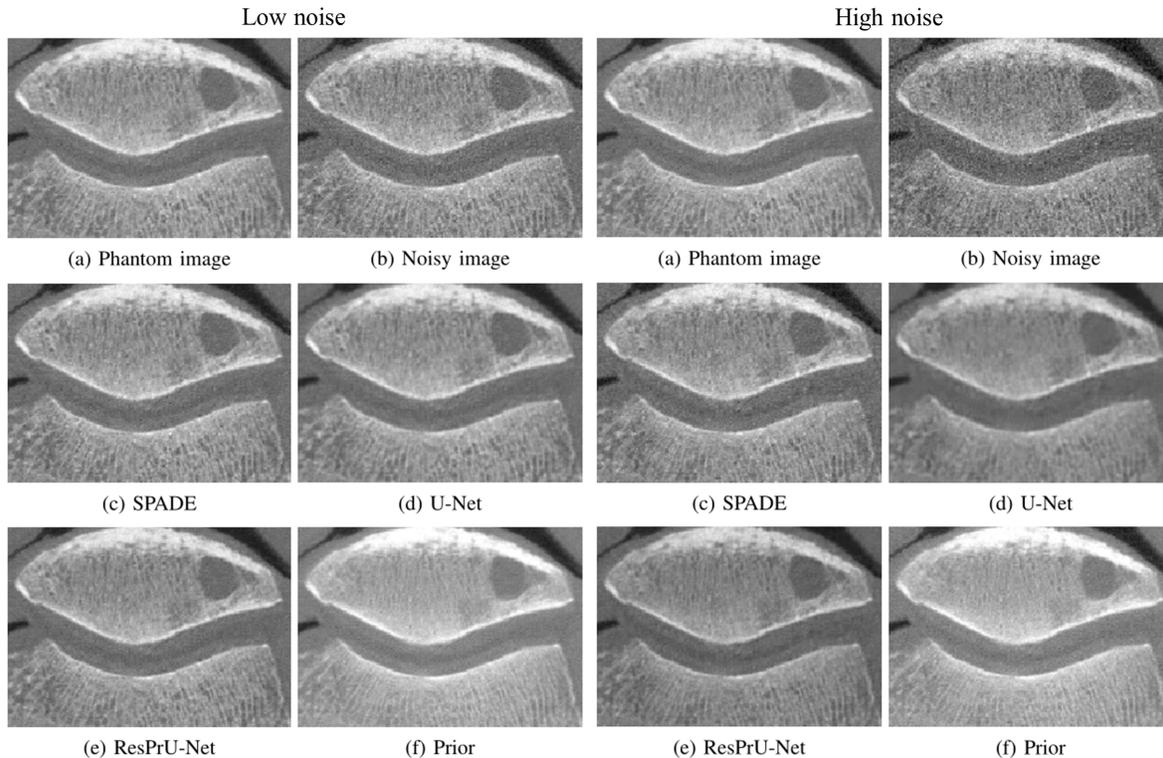


Fig. 3. Denoising results for knee phantom with low (left) and high (right) noise (2 % and 5 % additive Gaussian noise, respectively). a) Detail of knee phantom, b) noisy image, and denoised images by c) SPADE, d) U-Net and e) ResPrU-Net. f) Prior image used by SPADE and ResPrU-Net.

