Reinforced Hybrid Wiener Deconvolutional - Convolutional Autoencoders Based Image Deblurring

George F. Grosu
Communications Department
Technical University of Cluj-Napoca
Transylvanian Institute of Neuroscience
Cluj-Napoca, Romania
George.Grosu@com.utcluj.ro

Romulus Terebes
Communications Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
Romulus.Terebes@com.utcluj.ro

***Abstract*—This paper is addressing the issue of image deblurring by employing a novel reinforced hybrid Wiener deconvolutional-convolutional neural network (HWDCNN) in the context of images affected by Gaussian blur. The proposed method is set to explore the capabilities of a custom NN layer that performs the classical Wiener deconvolution operator in a trainable manner, hence, leading to a blind deblurring method where the weights represent the point spread functions (PSF) and signal-to-noise ratios (SNR) of the deconvolution layer which are optimized in a traditional deep learning paradigm. Additionally, the custom layer is set as a first layer on two of the branches of the model, followed by layers of transposed convolution to explore possible synergies. The results indicate that this method outperforms known state-of-the-art methods for the presented dataset setup in terms of both structural similarity index measure (SSIM) and peak SNR (PSNR) metrics.**

Keywords—Deblurring, Deconvolution, Deblurring, Neural Networks, Gaussian blur

I. Introduction

In general, any signal acquisition process is subjected to distorting effects due to setup and camera imperfections and scene's dynamicity. These imperfections can originate from various sources that add up to the measurements both linearly and nonlinearly [1]. For a single image, different patches at different scales can be affected by a variety of complex blur factors. Related to optics, focal length and aperture size of the measuring device can lead to different levels of Gaussian blur. Operation errors can also lead to out of focus blur, camera shake, motion blur or a mixture of them, both uniformly and nonuniformly distributed. On one hand, Gaussian blur can drastically lead to loss of edge information as the gradients are greatly reduced in affected image. On the other hand, motion blur results in ghost residuals that break the integral edge information, forming highly dense and nonlinear edges.

The issue of signal restoration aims to mitigate measurement distortions as adequately and efficiently as possible, such that the restored signal has a one-to-one match representation, ideally at the same instance of time when the measurement is performed. Addressing every source of distortion at once is difficult and usually proves to be an unrealistic approach, reason why various methods were developed to address separately distortion effects. Blurring effect is among the most common distortions encountered in analog and digital image acquisition, which is given by several reasons, from relativistic motion between the measuring device and the target objects, inadequate calibration of the measuring device focus system to lighting or electromagnetic radiation conditions which can limit the number of captured photons and increases the required exposure time, in which again, chances for pronounced relative motion are increased [2]. Image deblurring is classified as a low-level, non-trivial issue within the computer vision and image processing frameworks, marking an important process of high interest for various domains from medical and astronomy imaging to surveillance, automation systems and media production. The base assumption is that, for any arbitrary blurred image corresponds a single sharp, latent image. Nonetheless, blur inversion is essentially an ill-posed issue as for any blurred image exists not one, but at least set of images that can correspond to that blurry image. Gradually increasing the blur distortion will lead to an increase in size of the set of corresponding latent images. Moreover, the image can pass a critical point of entropy from which the set of latent sharp images can become unreasonably enormous. This phenomenon can scale from a local patch within the image up to the entire sample. To get a better view, one can consider a simple image which captures details at different scales and a type of blur is applied. By increasing the blurring effect, the finer details are easier distorted up to a level in which recovery of actual information is totally lost and the left information is equally correlated with a vast set of latent representations. Continuing to increase the effect this phenomenon scales up in a finer-to-coarse manner. This is directly related to the low-pass filtering characteristics of the blurring process, where the level of blurring effect is a function of the cut-off frequency, slope and attenuation level in spectral representation [3].

In conventional image processing approaches the problem is focused on identifying an inverse filtering operation for which a single or a set of PSF can reasonably reverse the blurring effects of the target images. Such a process is conventionally named deconvolution. This problem formulation is the most direct and intuitive approach in the framework of signal processing, where a blurred image is modeled as the convolution between a sharp image and a PSF, or a multiplication between the Fourier transform of the image and the transfer function. Depending on how the deconvolution operation is performed, this *first problem formulation* can prove to be ill-posed and unable to provide a unique solution. In a more abstract direction, *the second formulation* of deblurring is envisioned as an unknown function of a latent sharp image and a parameter vector. The deblurring process becomes the inverse of that function. This later formulation became a fit way to describe the issue from

978-1-6654-5150-5/22/$31.00 ©2022 IEEE

a machine learning (ML) perspective, especially in the blind approach. In this paper only the context of Gaussian deblurring is addressed through ML techniques.

