A lightweight convolutional neural network for image denoising with fine details preservation capability

3 authors, including:

Sutanu Bera
Indian Institute of Technology Kharagpur
4 PUBLICATIONS   **1** CITATION

A LIGHTWEIGHT CONVOLUTIONAL NEURAL NETWORK FOR IMAGE DENOISING WITH FINE DETAILS PRESERVATION CAPABILITY

Sutanu Bera, Avisek Lahiri, Prabir Kumar Biswas

Department of Electronics and Electrical Communication Engineering,
Indian Institute of Technology Kharagpur

ABSTRACT

Image denoising is a fundamental problem in image processing whose primary objective is to remove the noise while preserving the original image structure. In this work, we proposed a new architecture for image denoising. We have used several dense blocks to design our network. Additionally, we have forwarded feature extracted in the first layer to the input of every transition layer. Our experimental result suggests that the use of low-level feature helps in reconstructing better texture. Furthermore, we had trained our network with a combination of MSE and a differentiable multi-scale structural similarity index(MS-SSIM). With proper training, our proposed model with a much lower parameter can outperform other models which were with trained much higher parameters. We evaluated our algorithm on two grayscale benchmark dataset BSD68 and SET12. Our model had achieved similar PSNR with the current state of the art methods and most of the time better SSIM than other algorithms.

Index Terms— Image Denoising, MS-SSIM, Skip Connections, Low-level feature, Texture Preservation, CNN

1. INTRODUCTION

Image denoising is a fundamental image reconstruction problem, but still an active topic for low-level vision researchers. The main objective of image denoising is to recover the clean latent image f from the noise corrupted version g, which follows the image degradation model $g = f + \eta$, where η is the additive noise. It has been shown that a efficient denoising algorithm can solve many other image reconstruction problems such as super resolution, deblurring, inpainting, compression etc.[1]

Initially, learning the image prior was considered as the effective way of denoising the image. In particular, models like non-local similarity (NSS) models[2], Markov random field (MRF) models[3], Sparse models[4], etc. was used for denoising.

However, all the previous models require complex iterative computation, thus becomes time-consuming in the testing stage. On the contrary, for a real-time application, fast algorithms are needed. As a solution , discriminative models[5]
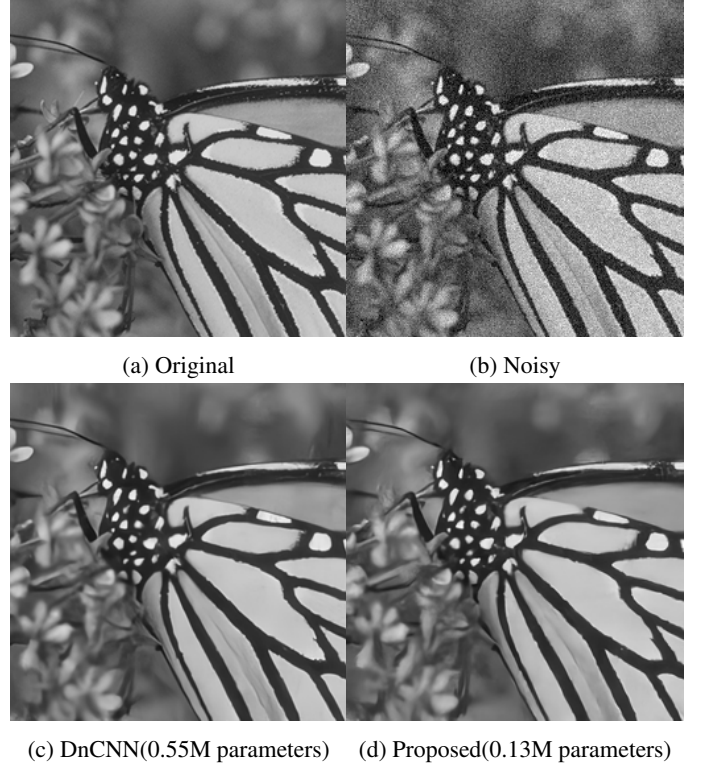

(a) Original   (b) Noisy

(c) DnCNN(0.55M parameters)   (d) Proposed(0.13M parameters)

Fig. 1: Denoising Result on noise level $\sigma = 25$, (a) Original, (b) Noisy (PSNR:24dB/SSIM:0.60) (c)DnCNN (PSNR:30.38dB/SSIM:0.91) (d) Proposed (PSNR:30.29dB/SSIM:0.91). Proposed model with 76% lesser parameter than current state of art[10] yielded visually similar image.

attracted attention because these models eliminated the iterative steps required at the testing stage.

