Mathematical Problems of Computer Science 55, 44--53, 2021

UDC 004.932

Application of Deep Learning-Based Methods to the Single Image Non-Uniform Blind Motion Deblurring Problem

Misak T. Shoyan[1], Robert G. Hakobyan[1] and Mekhak T. Shoyan[2]

[1] National Polytechnic University of Armenia

[2] Yerevan State University, Armenia

e-mail: misakshoyan@gmail.com, rob.hakobyan@gmail.com, mexakshoyan@gmail.com

Abstract

In this paper, we present deep learning-based blind image deblurring methods for estimating and removing a non-uniform motion blur from a single blurry image. We propose two fully convolutional neural networks (CNN) for solving the problem. The networks are trained end-to-end to reconstruct the latent sharp image directly from the given single blurry image without estimating and making any assumptions on the blur kernel, its uniformity, and noise. We demonstrate the performance of the proposed models and show that our approaches can effectively estimate and remove complex non-uniform motion blur from a single blurry image.

Keywords: Motion blur, Blind motion deblurring, Non-uniform blurring, Blur kernel.

1. Introduction

Motion blur is one of the most undesired types of image degradation when taking photos. The shake of the camera and the object motion during the exposure cause motion blurry images. Motion blur is an undesirable effect, particularly in photography, and still is considered an effect, which causes a significant distortion of an image. The process of recovering the latent sharp image from a single motion blurry image or from a sequence of blurry video frames is called motion deblurring. In practice, there are a large number of possible motion paths, and every motion-blurred image is uniquely blurred, thus motion deblurring is a common and challenging problem nowadays.
A high-level representation of the blurring process is the following model

$$b = I \otimes \mathrm{f} + \mathrm{n}, \qquad (1)$$

where I is the latent sharp image, f is the blur kernel, n denotes the noise, and $\otimes$ is the convolution operator. In the presence of only one blurry image, the problem is called single-image motion deblurring. In the case of multiple sequential blurry images, the problem is called multi-image/video motion deblurring. Our interest is mainly related to single-image motion deblurring.

If the blur kernel or point spread function (PSF) is shift-invariant in the sense that blurring is uniform, then the deblurring problem turns into the image deconvolution problem. When the point spread function (PSF) is shift-variant and therefore the blurring is non-uniform, then it is considered a deblurring problem.

Image deblurring is categorized as non-blind and blind cases. In the case of non-blind deblurring, the blur kernel is known, or there is a way to compute it using some prior knowledge, so the problem turns to estimate the latent sharp image given the known blur kernel. There are some difficulties to overcome even though it may seem not a hard task. For example, the presence of noise and possible ringing artifacts arising during deblurring make it a challenging problem.

There are some traditional methods such as Wiener deconvolution [1] which is expressed as

$$G(f) = \frac{H^*(f)\, S(f)}{|H(f)|^2\, S(f) + N(f)} \tag{2}$$

where f is the frequency in the frequency domain, G is the Fourier transform of the estimated kernel, which then is convolved with the blurry image to estimate the latent sharp image, H is the Fourier transform of the blur kernel, N and S are the mean power spectral density of the noise and latent sharp image respectively, $*$ denotes the complex conjugation. Iterative Richardson-Lucy (RL) [2, 3] deconvolution is another method, which is expressed as

$$I^{t+1} = I^t \left(PSF^T \otimes \left(\frac{B}{I^t \otimes PSF} \right) \right), \tag{3}$$

where I^t and I^{t+1} are t^{th} and $(t+1)^{th}$ estimations of the latent sharp image I, B is the blurry image and PSF^T is the flipped version of PSF.

These methods were presented decades ago. In further studies, the solution to the problem of non-blind deblurring tends to be based on many famous image priors, for example, sparse priors [4] and total variation [5], which have been introduced for regularization purposes to improve the quality of deconvolution in the presence of noise.

The blind deblurring [6] is a more challenging problem since in this case the blur kernel or PSF is also unknown in addition to the unknown latent sharp image. The blind deblurring problem consists of two stages: the PSF estimation and non-blind deconvolution. In contrast to non-blind deblurring, more sophisticated priors have been introduced here, such as norm-based prior [7], dark channel prior [8], reweighted graph total variation prior [9], etc.

Image deblurring methods are also categorized as deep learning-based (DL) and non-deep learning-based (non-DL) or optimization-based methods. Non-DL-based or optimization-based methods try to reconstruct the latent sharp image by minimizing the energy function [10, 11], using, for example, Gaussian or Poisson likelihoods in the scope of maximum-a-posteriori estimation [12].

