CSIT Conference 2021, Yerevan, Armenia, September 27 - October 1

Single Image Joint Motion Deblurring and Super-Resolution

Mekhak Shoyan
Yerevan State University,
Yerevan, Armenia
e-mail:
mexakshoyan@gmail.com

Misak Shoyan
National Polytechnic University
of Armenia, Yerevan, Armenia
e-mail:
misakshoyan@gmail.com

Shant Navasardyan
Picsart AI Research (PAIR),
Yerevan, Armenia
e-mail:
shant.navasardyan@picsart.com

Robert Hakobyan
National Polytechnic University
of Armenia, Yerevan, Armenia
e-mail:
rob.hakobyan@gmail.com

Abstract - **In the past decade, deep convolutional neural networks achieved great success in the task of single image super-resolution providing realistic texture and structure details. However, sometimes low-resolution images along with a lack of spatial information are degraded by complicated blur effects. Especially when images are taken in a moving environment or with camera shakes, one may end up with results suffering from motion blur. In this work, the joint problem of single image motion deblurring and super-resolution is addressed. We propose a single branch fully convolutional neural network to restore high-resolution sharp images from given motion blurry low-resolution images without estimating or making any assumptions on the blur uniformity, its spatially varying filter, and noise. Our end-to-end solution reuses the features extracted from the motion deblurring network in the super-resolution module resulting in an efficient model with high performance. Experiments show that our method outperforms the existing state-of-the-art solutions qualitatively and quantitatively by reaching 28.265 peak signal-to-noise ratio. Our implementation code is available at https://github.com/Mekhak/motion-deblur-sr.**

Keywords - **Motion deblurring, super-resolution, deep learning.**

I. INTRODUCTION

Super-resolution (SR) refers to the task of generating a high-resolution (HR) image with enhanced spatial and context information given the corresponding low-resolution (LR) image. The advance of SR techniques is beneficial for a wide range of applications such as surveillance, object detection, object tracking, face detection, etc. Image SR is an ill-posed problem because one low-resolution image can yield several possible high-resolution images. The classical approach of upscaling images is using interpolation techniques such as a nearest neighbor, bilinear, or bicubic interpolations [27, p. 77-78], new edge-directed interpolation [1], etc. Some learning-based methods propose to solve an optimization problem for each low-resolution input [2,3]. Later, deep-learning-based methods have demonstrated superiority over the conventional methods by recovering more texture details and resulting in sharper HR images [4-6].

Since image degradation is inevitable during the image capturing process, the LR images might also be degraded by various kinds of blur effects. This causes blurry HR results after upscaling LR images with a super-resolution technique. Real word images are often degraded by non-uniform motion blur due to camera or scene motion during the exposure. So a need arises to incorporate motion deblurring techniques. The problem of motion deblurring is also highly ill-posed since various deblurred images could correspond to a given motion blurry image.

Conventional deblurring methods assume the existence of the motion blur kernel and employ a variety of constraints or regularizations to estimate that kernel resulting in expensive optimization problems [7,8]. However, designing such regularizations is a challenging task and often not generalizable. Recently, convolutional neural networks (CNN) have achieved better generalization and tremendous performance over the conventional methods [9-13] in deblurring problems.

So, to get visually plausible results for motion blurry image super-resolution, one needs to neutralize the motion blur effect along with recovering the spatial information. In this work, we address the problem of recovering the HR sharp image from a given LR motion blurry image.

The existing SR algorithms [4-6] fail to handle the motion blur effects well (see Fig. 1 (c)). So it is natural to deal with the motion deblurring separately. The naive approach is to deblur the LR image with the methods described in [9, 11], then apply state-of-the-art SR techniques (e.g., HAN [6]) or vice-versa. However, this approach has several drawbacks: first, it is sub-optimal, since the feature extraction and image reconstruction phases will be performed twice in both motion deblurring and SR models, which is memory and time-consuming. Experiments show that the features extracted by image deblurring models encode similar information as the features extracted by the SR models. Hence, it will be reasonable and effective to reuse extracted features from the first model in the second one. Second, solving the SR and motion deblurring problems separately may result in error accumulation since the error estimated by the first model will be propagated through the second model generating unwanted artifacts or edge distortion (see Fig. 1).