In recent times, as a discriminative model, deep convolutional network[6],[7],[8],[9] has started gaining considerable attention because of its fast and efficient denoising capability. In this regard Kai Zhang et al.[10] proposed DnCNN network, they first utilized batch normalization and residual learning for denoising task. Their model otperformed many of benchmark model such as; BM3D[11], TNRD[12], etc.

Indeed, Deep CNN based models provide better performance than other models. However, these models comprise of a good deal of matrix multiplication. Therefore, good com-
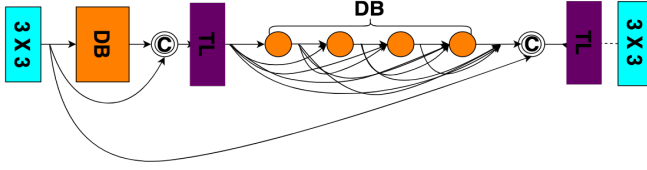
Fig. 2: Proposed Architecture DB: Dense Block, TL: Transition Layer

puting resources are required for implementation. Also, attempt to reduce the complexity of network results in a reduction in performance. In this study, we tried to make a tradeoff between the model complexity and performance.

Meanwhile, Mean Square Error (MSE) or L2 norm of the difference image remains the most preferred error measure in the researcher community; it may be because L2 norm is convex, differentiable, easier to optimize. Additionally, it also provides the maximum likelihood (ML) estimate solution. Still, as an error measure, the L2 norm ignores the local characteristic of the image like, contrast, luminance, structural information, etc., which are an integral part of the human visual system. In this study, we also tried to incorporate this measures into loss function to train our network.

Contribution: The main contributions of this study are as follows:

- We present a new CNN architecture for image denoising with lesser complexity than the benchmark architectures. We also propose an optional addition to the above mentioned architecture to further reduce the complexity and execution time. Our model with lesser number of parameter produced similar result like more complex deep CNN models.

- We forwarded the feature extracted in the first layer to next layer through skip connections. Our experimental result shows that, these low-level features contain complementary information that high level feature lacks for texture generation. Our solution reconstructs better texture than baseline solution.

- We also showed the effect of inclusion of local characteristic into the cost function by using MS-SSIM. We also propose specific sequence of multi stage training to generate optimal solution.

Remaining part of the paper is organized in the following way. In Section 2 the proposed architecture is introduced, then in Section 3 the training procedure and the cost function is explained, and in Section 4 the evaluation results are represented, and finally, Section 5 provides the conclusion about the findings of this study.

2. PROPOSED ARCHITECTURE

It is known that the initial layer of deep neural network learns low-level feature like edge information, corner points, etc, whereas the deeper layer learns more complex feature like face orientation, larger shapes, etc. Perhaps, these high-level features are useful in classification task, but for image restoration these features might become less significant. Indeed, the low-level features plays an important role in image restoration. With this foundation, we decided to propagate low-level features through the network and let the network infer the importance of these features for reconstruction. Figure 2 illustrates the proposed network. First layer is a convolution layer with 64 filters of 3×3 kernel size. Next, a series of dense block (DB)and transition layer (TL) is used to extract features, and then finally a convolution layer with 3×3 kernel to reconstruct the image. We used total 6 pairs of dense block and transition layer in our network. Inside the DB, the feature-maps of all preceding convolution layers are used as the input to current layer, and its own feature-maps are used as inputs into all subsequent layers. The growth rate of every DB was set to 16, and four convolution layer followed by batch normalization and ReLU non-linearity was used in each DB. After every dense block, a transition layer consisting 1×1 convolution layer followed by batch normalization and ReLU, had been used to reduce the depth of the feature map and, also to combine the feature extracted at different layer. According to the NIN paper[13], 1×1 convolution is similar to cross-channel parametric pooling. This cascaded cross channel parametric pooling structure allows complex and learnable interactions of cross channel information. The input of the transition layer is a mixture of the low-level and high-level feature over the volume. The 1×1 convolution may drop the low-level feature or can propagate them to the next layer.

Adapting dense net[14] architecture helped in reusing the feature map. As a result, the total number of parameter reduced significantly. Furthermore, a better receptive field was obtained by using more number of layers with fewer filters. Additionally, the use of skip connection also makes the error surface more smooth and convex[15]. In addition to this, traditional convolution can also be replaced with depthwise separable convolution[16] to further reduce the number of parameters drastically.