Even though non-DL-based methods are effective in image deblurring, they are usually based on relatively simplified assumptions on the blur model compared with DL-based methods. It is also worth mentioning the time-consuming hyperparameter tuning process for non-DL-based methods, which is significant in real-world cases. In recent years, DL-based approaches have become more and more applicable. DL-based methods use convolutional neural networks to reconstruct the latent sharp image [13]. Also, recurrent neural networks are used for single image deblurring [14]. In terms of both accuracy and efficiency, these methods exceed non-DL methods. So, we present deep learning-based blind image deblurring methods for estimating and removing non-uniform motion blur from a single blurry image.

2. Dataset

A common practice for creating a dataset for supervised image deblur problems is to synthetically generate blurry images by blurring latent sharp images with a kernel and then adding some noise [15, 16]. However, the blurry images generated in this way may differ from a real blurry image, and the dataset might not be representative enough.

A new kernel-free approach of dataset generation for supervised motion deblur problems was proposed in [17]. They used a GOPRO4 Hero Black camera for dataset generation. They record high-quality videos with 240 fps and then average sequential video frames of latent sharp images to produce motion blurry images [18]. The corresponding latent sharp image for the generated blurry image is chosen as the middle image of the sequence that is used to average and generate the blurry image.

When the motion blur is caused by the motion of an object, the blurriest part of the blurry image should be the object itself, leaving the background mostly the same as in the latent sharp image. The proposed kernel-free dataset generation method [17] for supervised motion deblur problems solves that problem unlike the other methods [15, 16].

We chose the GOPRO dataset [18] for training and evaluating our models. The dataset contains 3214 pairs of blurry and sharp images.

3. Proposed Methods

We propose two encoder-decoder architecture based fully convolutional neural networks.
The first one (ResnetEncDec) uses Resnet-50 [19] as an encoder. It receives a 3x256x256 RGB image as input. The first step is a convolution with a 7x7 kernel with stride 2 followed by max-pooling with stride 2. Then the Resnet-50 residual blocks follow, which use 1x1 and 3x3 convolutions. Each convolution layer is followed by a batch normalization layer [20] and ReLU activation. The encoder part outputs a 2048x8x8 feature map, which is used as an input of the decoder part.

The decoder part consists of transposed convolution and upsample layers. First, 3 decoder blocks follow, each of which consists of a transpose convolution layer followed by 2 convolutions. Then, 2 upsample layers follow, each of which performs a bilinear upsampling with a factor of 2 followed by 2 convolutions. Then, a 1x1 convolution follow to reduce the channels of the activation map to 3. Then, a sigmoid activation follow to output colors in [0, 1] range for each pixel of the output image. All the convolution and deconvolution layers are followed by batch normalization and ReLU activation (except the last convolution layer, which is followed by sigmoid activation).

The skip connections are used between the encoder and decoder layers inspired by the U-Net architecture [21]. The architecture of the network is shown in Figure 1.

The next proposed network is inspired by the real-time style transfer method proposed in [22]. They propose using an image transform network (TransformNet) for the style transfer problem to stylize the input content image with the style of the style image (Fig 2). Since the network performed well on style transfer image to image problem, thus, being able to generate an image that is some modified version of the input image, we proposed it for the motion deblur problem.

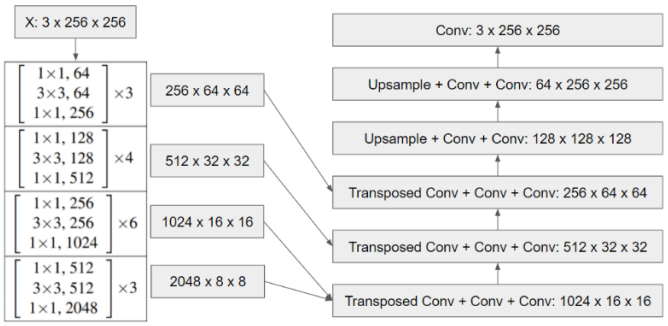


Fig. 1. The architecture of the ResnetEncDec fully convolutional network.

