2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC) | 978-1-6654-8872-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/NSS/MIC44845.2022.10399172

ESR-Net: An Efficient Image Super-resolution Network for SPECT Reconstruction

Zongyu Li, *Student Member, IEEE*, Yuni K. Dewaraja, *Member, IEEE*
and Jeffrey A. Fessler, *Fellow, IEEE*

***Abstract*—SPECT image reconstruction is challenging due to the poor spatial resolution of the SPECT camera. Incorporating machine learning (ML) based regularizers into reconstruction algorithms is receiving more attention recently. However, most previous works assume the true images used for training have the same voxel size as the reconstructed images, which can be suboptimal when training images with finer voxel sizes are available. Directly using ML-based algorithms with fine voxel sizes could be very computationally expensive due to the heavy computations involved in forward and backward projections. This paper proposes a novel, efficient image super-resolution reconstruction network (ESR-Net) that can improve the resolution by training the ML based regularizer using true activity maps having finer voxel sizes, while maintaining the computational efficiency by computing the forward and backward projections in coarser voxel sizes through downsampling and upsampling. Simulation results based on digital XCAT phantoms demonstrated that the proposed ESR-Net significantly outperformed other methods (OSEM and BCD-Net), when evaluated qualitatively by visualizing reconstructed images and line profiles, as well as quantitatively by mean recovery coefficients (MRC) and normalized root mean square error (NRMSE).**

***Index Terms*—Regularized model-based image reconstruction, image super-resolution, unrolled iterative algorithm, deep learning, Lu-177 SPECT.**

I. Introduction

SPECT is a nuclear medicine technique that images spatial distributions of radioisotopes for clinical diagnosis and to estimate radiation absorbed doses in nuclear medicine therapies. The SPECT reconstruction problem is challenging because of the limited spatial resolution of the collimator.

Traditional SPECT reconstruction methods, e.g., ordered subset expectation maximization (OSEM), suffer from a trade-off between recovery and noise. Regularizers may address that trade-off, but choosing an appropriate regularizer can be challenging. Conventional regularizers such as total variation (TV) and non-local means (NLM) rely on assumed image properties that may not hold in practice. Recently, regularizers learned from deep neural networks (DNN) have received much attention. Chun *et al.* [1] proposed to use matched encoders and decoders (known as BCD-Net) to learn regularizers for low-dose CT reconstruction; Lim *et al.* [2] refined the structure of BCD-Net and applied it to low-count PET reconstruction. Other methods such as FBSEM-Net [3], EM-Net [4], MAPEM-Net [5], have also been proposed recently. Even if originally proposed for CT or PET, these algorithms are applicable for SPECT reconstruction by using an appropriate system model [6]. However, one limitation of these previous methods is that they assume the true images used for training have the same voxel size as the reconstructed images, which can be suboptimal if true images having finer voxel sizes (or higher resolution) are available. Directly applying the algorithms mentioned above to images with fine voxel sizes is conceptually straightforward, but would be very computationally expensive due to the heavy computations in forward and backward projections.

This paper proposes a novel, efficient method that can enhance the resolution of the reconstruction by training the regularizer using training images having finer voxel sizes, while maintaining the computational efficiency by working with coarser voxels for forward and backward projections. We use downsampling and upsampling operators to handle the different voxel sizes. We call the proposed **E**fficient **S**uper-**R**esolution network "ESR-Net".

The rest of this paper is organized as follows: Section II introduces the proposed ESR-Net, digital phantom simulation and evaluation metrics. Section III provides simulation results and compares with previous methods both qualitatively and quantitatively. Section IV concludes this paper and provides future directions.

This work was supported in part by NIH Grants R01 EB022075 and NIH R01 CA240706.

Z. Li and J. A. Fessler are with the Department of Electrical and Computer Engineering, University of Michigan, (e-mails: zonyul@umich.edu, fessler@umich.edu).