Several recent works [14,15] jointly solve the image motion deblurring and SR problem. Zhang et al [14] propose a deblurring branch in a super-resolution network to extract the features for deblurring. However, they focus on uniform Gaussian blur degradations only. In [15] authors propose a dual branch architecture to simultaneously generate LR deblurred and sharp HR images. However, they use dual branch architecture extracting independent features for both deblurring and SR tasks, which is sub-optimal.

To tackle these problems, in this work, we propose a single branch network architecture to solve the complex joint problem by reusing the deblurring features extracted by the network proposed by Zamir et al (MPRNet) [9] to reconstruct the HR sharp image. We propose to refine the hierarchical features extracted from the stages of MPRNet with channel attention blocks [16] and passed them to the image reconstruction module. In summary, our contributions are the following: 1) We propose an end-to-end single-branch pipeline for the joint motion deblurring and image super-resolution task. 2) Our method propagates the image features extracted from the deblurring subnetwork to the subnetwork responsible for super-resolution.
3) Experiments show that our method outperforms the existing state-of-the-art solutions both quantitatively and qualitatively (see Section 4).

(a) Ground-truth HR (b) Blurry LR input (c) HAN (d) MPRNet + HAN (e) DMPHN + HAN (f) HAN + DMPHN (g) GFN (h) Ours

Figure 1. Demonstration of the results of HAN[6] SR algorithm (c) vs combinations of deblurring (MPRNet[9], DMPHN[11]) and SR algorithm (HAN[6]) ((d), (e), (f)) vs existing joint super-resolution and deblurring method GFN[15] (g) vs ours (h). Best viewed zoomed-in on screen.

II. Related Work

As we have already mentioned in Section 1, in this work we focus on the joint problem of single image motion deblurring and single image super-resolution. However, in computer vision, both the sub-tasks (motion deblurring and SR) are very popular. Hence, we start this section with a brief literature review on each of them.

A. Motion Deblurring The non-uniform blind (when the spatially varying blur kernel is unknown) deblurring problem for dynamic scenes is a challenging problem due to its ill-posedness. Conventional blind image motion deblurring methods [7,8] fail to remove non-uniform complex motion blur effects due to some non-generalizable assumptions on the blur kernel model. They are based on a variational model, the key component of which is the regularization term. The restoration quality depends on the prior regularization and its weights. Also, blur kernel estimation with traditional methods is computationally expensive and leads to a long processing time, thus making these methods non-applicable in real-world applications.

Recent works [9-13] are based on deep CNNs, which directly predict the restored sharp image given the blurry input without blur kernel estimation. Zhang et al [11] propose an end-to-end hierarchical model that performs deblurring in a fine-to-coarse manner. They use hierarchical multi-patch architecture to exploit both local and global blur information. Due to the multi-patch setup, each finer level of the network is connected to the corresponding coarser level in a residual manner, thus allowing the network to focus on different scales of the blur.

Zamir et al [9] proposed a multi-stage architecture that progressively learns restoration functions for the degraded inputs. Their proposed network exploits encoder-decoder architecture and single-scale pipeline to preserve both multi-scale contextual information and spatial details. They propose a supervised attention module (SAM) [9] to plug between every two stages to enable progressive learning. In the last state, the original-resolution subnetwork (ORSNet [9]) was used, which does not contain any downsampling operation and fully exploits the benefits of channel attention blocks (CAB [16]) to generate spatially enriched features. Residual connections [17] are added in each stage to force learning the input image modifications for each state. Their proposed network achieves state-of-the-art results not only in motion deblurring but also in image deraining and image denoising problems.