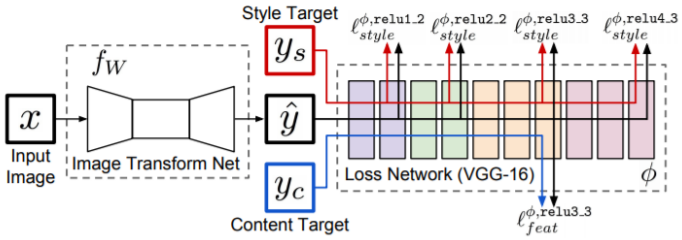
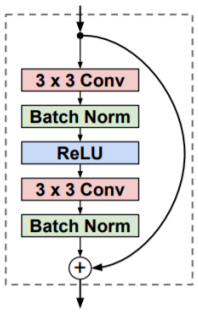


Fig. 2. The architecture of the style transfer network [22].

The first layer of the proposed Transform Net is a 9x9 convolution with stride 1. Then two 3x3 convolutions follow with stride 2. Then, 5 residual blocks follow, each of which consists of two 3x3 convolutions followed by batch normalization and ReLU activation (Fig. 3). Each residual block contains a residual connection between its input and output. After the 5 residual blocks, two 3x3 transposed convolution layers follow with stride 2. Then, a 9x9 convolution follow with stride 1. Finally, sigmoid activation follows to output colors in [0, 1] range for each pixel of the output image. Each convolution layer is followed by batch normalization and ReLU activation (except the last convolution layer, which is followed by sigmoid activation).

Layer	Activation size
Input	$3 \times 256 \times 256$
$32 \times 9 \times 9$ conv, stride 1	$32 \times 256 \times 256$
$64 \times 3 \times 3$ conv, stride 2	$64 \times 128 \times 128$
$128 \times 3 \times 3$ conv, stride 2	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
Residual block, 128 filters	$128 \times 64 \times 64$
$64 \times 3 \times 3$ conv, stride 1/2	$64 \times 128 \times 128$
$32 \times 3 \times 3$ conv, stride 1/2	$32 \times 256 \times 256$
$3 \times 9 \times 9$ conv, stride 1	$3 \times 256 \times 256$

(a)



(b)

Fig. 3. (a) The architecture of the TransformNet. [23] (b) The architecture of each residual block [23].

4. Training

Both proposed networks are trained on the GOPRO dataset with 256x256 resized images. Since we want to minimize the pixel-wise differences between the output and latent sharp image in the motion deblur problem, we chose MSE [24] and MAE [25] as loss functions:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \quad (4)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \quad (5)$$

where N is the number of pixels in the image, y is the pixel value of the sharp image and $\hat{y}$ is the predicted pixel value.

Our experiments showed that MSE performs better for both of the networks, at least at the early steps of training, so we used MSE for further experiments.

As evaluation metrics we chose PSNR (peak signal-to-noise ratio) [26] and MSE functions:

$$PSNR = 20\log_{10}\left(\frac{MAX_i}{\sqrt{MSE}}\right), \quad (6)$$

where MAX_i is the maximum possible pixel value of the image.

The Adam optimizer [27] was used with a learning rate of 0.001. Both networks are trained for 350 epochs with batch sizes 15 and 44 for ResnetEncDec and TransformNet correspondingly running on GeForce GTX 1070 Ti GPU. ImageNet [19] pre-trained weights are used to initialize the ResnetEncDec encoder part. For TransformNet, training continued additionally for 250 epochs with SGD optimizer [28] without momentum with a learning rate of 0.0001. However, it does not lead to significant improvements.

The learning curves of both networks are shown in Figure 4.

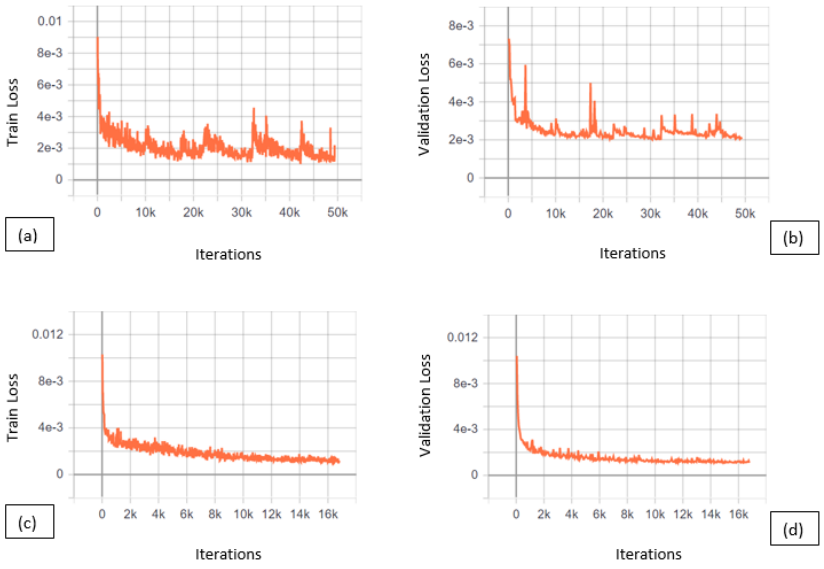


Fig 4. The learning curves of ResnetEncDec (a, b) and TransformNet(c, d).