Y. K. Dewaraja is with the Department of Radiology, University of Michigan, (e-mail: yuni@med.umich.edu).

978-1-6654-8872-3/22/$31.00 ©2022 IEEE

II. Methods

A. ESR-Net

In regularized model-based SPECT reconstruction of an image $\boldsymbol{x}$, we aim to solve

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\geq 0}{\arg\min}\, f(\boldsymbol{x}) + \beta R(\boldsymbol{x}),$$
$$f(\boldsymbol{x}) \triangleq \boldsymbol{1}' \left(\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}\right) - \boldsymbol{y}' \log(\boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{r}}), \qquad (1)$$

where $\boldsymbol{A}$ denotes the SPECT system model, $\boldsymbol{y}$ denotes noisy measurements that are assumed to follow i.i.d. Poisson distribution. $\bar{\boldsymbol{r}}$ denotes the mean background events such as scatters; $f(\boldsymbol{x})$ is the negative Poisson maximum likelihood (ML) function, $R(\boldsymbol{x})$ is the regularization function. This paper focuses on machine learning based $R(\boldsymbol{x})$.

The key idea of ESR-Net is to let $f(\boldsymbol{x})$ work with coarse voxel sizes whereas $R(\boldsymbol{x})$ works with finer voxels. So we modify the cost function (1) to be

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x}\geq 0}{\arg\min}\, f(\boldsymbol{T}\boldsymbol{x}) + \beta R(\boldsymbol{x}), \qquad (2)$$

where $\boldsymbol{T}$ denotes a 3D downsampling matrix, and $\boldsymbol{x}$ denotes a finely sampled image. Our implementation of the downsampling $\boldsymbol{T}\boldsymbol{x}$ uses average pooling.

To attack (2), one can use unfolded block coordinate descent (BCD) algorithm [2], which leads to the iteration update of the form

$$\boldsymbol{u}_{k+1} = r_{\boldsymbol{\theta}}(\boldsymbol{x}_k), \qquad (3)$$
$$\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x}\geq 0}{\arg\min}\, f(\boldsymbol{T}\boldsymbol{x}) + \frac{\beta}{2}\|\boldsymbol{x} - \boldsymbol{u}_{k+1}\|_2^2$$
$$\approx \tfrac{1}{2\beta}\left(\sqrt{h^2(\boldsymbol{u}_{k+1}) + 4\beta\boldsymbol{x}_k \odot e(\boldsymbol{x}_k)} - h(\boldsymbol{u}_{k+1})\right),$$

where $r_{\boldsymbol{\theta}}$ denotes a neural network with parameter $\boldsymbol{\theta}$; $(\cdot)^2$ and $\odot$ denotes element-wise square and multiplication, respectively. Subscript k denotes the iteration number, and functions $h(\cdot)$ and $e(\cdot)$ are

$$h(\boldsymbol{u}_{k+1}) \triangleq \boldsymbol{T}'\boldsymbol{A}'\boldsymbol{1} - \beta\boldsymbol{u}_{k+1},$$
$$e(\boldsymbol{x}_k) \triangleq \boldsymbol{T}'\boldsymbol{A}'\left(\boldsymbol{y} \oslash \left(\boldsymbol{A}\boldsymbol{T}\boldsymbol{x}_k + \bar{\boldsymbol{r}}\right)\right), \qquad (4)$$

where $\boldsymbol{T}'$ denotes the adjoint of the down-sampling operator $\boldsymbol{T}$, which is an up-sampling operation. We implemented this as an exact adjoint. In (3), we ran one iteration of regularized EM algorithm to approximate the minimizer per each outer iteration. Fig. 1 shows the proposed ESR-Net architecture.

B. Digital Phantom Simulation

We simulated 4 XCAT phantoms (of size 384×384×240, voxel size 1.6×1.6×1.6 mm^3) [7] as the true activity map in our experiment, which covered a diverse range with regard to lesions of different sizes and locations (within and outside the liver). Lesions with non-uniform activity distribution, such as necrotic (cold) sub regions (Fig. 2 (a)) were also included. Out of the 4 XCAT phantoms, we randomly selected two for training, one for validation and one for testing. To be clinically relevant, we assigned the activity distribution of XCAT phantoms to approximately follow the activity distribution of patients imaged with SPECT for the purposes of dosimetry following Lu-177 DOTATATE therapy. Table I shows the activity ratio using liver as reference (the activity of liver is normalized to 1) of XCAT phantoms.