3. TRAINING DETAILS

Use of residual learning in image reconstruction is already established[10]. According [20] to when the original mapping is more similar to identity mapping then it is more easy to learn and optimize the residual mapping. Then clean image can restored by subtracting the residual map(i.e. the noise map) from the noisy image.

Now, one obvious choice for error measure is L2 norm of dif-

Dataset	Noise Level	Method								
		BM3D[11]	WNNM[17]	EPLL[18]	MLP[19]	CSF[5]	TNRD[12]	DnCNN[10]	FFDNet[7]	Proposed v1
BSD68	15	31.07	31.37	-	31.24	31.42	31.73	31.75	31.63	31.70
	25	28.57	28.83	28.68	28.96	28.74	28.92	29.23	29.19	29.20
	50	25.62	25.87	25.67	26.03	-	25.97	26.23	26.29	26.25
SET12	15	32.37	32.69	32.13	-	32.31	32.50	32.88	32.75	32.82
	25	29.96	30.25	29.69	30.02	29.83	30.05	30.45	30.43	30.41
	50	26.72	27.05	26.47	26.78	-	26.81	27.23	27.32	27.17

Table 1: The average PSNR(dB) results of different methods, top 2 results are marked in red and green colour respectively.

(a) Original   (b) Noisy   (c) Using Low-level feature   (d) Without low-level feature
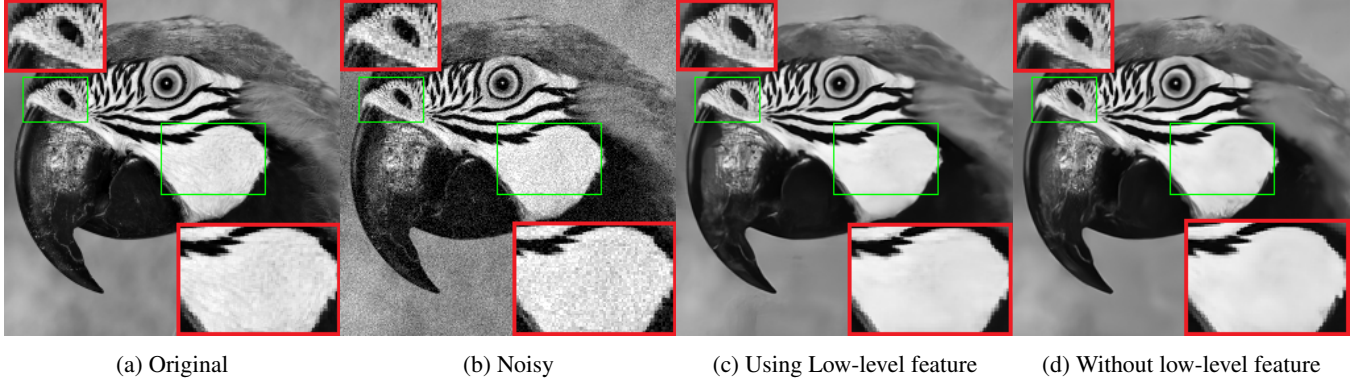
Fig. 3: Denoising Result of Parrot image on noise level $\sigma = 25$. (The reader is encouraged to zoom in for a better view). Reconstructed image using low-level feature has better texture than the other solution. (a) Original, (b) Noisy, (c) Proposed v1 (d) We removed the skip connection before all the transition layer so that low-level feature does not propagate through the network.

ference image or the Mean Square Error. However, L2 does not consider the local characteristic of the image. To overcame this drawback, we first trained the entire network using MSE. Then, retrained only the last layer of the network with the combination L2 norm and differentiable Multi Scale Structural Similarity Index (MS-SSIM) as a multi objective optimization problem keeping other layer untouched. Why we only trained the last layer only with MS-SSIM is discussed in the Section 4.

The cost function for training the final layer of the image is given by

$$l(\Theta) = (1 - MS - SSIM(x, (y - R(y; \Theta))))$$

$$+ \frac{1}{2N} \sum_{i=1}^{N} ||R(y_i; \Theta) - (y_i - x_i)||_F^2$$

The differentiable implementation of MS-SSIM can be found in the study by Hang Zhao *et al.*[21].

We trained our network for three specific noise level, particularly for $\sigma = 15, 25, 50$. For validation two benchmark dataset, Berkeley segmentation data set (BSD68) containing 68 natural images, and famous 12 images of SET12, is used.