5. Results

We evaluate the performance of our proposed models on the GOPRO dataset. The results are compared with one of the state-of-the-art methods [17]. The quantitative performance comparison of the proposed models is shown in Table 1 (note that we use 256x256 resized images, while in [17] they use images with an original size of 1280x720).

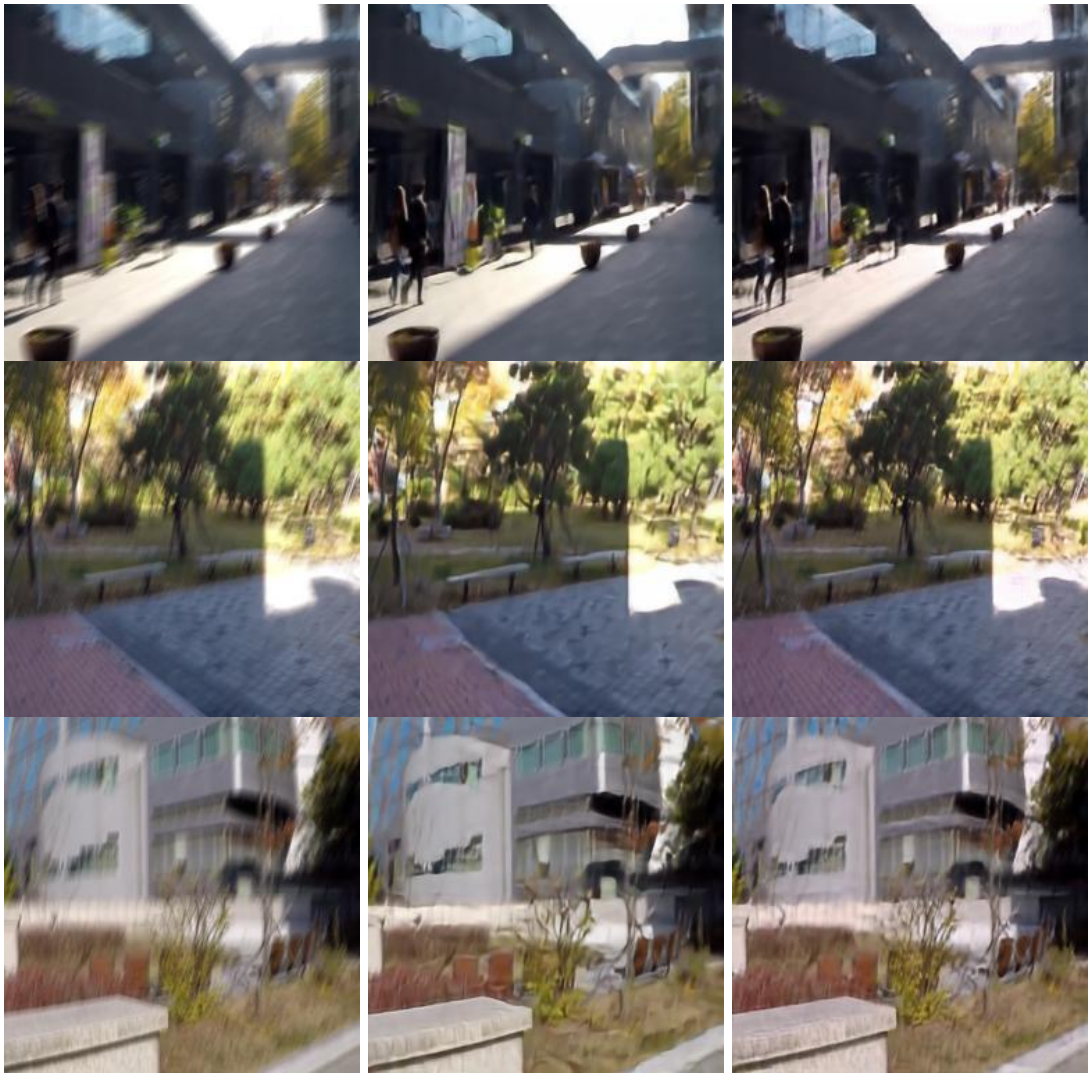
Table 1: Quantitative performance comparison of the models.