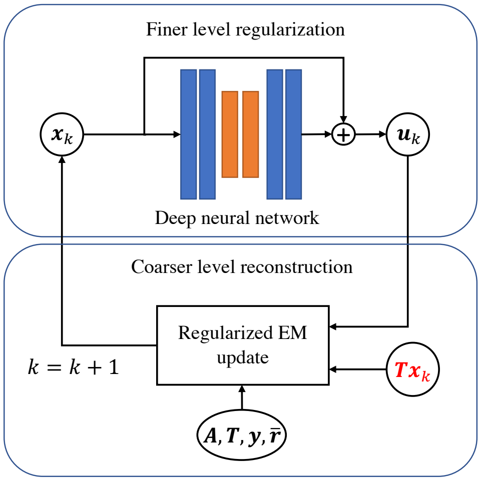


Fig. 1. Architecture of the proposed ESR-Net for SPECT reconstruction.

TABLE I
Activity concentration ratio (compared to liver) of XCAT phantoms.

Phantom/Organ	Lesion	Kidney cortex/medulla	Spleen	Lung
Train 1	10	2/1	1.5	0.1
Train 2	6.2	3.0/3.0	3.7	0.08
Val	3	1/0.25	1.5	0.1
Test	7	1/0.25	1.5	0.1

Next, Lu-177 SPECT projections corresponding to each XCAT phantom's activity/density maps were generated using the SIMIND MC code [8] simulating approximately 1 billion histories per projection. The SIMIND model parameters were based on Lu-177 patient imaging in our clinic (Siemens Intevo with medium energy collimators, a 5/8" crystal, a 20% photopeak window at 208 keV, and two adjacent 10% scatter windows). Poisson noise was simulated after the 128 projection views were scaled to a count-level in the range of 3–20 million total counts, corresponding to the range in post-therapy imaging. SPECT reconstruction used an OSEM algorithm in the Michigan Image Reconstruction Toolbox (MIRT)[1] with CT-based attenuation correction, triple energy window (TEW) scatter correction and collimator-detector response modeling (4 subsets and 16 iterations, $128 \times 128 \times 80$ matrix with voxel size $4.8 \times 4.8 \times 4.8$ mm^3, no Gaussian smoothing).

Since all activity maps must have the same units to be fairly compared, we scaled the true activity map by

$$\boldsymbol{x}_{\text{true}}^{\text{scaled}} \triangleq \frac{\boldsymbol{x}_{\text{true}}}{\sum \boldsymbol{x}_{\text{true}}} \cdot \text{CF} \cdot \text{IF} \cdot \text{CLF}, \tag{5}$$

where $\text{CF} = 15.65$ is the SIMIND calibration factor using a point source; IF stands for “interpolation factor”, which equals to 4.8mm^3/1.6mm^3=27; CLF stands for “count level factor”, which is the ratio between the scaled count level before adding Poisson noise and the total counts from SIMIND.

C. Neural Network

We performed sequential training, i.e., training every outer iteration of the BCD algorithm sequentially with non-shared weights DNNs, for the consideration of memory efficiency. Each DNN is a 3D U-Net [9] with 3 downsample-upsample pairs and 8 filters in the first convolutional layer. After each downsample layer, the number of filters at the next convolutional layer was increased by a factor of two. To potentially simplify the DNN training, we added the DNN input (i.e., $\boldsymbol{x}_k$) to the DNN output, as in the common residual learning approach [10]. We removed all the batch normalization layers in the DNN since we set the batch size to be one. Each DNN was trained for 500 epochs on a Nvidia RTX 3090 GPU by minimizing the mean square error (MSE) using AdamW optimizer [11] with learning rate 0.002. We implemented the DNN in PyTorch.

D. Evaluation Metric

We used mean recovery coefficient (MRC) and normalized root mean square error (NRMSE) as evaluation metrics, where MRC is

$$\text{MRC} \triangleq \frac{\frac{1}{n_p} \sum_{j \in \text{VOI}} \hat{\boldsymbol{x}}[j]}{\frac{1}{n_p} \sum_{j \in \text{VOI}} \boldsymbol{x}_{\text{true}}^{\text{scaled}}[j]} \times 100\%, \tag{6}$$

where n_p denotes number of voxels in the voxels of interest (VOI). NRMSE is defined as

$$\text{NRMSE} \triangleq \frac{\sqrt{\frac{1}{n_p} \sum_{j \in \text{VOI}} \left(\hat{x}[j] - \boldsymbol{x}_{\text{true}}^{\text{scaled}}[j] \right)^2}}{\sqrt{\frac{1}{n_p} \left(\boldsymbol{x}_{\text{true}}^{\text{scaled}}[j] \right)^2}} \times 100\%. \tag{7}$$