B. Super-resolution Single image super-resolution is also an ill-posed computer vision problem. Deep learning-based methods [4-6] achieve significant performance improvements among the conventional methods [1-3]. Ledig et al [4] proposed a generative adversarial network (GAN) for image super-resolution. In the proposed GAN, they replaced the MSE loss with a loss calculated on the feature maps of the VGG network [18]. They also propose a 16 blocks deep ResNet [17] (SRResNet) optimized with MSE achieving state-of-the-art results. Lim et al [5] further improve the performance of SRResNet by analyzing and removing unnecessary modules to simplify the network architecture. They also proposed a multi-scale system that reconstructs high-resolution images of different upscale factors in a single model.

Niu et al [6] propose a novel super-resolution algorithm introducing a layer attention module (LAM) [6] that captures the correlations and dependencies among different layers of the network.

However, the discussed methods aim to super-resolve images without taking into account the possible blur degradations such as motion blur. As experiments show, these methods amplify the motion blur effect if the input image is degraded by it. In comparison with them, our method successfully upscales input images while removing the motion blur effect (see Fig. 1, as well as Section 4).

C. Joint motion deblurring and super-resolution The joint problem of single image motion deblurring and super-resolution is more challenging than the individual problems, since the LR image is also degraded by motion blur. Xu et al [19] propose a GAN to learn a category-specific prior to solve the problem. However, they focus on text and face images only, which makes their method not applicable in the case of real-world images. Zhang et al [14] propose a dual branch architecture, where the deblurring branch shares the same features extracted from the super-resolution branch. However, they consider only image degradations caused by Gaussian blur, which makes their method non-applicable in real-world motion blurry images.

In [15], authors proposed a dual branch architecture, where the restoration branch performs motion deblurring while the base feature extraction branch extracts features for SR. The features extracted by the restoration branch contain local and global information of the input image, while the deblurring branch features focus on degraded regions and capture the blur information. The extracted features are then fused via the recursive gate module and then fed to the reconstruction module, which enlarges the spatial resolution of the input image 4 times. However, their method fails to recover enough sharpness for some cases (see Fig. 1, Fig. 4).

To address the limitations of the solutions mentioned above, we propose a single branch network architecture to solve the joint problem of single image motion deblurring and SR. Our network extracts contextually and spatially enriched features that are informative for both motion deblurring and SR subtasks. We reuse the features extracted by MPRNet [9] in the SR problem before refining them by CAB [16] modules. Experiments show that our method outperforms the existing state-of-the-art methods quantitatively (see Table 1) and qualitatively (see Fig. 4).

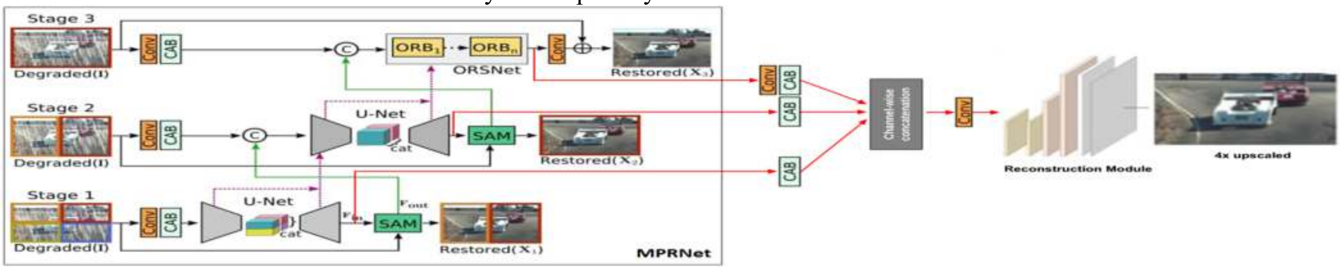


Figure 2. The architecture of the proposed network. Best viewed zoomed-in on screen.

III. Method

We propose a single-branch network architecture to solve the complex joint problem of single image motion deblurring and super-resolution. We suggest to reuse the motion deblurring features extracted by the hierarchical layers of the MPRNet [9] for image super-resolution. We propose a reconstruction module [15] that enlarges the spatial resolution 4 times. Channel attention blocks (CAB) [16] are used to refine the MPRNet-generated features before feeding them to the reconstruction module.

