2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) | 978-1-7281-9772-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/RTEICT49044.2020.9315623

Motion Deblurring of Faces

Pooja Anand
Department of Electronics and Communication Engineering
National Institute of Technology Karnataka, Surathkal, Mangalore, India
pooja.anand12195@gmail.com

Sumam David S
Department of Electronics and Communication Engineering
National Institute of Technology Karnataka, Surathkal, Mangalore, India sumam@nitk.edu.in

K S Sudeep
Intel Technology India Pvt Ltd
Bangalore,Karnataka, India
sudeep.k.s@intel.com

***Abstract*— This paper evaluates learning-based data-driven models for deblurring of facial images. Existing algorithms for deblurring, when used for facial images, often fail to preserve the facial shape and identity information. The best available models, which are used for general-purpose image deblurring, are pre-trained using only facial images. The Peak Signal to Noise Ratio (PSNR) Structural Similarity Index Measure (SSIM) and Time to deblur single images are the key metrics used for evaluating the models and for finding the most efficient model for deblurring facial images. From the results, the observation is that even though the PSNR value for DeblurGANv2 model is the highest, the best trade off between PSNR, SSIM, Time to deblur and visual quality is seen in DeblurGAN model.**

Keywords— Image deblurring, Deep Learning, SSIM, PSNR.

I. INTRODUCTION

The field of computer vision has greatly benefited from advancements in machine learning and neural networks. Artificial intelligence has gained a lot of importance in our daily lives. Spanning from the simple data collection to task-specific robots, the application of artificial intelligence is vast. Machine learning allows systems to learn from input data and to improve its performance without much human intervention. Neural networks can be considered as a subfield of machine learning.

Image deblurring can be considered as a process of inverse convolution. Conventional deblur models of single image deblurring have gained attention in recent times since a similar reference of the sharp image is not always available. Motion blurred images occur in video analysis. Failure to de-convolute these images leads to loss of information. This study aims to find the best efficient model for deblurring single face images.

Various deep learning techniques are used for deblurring images. However, facial deblurring is not well-studied due to the inadequate availability of large facial image datasets. The types of blur that we usually see in images are motion blur and defocus. Motion blur is caused by small changes in the relative position of the camera, for example, camera shake, object movement, or a combination of both.

The objective of this paper is to find an efficient neural network based algorithm to recover the underlying true image from the blurred image. It is carried out by specifically pre-training the models using only facial images and then comparing these results with the corresponding results obtained from the models which are not specifically trained for facial images.

Let y be the blurred image generated by convolving the true image x (also referred to as the latent image) with the kernel k and additive noise n. The whole blurring process can be represented as in equation (1). The aim is to get the sharp image x from the blurred image y.

$$y = k * x + n \tag{1}$$

The models used in this study are:

- DeepDeblur: Fast one-step blurry face images restoration [4].
- DeblurGAN: Blind motion deblurring using conditional adversarial networks [2].
- DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better [3].

Here the focus is on deblurring facial images. Motion blur and synthetic blur are used in this study. Motion blur is obtained by averaging consecutive frames. The disadvantage of such an approach is that a huge number of frames are required for single face identity. Such a vast amount of data is not readily available, and this is the main reason why deblurring of faces has not been well-studied. For motion blur, we assume to have around L number of frames per face identity where L is input by the user. But there are several restrictions such as each frame should differ by at least one-pixel value, and there should be a motion flow in the whole set of frames. In short, the first and last frames should contain an overlap of information. Synthetic blur is obtained by convolving the true image with a pre-defined blur kernel [1].

The rest of this section gives a brief introduction to different models used for deblurring of face images. Section II describes the methods used to adapt these models specifically for facial deblurring. Section III gives an overview of the datasets and the platform used in this work. Section IV shows the output of each model, and performance comparison of the models with respect to the key metrics, and pretrained model outputs.

A. DeepDeblur: Fast one-step blurry face images restoration

It is a one step restoring method for deblurring images. The training of this model is based on equation (2). Here, F denotes the neural network function as in [4].

$$\arg\min_F ||F(y) - x|| \tag{2}$$

The basic network used is the inception module. When convolution layers are stacked one over the other for getting deep into the network, the training will take more time, and also, after a certain point, the accuracy saturates. The Inception module, as the name suggests, goes deeper. A simple example of the inception module is shown in Fig. 1, where the convolution of the three different filters is performed on the same input, and then the outputs are concatenated and sent to the next corresponding layer, thus aids in extracting more information.