E. Compared Methods

We compared our proposed ESR-Net with the conventional unregularized OSEM algorithm and BCD-Net [2]. Both OSEM and BCD-Net work with 4.8mm^3 voxel sizes, whereas ESR-Net works with 1.6mm^3 voxels, so we resized the reconstructed images of OSEM and BCD-Net into 1.6mm^3 voxel size using trilinear interpolation before comparison. We trained BCD-Net and ESR-Net with the same regularization parameter $\beta = 0.1$ and the same DNN architecture; the only difference is that the BCD-Net was trained using activity maps of size $128 \times 128 \times 80$ (GT 128 in Fig. 2), which were downsampled from the original true activity maps. We used validation data to empirically choose the number of outer iterations as 4. The regularized EM algorithm for BCD-Net and ESR-Net was implemented in Julia using the “SPECTrecon” package[2] [6].

III. Result

A. Qualitative Comparison

Fig. 2 shows that the proposed ESR-Net visually improves the reconstruction of a necrotic tumor over the OSEM and the BCD-Net significantly. In particular, the OSEM and BCD-Net barely recovered the cold center; in contrast, the proposed ESR-Net showed much better recovery and was even comparable with the GT 128 regarding the line profile as demonstrated in Fig. 2 (f). Furthermore, Fig. 2 also demonstrates improvement in resolution for spleen.

B. Quantitative Comparison

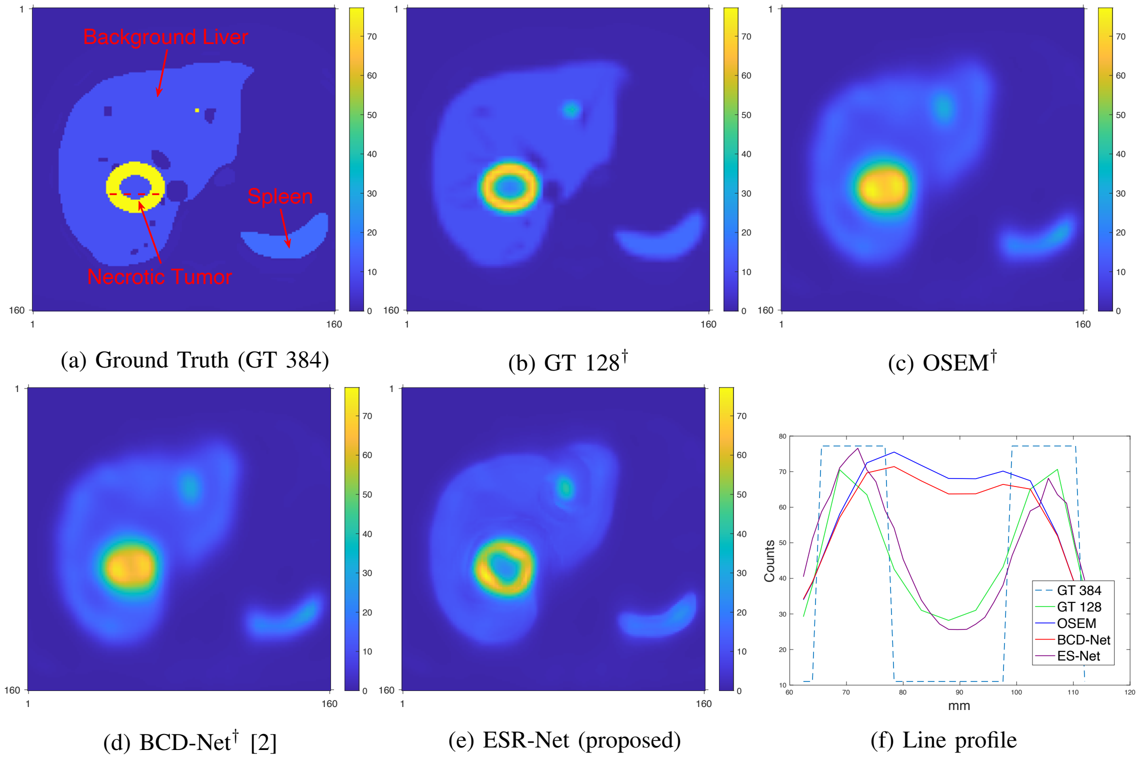
Table II and Table III show that the proposed ESR-Net consistently has the highest MRC over all lesions and the lowest NRMSE over all test organs, compared to OSEM and BCD-Net. The improvement by the ESR-Net for NRMSE can up to 10% (for spleen). On the other hand, the BCD-Net showed comparable performance as the OSEM, which could be due to the suboptimality of the ground truth used in training.