A. Network architecture The architecture of the proposed network is presented in Fig. 2. As mentioned in Section 2, the MPRNet achieves state-of-the-art results on image restoration tasks generating fine texture and structure details in resulting images. Hence, it extracts informative features for image restoration tasks in its deep layers. On the other hand, the SR problem can also be considered as a problem of recovering image details (after the image is upscaled or resampled), mainly the textural and structural details, so the features extracted by MPRNet might be useful for the SR problem.

Since we use the MPRNet [9] as the first part of our whole model, we start this section with a brief introduction to it. The MPRNet consists of three stages (see Fig. 2). The first two stages are based on encoder-decoder architecture to capture contextually informative features. The encoder-decoder subnetworks are based on standard U-Net [20] architecture. Both the encoder and decoder networks use 2 channel attention blocks (CAB) at each scale. The CAB module aims to exploit the dependencies among the channels of its input. It aims to refine the input feature map by suppressing less informative features and amplifying more informative ones. First, the channel-wise statistics (the mean pixel values for each channel) of the input feature map are obtained using global average pooling. Then, two 1x1 convolutions are performed followed by ReLU and sigmoid activations, correspondingly. Then, the obtained channel attention map is multiplicated (element-wise) with the input feature map along the channel dimension to obtain the refined feature map (see Fig. 3).

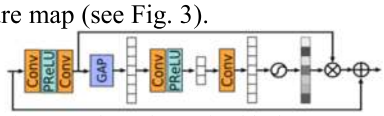


Figure 3. Channel attention block (CAB [9]).

In the last stage, the original-resolution subnetwork (ORSNet) [9] is used, which does not employ any downsampling operation to generate spatially enriched features.

We refined the features extracted by MPRNet with CAB blocks to suppress low informative features for SR and amplify the more informative ones. We use three CABs to refine features generated by all three stages of MPRNet. The refined features are then concatenated channel-wise followed by 3x3 convolution before feeding to the reconstruction module (see Fig. 2).

The reconstruction module performs 4x upscaling given refined MPRNet-generated features. It consists of 8 residual blocks [5] and two pixel-shuffling layers [21]. Each residual block consists of two 3x3 convolutions. The first convolution layer is followed by ReLU activation.

B. Loss functions Each stage of the MPRNet generates an LR deblurred image, thus MPRNet outputs three LR deblurred images ($\hat{L}_1$, $\hat{L}_2$, $\hat{L}_3$), while the reconstruction module outputs the desired 4x upscaled sharp image H. We train the proposed network with the following loss function:

$$\mathcal{L} = \mathcal{L}_{\mathrm{SR}} + \alpha \mathcal{L}_{\mathrm{DB}}$$

where $\alpha = 0.5$, $\mathcal{L}_{\mathrm{DB}}$ and $\mathcal{L}_{\mathrm{SR}}$ denote deblur and super-resolution losses, respectively and are defined as follows:

$$\mathcal{L}_{\mathrm{DB}} = \sum_{s=1}^{3} \mathcal{L}_{\mathrm{char}}(\hat{\mathrm{L}}_{\mathrm{s}}, \mathrm{L}) + \mathcal{L}_{\mathrm{edge}}(\hat{\mathrm{L}}_{\mathrm{s}}, \mathrm{L})$$

$$\mathcal{L}_{\mathrm{SR}} = \mathrm{MSE}(\mathrm{H}, \hat{\mathrm{H}})$$

Here $\hat{\mathrm{H}}$ denotes the reconstructed high-resolution image, L is the ground truth low-resolution image. $\mathcal{L}_{\mathrm{char}}$ is Charbonnier loss [22] and $\mathcal{L}_{\mathrm{edge}}$ is the edge loss [9] (Δ denotes the Laplacian operator):

$$\mathcal{L}_{\mathrm{char}} = \sqrt{||\hat{\mathrm{L}}_{\mathrm{s}} - \mathrm{L}||^2 + \varepsilon^2}$$

$$\mathcal{L}_{\mathrm{edge}} = \sqrt{||\Delta\hat{\mathrm{L}}_{\mathrm{s}} - \Delta\mathrm{L}||^2 + \varepsilon^2}$$