Metrics	*ResNetEncDec*	*TransformNet*	*Nah et al. [17]*
PSNR	24.98	26.26	28.93
MSE	0.0033	0.00245	-

Some deblurring results are shown in Fig. 5.

In terms of performance and memory usage, the TransformNet and ResNetEncDec are lightweight networks compared to [17], since [17] relies on a deep multi-scale architecture.

At the same time, as it is obvious from the architectures of the proposed networks, the TransformNet is more lightweight and requires less computational time and resources than the ResNetEncDec.



Input image ResnetEncDec result TransformNet result

Fig 5. The results on GOPRO test dataset.

6. Conclusion

In this paper, two deep learning-based blind motion deblurring methods were presented to reconstruct the latent sharp image from a single motion blurry image without having any information about the blur kernel, its uniformity, and existing noise. The proposed methods, which are encoder-decoder architecture-based fully convolutional neural networks, were trained, validated and evaluated on the GOPRO dataset [18] (using 256x256 resized images) and compared with one of the state-of-the-art methods presented in [17]. Based on the results shown in Table 1 and Figure 5, it becomes clear that the proposed methods can effectively remove complex non-uniform motion blur demonstrating acceptable results. The code and results are available at https://github.com/Mekhak/motion_deblur_dl.

Future work should address improving the accuracy of the proposed methods.

References

[1] *Wikipedia, (2008) Wiener Deconvolution. [Online]. Available:* https://en.wikipedia.org/wiki/Wiener_deconvolution

[2] W. Richardson, "Bayesian-based iterative method of image restoration", *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55-59, 1972.

[3] L. Lucy, "An iterative technique for the rectification of observed distributions", *The Astronomical Journal*, vol. 79, no. 6, pp. 745-754, 1974.

[4] D. Krishnan and R. Fergus, "Fast image deconvolution using hyperlaplacian priors", *Proceedings of the 23rd International Conference on Neural Information Processing Systems*, Vancouver, Canada, pp. 1033–1041, 2009.

[5] L. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms", *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[6] A. Levin, Y. Weiss, F. Durand and W. Freeman, "Understanding blind deconvolution algorithms", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2354–2367, 2011.

[7] J. Pan, Z. Hu, Z. Su and M. Yang, "L0-regularized intensity and gradient prior for deblurring text images and beyond", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 2, pp. 342-355, 2017.

[8] J. Pan, D. Sun, H. Pfister and M. Yang, "Blind image deblurring using dark channel prior", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 1628-1636, 2016.

[9] Y. Bai, G. Cheung, X. Liu and W. Gao, "Graph-Based Blind Image Deblurring From a Single Photograph", *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1404-1418, 2019.

[10] S. Cho and S. Lee, "Fast motion deblurring", *ACM Transactions on Graphics*, vol. 28, no. 5, article 145, pp. 1-8, 2009.

[11] S. Zheng, L. Xu and J. Jia, "Forward motion deblurring", *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Sydney, Australia, pp. 1465-1472, 2013.

[12] Wikipedia, (2016) The maximum-a-posteriori estimation. [Online]. Available: https://en.wikipedia.org/wiki/Maximum_a_posteriori_estimation

[13] L. Xu, J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution", *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Montreal, Canada, pp. 1790–1798, 2014.

[14] J. Zhang, J. Pan, J. Ren, et al., "Dynamic scene deblurring using spatially variant recurrent neural networks", *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, USA, pp. 2521-2529, 2018.

[15] T. Nimisha, V. Rengarajan and R. Ambasamudram, "Semi-Supervised Learning of Camera Motion from a Blurred Image", *Proceedings of the 25th IEEE International Conference on Image Processing (ICIP)*, Athens, Greece, pp. 803-807, 2018.

[16] J. Sun, W. Cao, Z. Xu and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, USA, pp. 769-777, 2015.

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

[18] S. Nah, (2017) The GOPRO dataset. [Online]. Available: https://seungjunnah.github.io/Datasets/gopro

[19] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, pp. 770-778, 2016.

[20] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *Proceedings of the 32nd International Conference on Machine Learning*, Lille, France, pp. 448-456, 2015.