		METHOD			
Dataset	Noise Level	BM3D	TNRD	DnCNN	Proposed v1
BSD 68	σ=15	0.8741	0.8947	0.8947	**0.8949**
	σ=25	0.8025	0.8206	0.8321	**0.8326**
	σ=50	0.6744	0.7104	0.7252	**0.7289**
SET 12	σ=15	0.8989	0.9004	0.9073	**0.9099**
	σ=25	0.8553	0.8573	0.8675	**0.8678**
	σ=50	0.7679	0.7753	**0.7913**	**0.7913**

Table 2: Comparision of SSIM with other state of art models.

4. RESULT AND DISCUSSION

We evaluated both of our model namely; proposed v1: as described above section, and proposed v2: replacing the convolution layer by depthwise separable convolution. Table 1 depicts the comparison of the proposed v1 network with other method applied in this two dataset. It can be seen that proposed method has outperformed many benchmark algorithm like BM3D, TNRD, WNMM, CSF and also yielded comparable PNSR with DnCNN algorithm.

Figure 3 gives an example to show the effect of adding low-level feature to the input of every transition layer. Reconstructing the texture or the fine details is one of the most challenging tasks for image denoising models. Indeed, most of the algorithm produces a smoothed version of the image. However, it can be seen that adding low-level feature provides better texture. In Figure 3 the fine texture near the nose
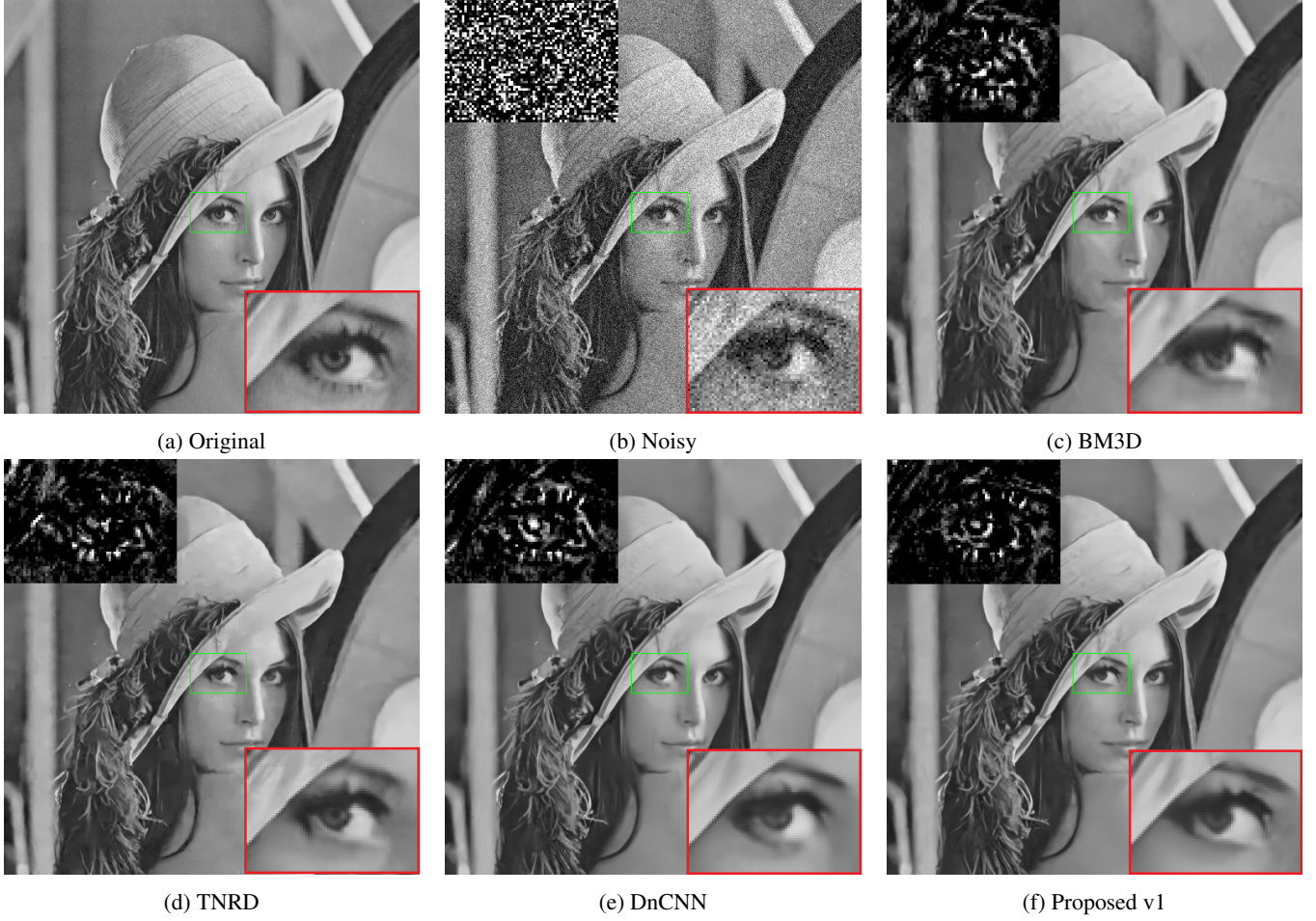
(a) Original   (b) Noisy   (c) BM3D