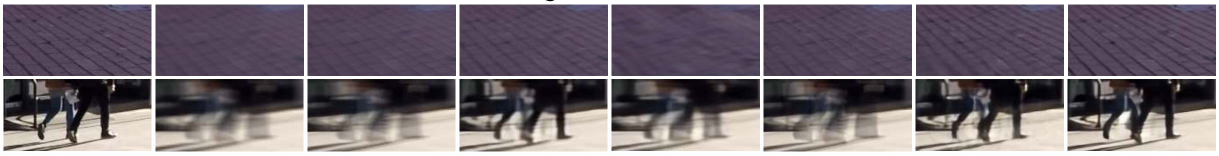
where ε is set to 10^{-3} empirically for all experiments.

C. Implementation details We use the GOPRO dataset [10] to generate training data for the joint image deblurring and SR problem. The GOPRO dataset consists of 2103 blurry and sharp high-resolution image pairs. We crop high-resolution blurry and sharp image pairs into 256x256 patches with stride 128 to generate H_{blur} and H images from the original blurry and sharp images, respectively. Then we get $[L_{blur}, L, H]$ image triplets by bicubically downsampling the H_{blur} and H images 4 times to get L_{blur} and L, respectively.

We trained our network for 21 epochs. To exploit the benefits of transfer learning, we initialized the weight of the MPRNet subnetwork with the GOPRO pre-trained weights. We used Adam optimizer [23] with the initial learning rate 1e-4. We trained our network on a single NVIDIA Geforce 1080 GPU.

IV. RESULTS

We evaluate the performance of the proposed network on the GOPRO test dataset. We evaluate the results using the PSNR [24], SSIM [25], and LPIPS [26] evaluation metrics. We demonstrate the benefits of jointly solving image SR and motion deblurring problems using a single model over running separate SR (EDSR [5], HAN [6]) and motion deblur (MPRNet [9], DMPHN [11]) models sequentially. Our model also achieves significant improvements over the state-of-the-art GFN [15] network, which solves the joint problem of single image motion deblurring and super-resolution using the dual-branch network architecture. The quantitative comparison results on the GOPRO test dataset are shown in Table 1.



(a) Ground-truth HR (b) Blurry LR input (c) HAN (d) MPRNet + HAN (e) DMPHN + HAN (f) HAN + DMPHN (g) GFN (h) Ours

Figure 4. Qualitative comparison results. Best viewed zoomed-in on screen.

Method	PSNR	SSIM	LPIPS	#Params
EDSR (SR)	24.948	0.84	0.37149	1.517.571
HAN (SR)	25.001	0.841	0.36856	16.071.745
MPRNet(DB) + EDSR(SR)	26.521	0.873	0.33975	21.644.644
MPRNet(DB) + HAN(SR)	26.571	0.874	0.33763	36.198.818
DMPHN(DB) + EDSR(SR)	24.952	0.832	0.41353	8.463.875
EDSR(SR) + DMPHN(DB)	24.843	0.835	0.35037	8.463.875
DMPHN(DB) + HAN(SR)	24.938	0.832	0.41495	23.018.049
HAN(SR) + DMPHN(DB)	24.991	0.84	0.33964	23.018.049
GFN	27.91	0.902	0.29642	12.210.438
Ours	**28.265**	**0.9086**	**0.28522**	21.839.972

Table 1. Quantitative comparison results

Some visual comparisons on the GOPRO test dataset are presented in Fig. 4. As can be seen in Fig. 4, the concatenations of the state-of-the-art SR and deblurring models fail to remove complex motion blur effects and reconstruct fine image details. They often introduce undesired artifacts by amplifying the blur effects. The images generated by our model are sharper and more close to the ground truth compared to the images generated by the existing state-of-the-art joint super-resolution and deblurring method GFN.

V. CONCLUSION

In this work, we propose an end-to-end single branch pipeline for the joint motion deblurring and super-resolution task. Our network effectively reconstructs a sharp high-resolution image given the corresponding motion blurry low-resolution image. Unlike the previous solutions, our network uses a single-branch architecture, thus reusing the features responsible for motion deblurring task in super-resolution problem. Experiments show that our method outperforms the existing state-of-the-art solutions both qualitatively and quantitatively.