The rest of the paper is continued according to the following organization: Section II consists mainly of a brief presentation of the existing related work on the issue of image deblurring and proposed approaches, Section III presents the employed methods for the work of this paper, Section IV is reserved for the results of this work, and finally, Section V presents the conclusions related to this work and the directions for future work.

II. Related work

With the increasing interest from a vast set of technological and scientific domains into the subject of deblurring, a substantial level of efforts led in impressive results. Traditional methods approached this issue by separating the problem into two stages: firstly, the estimation of the blurring PSF, and secondly, the restoration of the image. This approach is known as the non-blind deblurring, whereas blind deblurring aims to restore an image without knowledge about the actual blurring PSF.

Among the early and most common deblurring methods that emerged from the first type of problem formulation, deblurring as a deconvolution process, are the popular Wiener deconvolution and Lucy-Richardson (LR) deconvolution. The Wiener deconvolution method is a relatively simple, linear operation that inverses the blurring effect by a simple division of the image to a regularized estimated PSF in spectral domain. On the other hand, the LR method is a nonlinear, iterative method that can achieve greater performance, but becomes inefficient if the PSF is a function of unknown variables. Among these, another highly popular and widely employed nonlinear method is the Total Variation deconvolution algorithm which is specifically specialized in recovering sharp edges. Later, more sophisticated variations of these methods emerged, progressively introducing computational complexity with yet limited capabilities.

With the advancement of deep learning methods, a rich repertoire of blind & non-blind deblurring solutions became available, making up for the vast majority of the related state-of-the-art [4]. Initially, convolutional neural networks (CNN) were employed such that the convolution kernels estimate the PSF, constricting two levels, firstly for estimation and secondly for deconvolution [5]. Soon after, more straightforward approaches were developed in which CNNs were employed to directly identify a mapping between the blurred and sharp image. The work presented in [6] introduced DeepDeblur, a multi-scale (MS) CNN architecture to solve dynamic scene blurring of multiple sources. The method is based on scale specialized stacked subnetworks that restore the latent image at downscaled versions in a coarse-to-fine procedure. Inspired by the success of the pyramidal model architecture and coarse-to-fine multiscale processing approach, other forms of this architecture have been proposed. In the work of [7] a similar scale recurrent NN (SRN) version of such pyramidal architecture is employed, matching and even outperforming the original with a more lightweight model. Another alternative based on recurrent NN is given by [8] where instead of a multi-scale training approach, a multi-temporal progressive training approach is provided. The relatively recent work of [9] addressed the issue by proposing a totally different approach with the motivation to reduce model size by introducing modulated deformable convolutions for adaptive receptive fields which are argued to estimate the blurring PSF. Approaches based on generative adversarial networks (GANs) are of high popularity, managing to situate themselves among the top results [10]. Nonetheless, such approaches are difficult to train due to high instability and result in heavy, computationally demanding models. State-of-the-art results have been achieved by exploiting a model architecture initially proposed for image segmentation tasks such as the increasingly popular U-Net [11] which is composed of contracting paths capable to capture context and of symmetric expanding paths capable of precision localization. A modified version of U-Net, namely multiple input, multiple output (MIMO) UNet++ is proposed in the work of [12] where the encoder of UNet is adjusted to have as input multiscale images, similar to previous MS approaches, the decoder outputs deblurred images for a set of scales and an asymmetric feature fusion operation is introduced with the role to fuse the resulting multiscale features. The results indicate that the model outperforms the compared methods in terms of quality of the deblurred image and the computation time. In terms of images affected strictly by Gaussian blur and noise the work of [13] proposed a denoising prior driven deep NN (DPDNN), a model composed of denoising modules interleaved with back-projections, trained on the unfolded version of a custom iterative denoising algorithm. In paper [14] a multi-stream with attentional module and global information-based fusion network, namely MBANet, is proposed within a benchmark specifically designed for restoring images affected by Gaussian blur, obtaining results similar to the state-of-the-art under the similar setup configuration.

III. Methods

A novel type of model is proposed which aims to exploit synergies from the interplay between a custom trainable Wiener 2D depth-wise separable deconvolution layer, transposed convolution layers and common convolution layers to address the issue of image deblurring.