TABLE II
MRC of lesions. *: Lesion 1 has a 19 mL necrotic center.

Lesion/MRC(%)	Vol (mL)	OSEM	BCD-Net	ESR-Net
Lesion 1*	67.5	69.6	68.4	**75.9**
Lesion 2	10.1	92.7	89.4	**97.1**
Lesion 3	9.1	93.2	89.9	**98.1**

[1] Available at https://web.eecs.umich.edu/~fessler/code/index.html.

[2] Available at https://github.com/JuliaImageRecon/SPECTrecon.jl.



(a) Ground Truth (GT 384)

(b) GT 128†

(c) OSEM†

(d) BCD-Net† [2]

(e) ESR-Net (proposed)

(f) Line profile

Fig. 2. Qualitative comparison of different methods on test XCAT phantom, where † denotes after interpolation (image size $128\times128\times80\rightarrow384\times384\times240$ with voxel size $4.8\text{mm}^3 \rightarrow 1.6\text{mm}^3$). Subfigure (f) shows the line profile over a necrotic tumor marked by the dashed line in (a).

TABLE III
NRMSE of different organs. "Lesion*" denotes averaging across all lesions.

Organ/NRMSE(%)	OSEM	BCD-Net	ESR-Net
Lesion*	37.4	36.1	**31.4**
Kidney	53.8	53.2	**47.3**
Liver	47.7	46.9	**41.7**
Spleen	42.2	41.3	**31.8**
Lung	53.3	53.4	**44.6**

We also compared the NRMSE vs. iterations between different algorithms. We first tested the NRMSE of ESR-Net on the validation data. The NRMSE curve (Fig. 3) went down initially and up after 4 iterations, implying that DNN might start overfitting to the training data after 4 iterations. Based on the validation results, we ran 4 iterations for all methods. Fig. 3 also demonstrates that the proposed ESR-Net consistently outperformed OSEM and BCD-Net at any iteration; whereas the BCD-Net was comparable to the unregularized OSEM.

C. Run Time Comparison

Table IV compares the run time of a single forward projection operation (ran in Julia with "SPECTrecon" package using 8 threads of an Intel Core i9-10920X CPU @ 3.50 GHz) for different image sizes. Table IV demonstrates that directly applying algorithms such as BCD-Net to larger image size (finer voxel size) required much more compute time and hence can be impractical. This motivates the idea of ESR-Net that uses finer voxel sizes only in DNN regularization because efficient implementations of matrix multiplications are available on GPU; while working with coarser voxel size in forward and backward projections to save compute time.

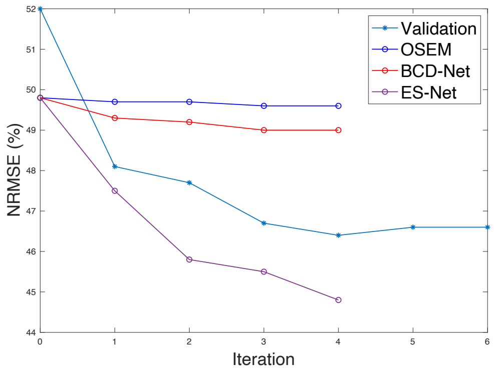


Fig. 3. NRMSE vs. iteration between different algorithms. "Validation" denotes the NRMSE of "ESR-Net" tested on the validation data.