REFERENCES

[1] X. Li and M. Orchard, "New edge-directed interpolation," *IEEE Transactions on Image Processing*, vol. 10, no. 10, pp. 1521-1527, 2001.

[2] H. Chang, D. Yeung and Y. Xiong, "Super-resolution through neighbor embedding", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, USA, pp. 1275-1282, 2004.

[3] J. Yang et al., "Coupled Dictionary Training for Image Super-Resolution", *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467-3478, 2012.

[4] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, pp. 105-114, 2017

[5] B. Lim et al., "Enhanced Deep Residual Networks for Single Image Super-Resolution," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, USA, pp. 1132-1140, 2017

[6] B. Niu et al., “Single image super-resolution via a holistic attention network”, *Proceedings of European Conference on Computer Vision (ECCV)*, Glasgow, UK, pp. 191–207, 2020.

[7] J. Pan, D. Sun, H. Pfister and M. Yang, "Blind Image Deblurring Using Dark Channel Prior", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 1628-1636, 2016.

[8] J. Pan, Z. Hu, Z. Su, M. Yang; “Deblurring Text Images via L0-Regularized Intensity and Gradient Prior”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, USA, pp. 2901-2908, 2014.

[9] S.W. Zamir et al., “Multi-Stage Progressive Image Restoration”. *arXiv preprint arXiv:2102.02808v2*, 2021.

[10] S. Nah, T. Kim, K. Lee, “Deep Multi-Scale Convolutional Neural Network for Dynamic Scene Deblurring”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, pp. 257-265, 2017.

[11] H. Zhang, Y. Dai, H. Li, P. Koniusz, “Deep Stacked Hierarchical Multi-Patch Network for Image Deblurring”, *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, USA, pp. 5971-5979, 2019.

[12] M. Suin, K. Purohit, A. Rajagopalan, “Spatially-Attentive Patch-Hierarchical Network for Adaptive Motion Deblurring”, *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, USA, pp. 3603-3612, 2020.

[13] F-J. Tsai et al., “BANet: Blur-aware Attention Networks for Dynamic Scene Deblurring”, *arXiv preprint arXiv:2101.07518*, 2021.

[14] X. Zhang, F. Wang, H. Dong and Y. Guo, "A Deep Encoder-Decoder Networks for Joint Deblurring and Super-Resolution", *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, Canada, pp. 1448-1452, 2018.

[15] X. Zhang et al. "Gated Fusion Network for Degraded Image Super Resolution", *International Journal of Computer Vision,* vol. 128, no 1, pp. 1699–1721, 2020.

[16] Y. Zhang et al., “Image Super-Resolution Using Very Deep Residual Channel Attention Networks”, *Proceedings of European Conference on Computer Vision (ECCV)*, Munich, Germany, pp. 294-310, 2018.

[17] K. He et al, “Deep residual learning for image recognition”, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 770-778, 2016.

[18] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, *arXiv preprint arXiv:1409.1556,* 2015.

[19] X. Xu et al., "Learning to Super-Resolve Blurry Face and Text Images", *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 251-260, 2017.

[20] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Lima, Peru, pp. 234-241, 2015.

[21] W. Shi et al., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 1874-1883, 2016.

[22] P. Charbonnier et al, "Two deterministic half-quadratic regularization algorithms for computed imaging", *Proceedings of IEEE conference on Image Processing*, Austin, USA, pp. 168-172, 1994.

[23] D. Kingma, J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

[24] Wikipedia, (2014) The peak signal-to-noise ratio. [Online]. Available: https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

[25] Z. Wang et al., "Image quality assessment: from error visibility to structural similarity", *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, 2004.

[26] R. Zhang et al., "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric", *Proceedings of IEEE/CVF Conference on CVPR*, Salt Lake City, USA, pp. 586-595, 2018.

[27] R. Gonzalez, R. Woods, *Digital Image Processing*, 4th ed., Pearson, New York, 2018