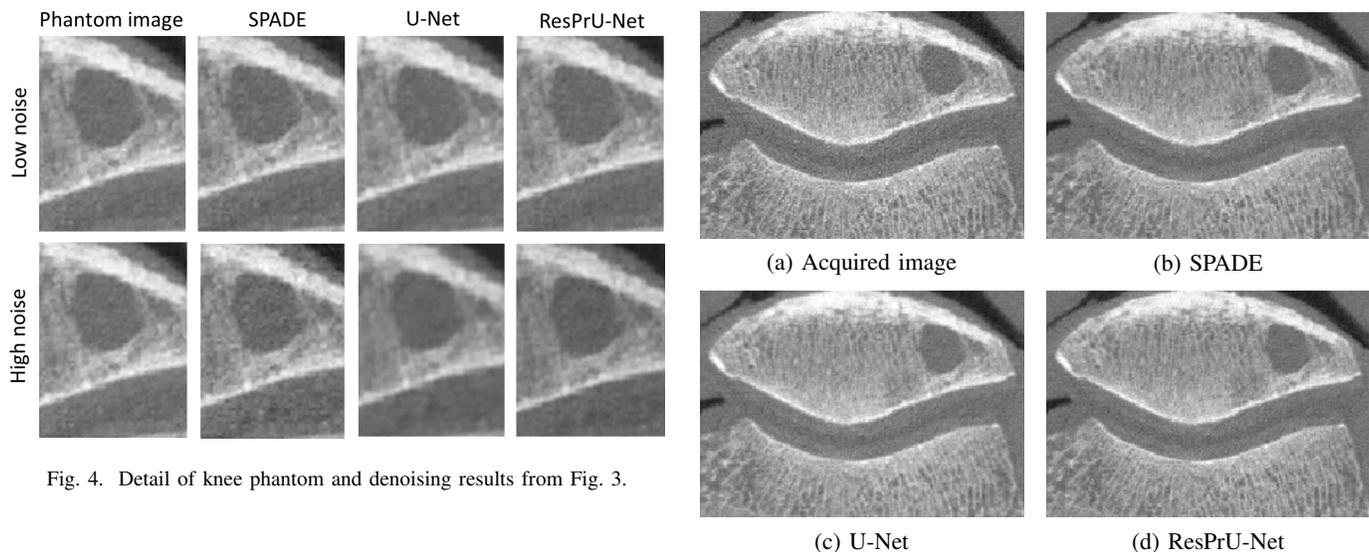


Fig. 4. Detail of knee phantom and denoising results from Fig. 3.

IV. DISCUSSION

We have proposed a novel deep residual learning method for 3D image denoising that allows exploiting a prior image and has been validated on spectral CT data. The proposed architecture - ResPrU-Net - builds from U-Net and includes a residual connection from the input prior image to the output, inspired on the previously proposed prior-based method SPADE [6]. ResPrU-Net is compared to U-Net and SPADE.

ResPrU-Net outperforms both U-Net and SPADE when the image is highly corrupted by noise. In this case, U-Net leads

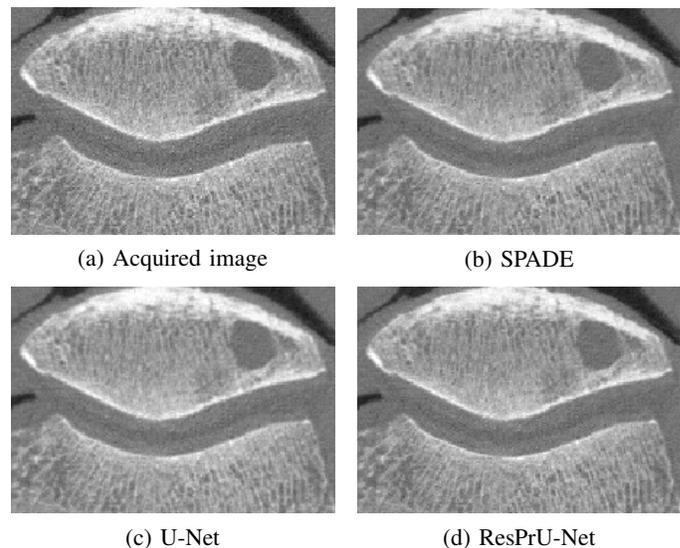


Fig. 5. Experimental data. a) Detail of a slice of acquired data set and denoised results by b) SPADE, c) U-Net and d) ResPrU-Net.

to significantly worse results, providing excessively blurred images. We remark that for the same level of noise other state-of-the-art methods, such as TV and dictionary learning methods provided poor results as presented in [6]. We also remark that U-Net is a 2D denoising method and we did not compare to 3D-U-Net with 3D convolutions. Nevertheless, in

[6] it was shown that neither 2D-TV nor 3D-TV provided similar image quality as SPADE method. SPADE is guided by the prior image, which provides two benefits. On the one hand, it allows SPADE to preserve the image quality of the prior image, and, on the other hand, it stabilizes the convergence allowing to SPADE to do further iterations than TV in order to further recover details without leading to TV spurious artefacts. In the current work, we show that ResPrU-Net not only presents similar results than SPADE in terms of removing noise and recovering image details, but it also leads to images with lower noise than SPADE. SPADE can present slight noise pattern when it is highly iterated. This result was unexpected as the target for ResPrU-Net is the output of SPADE denoising on the acquired image. This could be explained by the fact that ResPrU-Net is learning the probability distribution of the target conditioned to the given prior image, which does not have noise. Another possible explanation is that we are using an early stopping criterion on a encoder-decoder network. For instance, U-Net provide extremely blurred images at the first epochs and images with refined details before 50 epochs. ResPrU-Net provides images similar to prior image at the first epochs and well restored images that are different from the prior at higher epoch number.