A. Trainable 2D Wiener deconvolution layer

Despite its simplicity, the Wiener deconvolution method in its logic-based form, can offer reasonable results in image deblurring, especially if the PSF and the SNR values are rightly chosen. This aspect leads to the idea of parameter optimization for the extraction of useful features. On this assumption, a novel layer is proposed that implements the Wiener deconvolution method, optimizing in the same instance the weights representing the PSFs and the SNRs of the operator in tensorial form. The Wiener deconvolution method follows the first formulation paradigm, where the PSF is either known or estimated to inverse the effects blur effects on the image. An image, or in general any measurement acquired in a real-world scenario can be described mathematically as in equation (1) for the corresponding time representation:

$$y(t) = x(t) * h(t) + n(t) \tag{1}$$

where $y(t)$ is the recorded measurement, $x(t)$ is the true, latent representation of the measurement, $h(t)$ is the system's kernel or PSF from which distortions such as blurring can result and $n(t)$ is the additive channel's noise, all of them at given arbitrary time instance t. In frequency representation the convolution operation becomes multiplication, described in equation (2) where every variable is represented by the

Fourier transform. From this point of view, one is expected to recover the latent image $X(\omega)$ by applying equation (3).

$$Y(\omega) = X(\omega) \cdot H(\omega) + N(\omega) \quad (2)$$

$$X(\omega) = \frac{Y(\omega) - N(\omega)}{H(\omega)} \quad (3)$$

But equation (3) implies firstly that $H(\omega)$ is known, which rarely is the case. Secondly, the higher spectra of the noise $N(\omega)$ can be drastically amplified by dividing with the PSF, drowning the actual measurement in such noise. Thirdly, the deconvolution in this form allows for multiple solutions and requires various constraints and regularizations. Hence, the Wiener deconvolution aims to overcome this issue by estimating an appropriate kernel $\widehat{H}(\omega)$ such that the product of $\widehat{H}(\omega) \cdot Y(\omega)$ gives a reasonable approximation $\widehat{X}(\omega)$ of the measured target. Therefore, the trainable 2D Wiener deconvolution layer implements the operation as described in equation (4), where $H(\omega)$ and the SNR represent the trainable weights and $Y(\omega)$ is the input image or feature map.

$$\widehat{X}(\omega) = \Re(Y(\omega) \cdot \frac{1}{H(\omega)} \cdot \frac{1}{1 + \frac{1}{|H(\omega)|^2 \cdot SNR}}) \quad (4)$$

Conventional 2D convolution layers perform the operation on all image channels by summing the convolved channels into a single channel, resulting in a single channel for each filter. The proposed 2D Wiener deconvolution layer performs the deconvolution separately for each channel of the input image, resulting $\widehat{X}(\omega)$. The layer defines the PSF size as equal to the size of the input image and performs the deconvolution at once with the entire image. The output is transformed back to real values by taking only the real parts of the complex coefficients. The layer is implemented with TensorFlow 2.9.1.

B. Model architecture

The architecture of the model is constructed based on the coarse-to-fine approach, from which subnetworks are stacked to process input images at different sizes. Instead of taking as input images at different sizes or sequentially stacking subnetworks to operate at different scales, the proposed model takes a single image as input which is divergently propagated in 6 branches of autoencoders. Each autoencoder is composed firstly by 3 downsampling blocks that perform downsampling by applying a 2D convolution. The second and fourth branches are equipped with a deconvolution block that explore possible synergies and more complex operations. The second part of the autoencoders is composed of 3 upsampling blocks on every branch. Each branch (top-down) progressively defines greater kernels (for down and upsample blocks) to capture different scales within the input image. The latent vector is discarded from the architecture, as the downsampled feature maps are considered sufficient for proper restoration. After the first and second upsampling block, the outputs are concatenated with the feature maps from the second, respectively the first downsampling blocks from the corresponding branch through residual connectivity. The outputs of the odd autoencoders are summed up, similarly as the outputs of the even autoencoders. The summed outputs are concatenated with an upsampled feature map of the input. Finally, the resulted tensor is again convolved and downsampled to the size of the original image, which is expected to resemble the latent, sharp image. *The final proposed model is composed of two sequentially stacked modules of the structure presented in Figure 1.*

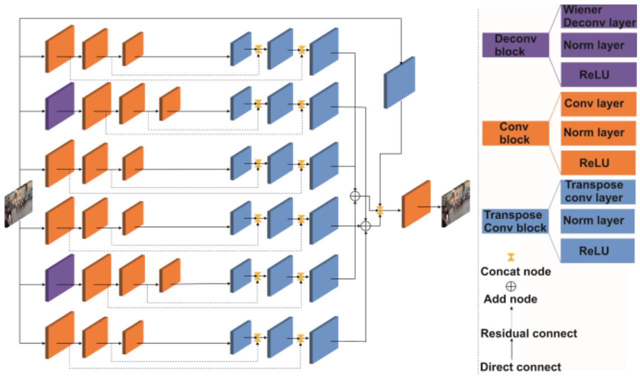

Figure 1. Structure of the module for the proposed deblurring model. The model is constructed of two sequentially connected such modules.