978-1-7281-9772-2/20/$31.00 ©2020 IEEE

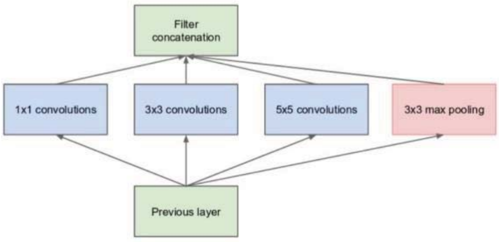

Fig 1. Inception Module [7]

As shown in Fig. 2, six inception modules are used in this network, with modules stacked one upon the other. The first layer has *3 x 3* convolutional kernels and outputs 64 channels. The last layer, which restores the output has *1 x 1* convolutional kernel which outputs 3 channels.

Extracting edge details and other long-distance movements from the blurred image is the basic network logic. Each inception block in Fig. 2 has five convolution kernel scales ranging from kernel size 1 to kernel size 14. To extract texture details smaller kernels of sizes *3 x 3* and *5 x 5* are used, whereas to extract abstract feature information, a larger kernel of size *14 x 14* is used. This network architecture helps in non-linear combining of feature maps and thus the deblurred image can be restored. The multi-scale convolutional network design using *1 x 1* kernel size helps in reducing the growth of model parameters which enables controlled training process and indeed computational complexity is reduced as cited in [4]. RMSProp optimizer is used with a learning rate of 0.001.

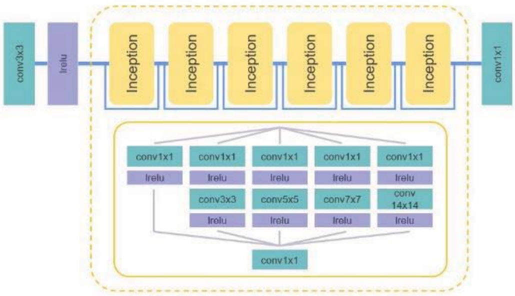

Fig 2. Architecture of DeepDeblur [4]

B. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks

As the name suggests, DeblurGAN uses conditional generative adversarial networks (GAN) to deblur single blurred images. Generative adversarial networks are deep neural networks consisting of generator and discriminator networks.

The main aim of the generator is to mimic the input data and pass it on to the discriminator. The discriminator works to distinguish between the real and fake data. A block schematic of GAN is shown in Fig. 3.

The proposed methodology used in this model is the initial training of the generator network, for each blurry image it estimates corresponding sharp image. Also, during the training phase, we introduce critic function for discriminator and train both networks in an adversarial manner.

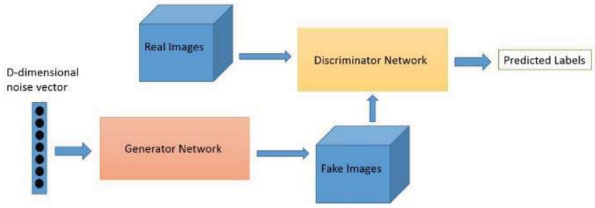

Fig 3. Generative Adversarial Networks [8]

The DeblurGAN network architecture shown in Fig. 4 contains two convolution blocks with stride 1/2, nine residual blocks (ResBlocks), and two transposed convolution blocks. The residual blocks have convolution layer, instance normalization layer, and ReLU activation. Also a dropout regularization with a 0.5 probability is added after the first convolution layer in each residual block. A global skip connection ResOut is also added to the network. Except for the last convolutional layer, all the convolutional layers are followed by an instance normalization layer and leaky ReLU layer with α = 0.2, Adam optimizer with the learning rate of 0.0001 for both generator and critic as cited in [2].

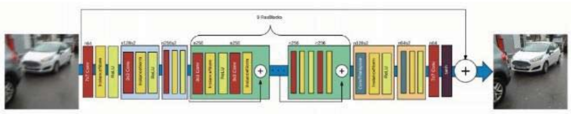

Fig 4. DeblurGAN Model [2]

C. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better

DeblurGAN-v2 is a new framework to showcase the improvement of the previous model DeblurGAN. This model focuses on improving two aspects - deblurring performance improvement and the inference efficiency. Feature pyramid networks (FPNs) are the basic building network for image deblurring. Upsampling layers and convolutional layers are also added to this model. Pretrained Imagenet model aids in deblurring, while the backbone logic is inception Resnets.