(d) TNRD   (e) DnCNN   (f) Proposed v1

Fig. 4: Denoising result on lena image with noise level 25 by different method. Proposed method performed better in preserving fine details while removing noise. The lower inset is the zoomed version of marked area, and the upper inset is the difference image of same region.

Dataset	Noise Level	DnCNN	Proposed v1	Proposed v2
BSD68	15	31.75	31.70	31.62
	25	29.23	29.20	29.08
	50	26.23	26.25	26.12
SET12	15	32.88	32.82	32.69
	25	30.45	30.41	30.32
	50	27.23	27.17	27.09
Parameter			556032	382080	133248

Table 3: Comparison of Parameter and PSNR(dB)

and also in the area near chin is noticable. Table 2 reports comparative SSIM of our model with other three benchmark algorithm. In most of the times, We had scored best SSIM. Retraining last layer with MS-SSIM helped in achieving this. Furthermore, we also witnessed a small improvement in PSNR after training with MS-SSIM. We also tried to retrain all the layer with MS-SSIM and, but doing so resulted in similar SSIM but lower PSNR images.

Table 3 shows comparison between the number of trainable parameter and PSNR. Model trained with depthwise sepa-rable convolution has nearly 76% less parameter than the DnCNN model, but has achieved similar PSNR, also our model with normal convolution has 30% less parameter than DnCNN.

The visual comparisons of different methods are given in Figure 4. Proposed model has kept fine details better than the other model. The shape of the reconstructed eye is better our model, it can be also verified by examining the difference image. Overall, proposed solution provides good perceptual quality image.

5. CONCLUSION

In this paper a new CNN network for image denoising is pro-posed. This method has less complexity than the state of the art method, but still resultant images are similar with better texture. This light weight model can be utilised in low re-sources devices, such as smart phones, to perform state of art denoising.

6. REFERENCES

[1] Yaniv Romano, Michael Elad, and Peyman Milanfar, "The little engine that could: Regularization by denoising (RED)," *CoRR*, vol. abs/1611.02862, 2016.

[2] Antoni Buades, Bartomeu Coll, and J-M Morel, "A non-local algorithm for image denoising," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 2, pp. 60–65.

[3] Stan Z Li, *Markov random field modeling in image analysis*, Springer Science & Business Media, 2009.

[4] Michael Elad and Michal Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[5] Uwe Schmidt and Stefan Roth, "Shrinkage fields for effective image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2774–2781.

[6] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," *arXiv preprint arXiv:1606.08921*, 2016.

[7] Kai Zhang, Wangmeng Zuo, and Lei Zhang, "Ffdnet: Toward a fast and flexible solution for cnn based image denoising," *IEEE Transactions on Image Processing*, 2018.

[8] Tal Remez, Or Litany, Raja Giryes, and Alex M Bronstein, "Deep class-aware image denoising," in *Sampling Theory and Applications (SampTA), 2017 International Conference on*. IEEE, 2017, pp. 138–142.

[9] B. V. Somasundaran, R. Soundararajan, and S. Biswas, "Image denoising for image retrieval by cascading a deep quality assessment network," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 525–529.

[10] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[11] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising with block-matching and 3d filtering," in *Image Processing: Algorithms and Systems, Neural Networks, and Machine Learning*. International Society for Optics and Photonics, 2006, vol. 6064, p. 606414.

[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] Min Lin, Qiang Chen, and Shuicheng Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[14] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[15] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein, "Visualizing the loss landscape of neural nets," in *Advances in Neural Information Processing Systems*, 2018, pp. 6391–6401.

[16] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[17] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2862–2869.

[18] Daniel Zoran and Yair Weiss, "From learning models of natural image patches to whole image restoration," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 479–486.

[19] Harold C Burger, Christian J Schuler, and Stefan Harmeling, "Image denoising: Can plain neural networks compete with bm3d?," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2392–2399.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[21] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.