[21] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, pp. 234-241, 2015.

[22] J. Johnson, A. Alahi, and L. Fei, "Perceptual losses for real-time style transfer and super-resolution", *Proceedings of the European Conference on Computer Vision (ECCV)*, Amsterdam, The Netherlands, pp. 694-711, 2016.

[23] J. Johnson, (2016) Perceptual Losses for Real-Time Style Transfer and Super-Resolution: Supplementary Material. Link for Fig. 3 a-b. [Online]. Available: https://cs.stanford.edu/people/jcjohns/papers/fast-style/fast-style-supp.pdf

[24] Wikipedia, (2019) The mean squared error. [Online]. Available: https://en.wikipedia.org/wiki/Mean_squared_error

[25] Wikipedia, (2017) The mean absolute error. [Online]. Available: https://en.wikipedia.org/wiki/Mean_absolute_error

[26] Wikipedia, (2013) The peak signal-to-noise ratio. [Online]. Available: https://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio

[27] D. Kingma and J. Ba, (2017) arXiv paper page - Adam: A Method for Stochastic Optimization. [Online]. Available: https://arxiv.org/abs/1412.6980v5

[28] Wikipedia, (2020) The stochastic gradient descent. [Online]. Available: https://en.wikipedia.org/wiki/Stochastic_gradient_descent

Submitted 18.12.2020, accepted 22.03.2021

Խորը ուսուցման վրա հիմնված մեթոդների կիրառումը պատկերում շարժման հետևանքով առաջացած ոչ միատարր պղտորման հեռացման կույր խնդրում

Միսակ Տ. Սհոյան[1], Ռոբերտ Գ. Հակոբյան[1] և Մեխակ Տ. Սհոյան[2]

[1] Հայաստանի Ազգային Պոլիտեխնիկական Համալսարան

[2] Երևանի Պետական Համալսարան

e-mail: misakshoyan@gmail.com, rob.hakobyan@gmail.com, mexakshoyan@gmail.com

Ամփոփում

Հոդվածում ներկայացվում են խորը ուսուցման վրա հիմնված պատկերից պղտորման հեռացման կույր մեթոդներ՝ պատկերում շարժման հետևանքով առաջացած ոչ միատարր պղտորման գնահատման և հեռացման համար: Խնդրի լուծման համար առաջարկվում են երկու լրիվ փաթույթային նեյրոնային ցանցեր (CNN): Տրված պղտորված պատկերից սկզբնական սուր պատկերը վերականգնելու համար ներկայացված ցանցերը ուսուցանվում են ամբողջապես՝ առանց գնահատելու և որևէ ենթադրություններ անելու պղտորման միջուկի, նրա միատարրության և առկա աղմուկի վերաբերյալ: Ցուցադրվում է առաջարկվող մոդելների արտադրողականությունը և ցույց է տրվում, որ առաջարկվող մոտեցումները կարող են արդյունավետորեն գնահատել և հեռացնել շարժման հետևանքով պատկերում առաջացած բարդ, ոչ միատարր պղտորումը:

Բանալի բառեր՝ Շարժման հետևանքով առաջացած պղտորում, շարժման հետևանքով առաջացած պղտորման հեռացում, ոչ միատարր պղտորում, պղտորման միջուկ:

Применение методов глубокого обучения в задаче слепого устранения размытости вслед за движением из одного неоднородно размытого изображения

Мисак Т. Сгоян[1], Роберт Г. Акопян[1], Мехак Т. Сгоян[2]

[1] Национальный Политехнический Университет Армении

[2] Ереванский Государственный Университет

e-mail: misakshoyan@gmail.com, rob.hakobyan@gmail.com, mexakshoyan@gmail.com

Аннотация

В этой статье представляются слепые методы устранения размытости изображения основанные на глубоком обучении – для оценки и удаления неоднородного размытия вслед за движением из одного размытого изображения. Для решения задачи предлагаются две полностью сверточные нейронные сети (CNN). Сети, предназначенные для восстановления исходного резкого изображения из размытого изображения, обучаются полностью – без оценки и каких-либо предположений о кернеле размытия, его однородности и присутствующего шума. Демонстрируется производительность предложенных моделей и показано, что предложенные методы могут эффективно оценивать и устранить сложное неоднородное размытие вслед за движением из одного размытого изображения.

Ключевые слова: Размытие из за движения, слепое устранение размытости вслед за движением, неоднородное размытие, кернел размытия.