C. Dataset and model training

For the work of this paper the *GoPro dataset* is employed as it became popular for deblurring benchmarks, both by considering the original data or altering the sharp data. The dataset contains 2103 sharp images with a resolution of 720 × 1280 and the corresponding images with simulated nonuniform motion blur for training, respectively 1111 images for testing. For this paper the training and test data considers only the custom Gaussian blurring distortions by convolving the sharp images with a PSF of size 17 × 17 for a range of standard deviations, from 1.6 to 2.4, with an increment of 0.2, as proposed in [14]. The final model is trained in an iterative fashion, where a first module is trained on the data formed with blur from the entire range of standard deviations. The weights from the trained module are frozen and another module is trained on top with data separately for each standard deviation. The Adam optimizer, decaying learning rate on plateau and a starting learning rate of 1e-4 are set as optimizing strategies. The loss function is defined as $1 - SSIM$. Training time of the proposed method summed up to 29 hours for a notebook system with AMD Ryzen 9 5900HX and Nvidia RTX 3080 16GB VRAM, inferring on a single sample with an average time of 135ms, with the consideration that the novel layer is not CUDA optimized, whereas the method in [14] trained for 1.5 days on a different system, achieving overall lower quality results in terms of PSNR and SSIM metrics in comparison to proposed method.

IV. Results

Method is employed on the reserved GoPro test set with blurring setup discussed previously. To evaluate the performance of the proposed model, two standard metrics are considered, the *PSNR* and the *SSIM*. The results of the proposed method are compared with the results of the methods presented in [14] in Table 1 and Table 2.

Table 1. *PSNR metric results of state-of-the-art deep learning-based methods and the proposed method on the test data for the range of standard deviations of applied Gaussian blur.*

Methods	Standard deviations					Mean
	1.6	***1.8***	***2.0***	***2.2***	***2.4***	
SRN [14]	37.17	34.97	34.52	32.99	32.02	34.33
DPDNN [14]	37.58	36.49	35.10	33.82	33.08	35.21
MBANet [14]	37.99	36.48	35.16	34.04	32.99	35.33
Proposed	**38.87**	**37.71**	**36.84**	**35.56**	**34.39**	**36.67**

Table 2. *SSIM metric results of state-of-the-art deep learning-based methods and the proposed method on the test data for the range of standard deviations of applied Gaussian blur.*

Methods	Standard deviations					Mean
	1.6	***1.8***	***2.0***	***2.2***	***2.4***	
SRN [14]	0.9777	0.9687	0.9648	0.9529	0.9430	0.9614
DPDNN [14]	0.9808	0.9760	0.9691	0.9598	0.9549	0.9681
MBANet [14]	**0.9829**	0.9760	0.9699	0.9619	0.9524	0.9686
Proposed	0.9812	**0.9765**	**0.9729**	**0.9626**	**0.9552**	**0.9696**

From the results presented in Table 1 and Table 2, the proposed method outperforms the previous methods analyzed in [14] for the suggested metrics, exception being the SSIM metric for the setup corresponding to 1.6 standard deviation. Although trained on the SSIM loss function, interestingly the method manages to outperform more on the PSNR metric. In Figure 2 are presented the deblurring comparison results of the model and compared methods on an arbitrary test sample. In terms of visual perception there are two main differences brought by the proposed method: firstly, edges become more prominent in comparison with the other methods; secondly, the restored test image with the proposed method resulted in a slight color shift from the original image, suggesting that the method requires more training time or the statistics of the test sample apart from the statistics of the training set.