For low noise, all methods provided very similar results, as deep learning methods have the output of SPADE denoising as target and in this case SPADE did not present the noise pattern that appeared in the high noise case. We expect that deep learning strategies would lead to superior results if a better target is provided. We will explore this in future work. The easiest solution is to acquire some new samples with higher dose if possible. Nevertheless, recent work has provided an alternative solution that can provide further improvement by learning from noisy images [20].

This work is subject to few limitations. The prior image used for both SPADE and ResPrU-Net has not been optimized. It could be improved, for instance, by learning the weights used for the running average for a specific application. The prior could be also learned by another network. Another obvious improvement is to use a 3D U-Net architecture to further exploit 3D information. There is also the possibility of incorporating the proposed framework based on a prior image and residual connection on previously proposed network configurations [10], [12]. With regard to training data, it consisted in numerical phantoms built from experimental knee images corrupted with additive Gaussian noise. Although a Gaussian noise model could be little too simplistic, for high noise, U-Net provided unsatisfactory results. This may be because the inherent noise and image details present in the knee images make the denoising task more challenging. On experimental data, all methods did extremely well but these data had low noise. More realistic simulated noise, noisier experimental data and learning directly from experimental data will be considered in future work. In this work, we have compared ResPrU-Net to U-Net and to the prior-based denoising method SPADE. In previous work, we found SPADE to provide superior results than TV and patch-based dictionary

learning method [5]. In future work, we will compare to other state-of-the art patch-based methods such as BM3D [21] as well as other deep learning strategies.

In conclusion, we have proposed a novel residual deep learning architecture that can exploit a prior image for 3D image denoising and have validated it on spectral CT images. Restored images by the proposed method provide better visualization of bone and cartilage tissue, which is relevant for diagnosis of knee osteoarthritis. Given the significant improvement with respect to U-Net, we can expect this architecture to be widely used for image denoising and for other image processing tasks.

REFERENCES

- [1] L. I. Rudin, S. Osher *et al.*, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259-268, 1992.
- [2] K. Bredies, K. Kunisch, and T. Pock, "Total Generalized Variation," *SIAM J. Imaging. Sci.*, vol. 3, no. 3, pp. 492-526, 2010.
- [3] X. Zhang, M. Burger *et al.*, "Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction," *SIAM J. Imaging. Sci.*, vol. 3, no. 3, pp. 253-276, 2010.
- [4] H. Schaeffer and S. Osher, "A Low Patch-Rank Interpretation of Texture," *SIAM J. Imaging. Sci.*, vol. 6, no. 1, pp. 226-262, 2013.
- [5] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311-4322, 2006.
- [6] J. F. Abascal, S. Si-Mohamed, P. Douek, C. Chappard, and F. Peyrin, "A sparse and prior based method for 3D image denoising," in *Proc. Eur. Signal Process. Conf. EUSIPCO*, 2019.
- [7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.
- [8] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Probl.*, 2017.
- [9] M. T. McCann, K. H. Jin, and M. Unser, "A Review of Convolutional Neural Networks for Inverse Problems in Imaging," 2017.
- [10] Q. Yang, P. Yan, Y. Zhang *et al.*, "Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, 2018.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation." Springer, Cham, 2015, pp. 234-241.
- [12] "Improving Low-Dose CT Image Using Residual Convolutional Network," *IEEE Access*, vol. 5, pp. 24 698-24 705, 2017.
- [13] A. Mousavi and R. G. Baraniuk, "Learning to invert: Signal recovery via Deep Convolutional Networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 2272-2276.
- [14] K. H. Jin, M. T. McCann, E. Froustey *et al.*, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509-4522, 2017.
- [15] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360-e375, 2017.
- [16] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *Journal of Machine Learning Research*, vol. 12, pp. 2121-2159, 2011.
- [17] M. Abadi, A. Agarwal, P. Barham *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," Tech. Rep., 2016.
- [18] S. Si-Mohamed, D. Bar-Ness *et al.*, "Review of an initial experience with an experimental spectral photon-counting computed tomography system," *Nucl. Instr. Meth. Phys. Res. B*, vol. 873, pp. 27-35, 2017.
- [19] J. Abascal, M. Abella *et al.*, "A novel prior- and motion-based compressed sensing method for small-animal respiratory gated CT," *PLoS One*, vol. 11, no. 3, 2016.
- [20] N. Yuan, J. Zhou, and J. Qi, "Low-dose CT image denoising without high-dose reference images," in *Fully Three-Dimensional Image Reconstr. Radiol. Nucl. Med.*, vol. 11072. SPIE, 2019.
- [21] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3D transform-domain collaborative filtering," *IEEE Trans Image Process*, vol. 16, no. 8, 2007.