TABLE IV
RUN TIME FOR A SINGLE FORWARD PROJECTION, WHERE "1X" DENOTES $128\times128\times80$; "2X" AND "3X" DENOTE THE IMAGE SIZE INCREASED BY A FACTOR OF 2 AND 3 AT EACH DIMENSION, RESPECTIVELY.

Image size	1x	2x	3x
Run time (s)	4.6	50.6	189.4

IV. Discussion and Conclusion

This paper proposed a novel, efficient image super-resolution network (ESR-Net) for SPECT reconstruction. The proposed ESR-Net is trained with finer voxel size ground truth (GT) but maintains computational efficiency by incorporating downsampling and upsampling during forward and backward projections. Simulation results based on digital XCAT phantoms showed that the proposed ESR-Net outperformed the OSEM and BCD-Net both qualitatively and quantitatively.

We also tested the ESR-Net on virtual patient phantoms and we found ESR-Net shows limited improvement compared to the BCD-Net. This is attributed to the resolution of true activity maps. For virtual patient phantoms, the true activity maps were defined from reconstructed patient images that were degraded by the camera resolution effects and hence limited the performance of ESR-Net. To address this issue, we plan to investigate 3D image deblurring algorithms for patient images to provide better training data in the future [12]. Future work also includes training and testing on an expanded dataset including patients, and generalizing ESR-Net for other radionuclides such as Y-90.

References

[1] Y. Chun and J. A. Fessler. "Deep BCD-Net Using Identical Encoding-Decoding CNN Structures for Iterative Image Recovery". In: *2018 IEEE IVMSP*. 2018, pp. 1–5. DOI: 10.1109/IVMSPW.2018.8448694.

[2] H. Lim, I. Y. Chun, Y. K. Dewaraja, and J. A. Fessler. "Improved Low-Count Quantitative PET Reconstruction With an Iterative Neural Network". In: *IEEE Trans. Med. Imag.* 39.11 (2020), pp. 3512–3522. DOI: 10.1109/TMI.2020.2998480.

[3] A. Mehranian and A. J. Reader. "Model-Based Deep Learning PET Image Reconstruction Using Forward–Backward Splitting Expectation–Maximization". In: *IEEE Trans. Rad. Plas. Med. Sci.* 5.1 (2021), pp. 54–64. DOI: 10.1109/TRPMS.2020.3004408.

[4] K. Gong, D. Wu, K. Kim, J. Yang, G. E. Fakhri, Y. Seo, and Q. Li. "EMnet: an unrolled deep neural network for PET image reconstruction". In: *Med. Imag. 2019*. Vol. 10948. SPIE, 2019, pp. 1203–1208. DOI: 10.1117/12.2513096. URL: https://doi.org/10.1117/12.2513096.

[5] K. Gong, D. Wu, K. Kim, J. Yang, T. Sun, G. E. Fakhri, Y. Seo, and Q. Li. "MAPEM-Net: an unrolled neural network for Fully 3D PET image reconstruction". In: *Fully3D*. Vol. 11072. SPIE, 2019, pp. 109–113. DOI: 10.1117/12.2534904. URL: https://doi.org/10.1117/12.2534904.

[6] Z. Li, Y. K. Dewaraja, and J. A. Fessler. "Training End-to-End Unrolled Iterative Neural Networks for SPECT Image Reconstruction". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* (2023), pp. 1–1. DOI: 10.1109/TRPMS.2023.3240934.

[7] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui. "4D XCAT phantom for multimodality imaging research". In: *Med. Phys.* 37.9 (2010), pp. 4902–4915.

[8] M. Ljungberg. *The SIMIND Monte Carlo program*. 2012. DOI: https://doi.org/10.1201/b13073-8.

[9] O. Ronneberger, P. Fischer, and T. Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *MICCAI*. 2015, pp. 234–241.

[10] K. He, X. Zhang, S. Ren, and J. Sun. "Deep Residual Learning for Image Recognition". In: IEEE, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[11] I. Loshchilov and F. Hutter. "Decoupled Weight Decay Regularization". In: *ICLR*. 2019.

[12] F. Wen, R. Ying, Y. Liu, P. Liu, and T.-K. Truong. "A Simple Local Minimal Intensity Prior and an Improved Algorithm for Blind Image Deblurring". In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.8 (2021), pp. 2923–2937. DOI: 10.1109/TCSVT.2020.3034137.