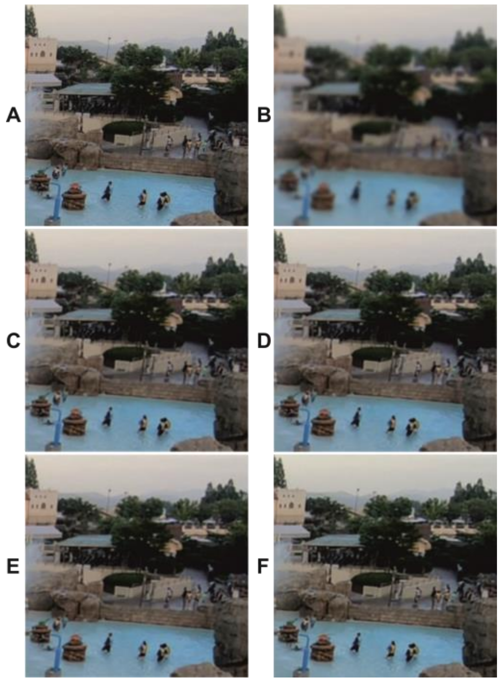

Figure 2. A) Cropped patch of sharp test sample. B) Gaussian blurred patch of test sample with 1.8 standard deviation. C) Restored patch by SRN. D) Restored patch by DPDNN. E) Restored patch by MBANet. F) Restored patch by proposed method.

V. Conclusions

The work of this paper addressed the issue of restoring images affected by Gaussian blur in 5 different scenarios with varying standard deviation by proposing a HWDCNN with parallel branching of autoencoders and enhanced with a novel Wiener deconvolution layer. In addition to the novel layer, the model is reinforced by an iterative training approach, stacking and training the same module on top of a previously trained module, method based on the assumption that the new module can exploit the information provided by a previous restoration module. The obtained results suggest that the proposed method can outperform the state-of-the-art methods on the same setup of the training and testing dataset.

Future work will consist of extending the restoration method for performing on a higher diversity of blurring effects and characteristics, such as motion blur, shake blur and mixed nonuniform blur with saturation artefacts.

References

[1] Y. Chen *et al.*, "The face image super-resolution algorithm based on combined representation learning," *Multimed Tools Appl*, vol. 80, no. 20, pp. 30839–30861, Aug. 2021, doi: 10.1007/s11042-020-09969-1.

[2] A. H. Abdulkhaleq, N. K. El Abbadi, and S. A. Al Hassani, "A Survey on Blind De-Blurring of Digital Image," *eijs*, pp. 338–352, Jan. 2022, doi: 10.24996/ijs.2022.63.1.32.

[3] J. C. Russ, "Image Processing," in *Computer-Assisted Microscopy*, Boston, MA: Springer US, 1990, pp. 33–69. doi: 10.1007/978-1-4613-0563-7_3.

[4] J. Koh, J. Lee, and S. Yoon, "Single-image deblurring with neural networks: A comparative survey," *Computer Vision and Image Understanding*, vol. 203, p. 103134, Feb. 2021, doi: 10.1016/j.cviu.2020.103134.

[5] J. Sun, Wenfei Cao, Zongben Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 769–777. doi: 10.1109/CVPR.2015.7298677.

[6] S. Nah, T. H. Kim, and K. M. Lee, "Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, pp. 257–265. doi: 10.1109/CVPR.2017.35.

[7] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-Recurrent Network for Deep Image Deblurring," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, Jun. 2018, pp. 8174–8182. doi: 10.1109/CVPR.2018.00853.

[8] D. Park, D. U. Kang, J. Kim, and S. Y. Chun, "Multi-Temporal Recurrent Neural Networks for Progressive Non-uniform Single Image Deblurring with Incremental Temporal Training," in *Computer Vision – ECCV 2020*, vol. 12351, A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, Eds. Cham: Springer International Publishing, 2020, pp. 327–343. doi: 10.1007/978-3-030-58539-6_20.

[9] Y. Yuan, W. Su, and D. Ma, "Efficient Dynamic Scene Deblurring Using Spatially Variant Deconvolution Network With Optical Flow Guided Training," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 3552–3561. doi: 10.1109/CVPR42600.2020.00361.

[10] Y. Zhang, S. Y. Ma, X. Zhang, L. Li, W. H. Ip, and K. L. Yung, "EDGAN: motion deblurring algorithm based on enhanced generative adversarial networks," *J Supercomput*, vol. 76, no. 11, pp. 8922–8937, Nov. 2020, doi: 10.1007/s11227-020-03189-y.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[12] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking Coarse-to-Fine Approach in Single Image Deblurring." arXiv, Sep. 16, 2021. Accessed: Aug. 04, 2022. [Online]. Available: http://arxiv.org/abs/2108.05054

[13] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, "Denoising Prior Driven Deep Neural Network for Image Restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, Oct. 2019, doi: 10.1109/TPAMI.2018.2873610.

[14] Q. Zhou, M. Ding, and X. Zhang, "Image Deblurring Using Multi-Stream Bottom-Top-Bottom Attention Network and Global Information-Based Fusion and Reconstruction Network," *Sensors*, vol. 20, no. 13, p. 3724, Jul. 2020, doi: 10.3390/s20133724.