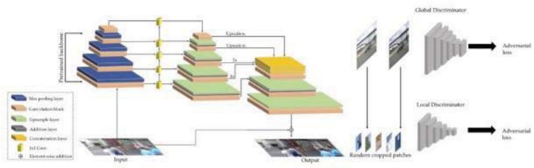

Fig 5. DeblurGANv2 model [3]

Figure 5 represents the entire architecture of the DeblurGAN v2 model. Here the Feature pyramid networks (FPN) architecture for feature extraction is used. This architecture helps in including multi-scale features. The deeper the network, the better quality of information is extracted, the same logic is used in FPN architecture thus aiding in extracting multiple feature maps for various semantics. FPN consists of both bottom-up and a top-down pathway. The bottom-up pathway is the usual convolutional network for feature extraction. As we go up, the semantic value of each layer is higher, and the spatial resolution gets downsampled. Reconstruction of higher spatial resolution from semantically rich layers occurs through the top-down pathway.

Five feature maps of different scales are taken from the feature pyramid network. Those features are later up-sampled to the same 1/4 input size and concatenated into one tensor which contains the semantic information on different levels. Two upsampling and convolutional layers are added at the end of the network to restore the original image size and to reduce other artifacts. A direct skip connection between input and output is also added to aid learning based on residue. *Tanh* activation layer is also used for output images to fall in the small range. Adam optimizer with the learning rate of 0.0001 is used [3].

II. Methodology

Efficient image deblurring methods should always aim to extract the finer details of the image. Edge prediction is also used to provide robust results in the case of synthetically blurred images, as it can be used to get the precise knowledge of the kernel used. As in any machine learning algorithms, datasets are prepared and preprocessed. Data retrieval includes feature extraction, feature scaling and selection. The next step is modeling, where model evaluation and tuning are done, loss functions are calculated and model is trained until the loss function between the expected and actual outputs is nearly zero or nil [5].

This paper uses three different models. As described in Section I, each model has different network architecture. Each model is pre-trained for deblurring of any image in general, but here the focus is on deblurring facial images. Each model is trained for facial images. Datasets used for the model and its details are mentioned in Section III. The pre-trained model for DeblurGAN [15] and DeblurGAN-v2 [16] are available and they have been used as a reference for comparison in section IV with our facial image model.

III. Implementation

Each model is trained in a similar end-to-end fashion. The training datasets used in this study are a mixture of synthetic blurred facial images and motion blurred facial images. Depending on the model, the batch size of training datasets is varied. The training dataset consist of pairs of blurred and sharp images. The whole dataset is split into training and testing datasets following the generic rule of 70:30 ratio. The datasets used in the study are shown in Table I.

TABLE I. Datasets

Sl. No.	Dataset	No. of Images	Resolution	Type of Blur
1	Youtubefaces [9]	3424	480 x 360	Motion
2	celebA [10]	2,02,602	178 x 218	Synthetic
3	Faces94,Faces95, Faces96, Grimace [11]	395	180 x 200	Motion
4	Labeled Faces in the Wild [12]	5713	250 x 250	Synthetic
5	https://youtu.be /Mz0918XdDew [13]	34	640 x 346	Motion

Youtubefaces, Faces94, Faces95, Faces96, Grimace and https://youtu.be/Mz0918XdDew are video datasets. These are converted to frames - each identity shown in these datasets has varying number of still image frames. For a single identity, the average of all the frames in the sequence is taken as the blurred image, and the middle frame in the sequence is taken as the corresponding sharp image. celebA and Labeled Faces in the Wild are still-image frame datasets, i.e., each identity has a single image. Each identity image is convolved with a blur kernel to obtain corresponding synthetically blurred images. Python programming language is used because Python language offers a wide range of user-friendly libraries like Keras, Pytorch and Tensorflow in the field of machine learning.

A. *Algorithm- Training Algorithm for Deblurring Facial Images.*

- Input: Set of blurred facial images and the corresponding sharp images.
- Let dataset A be the set of blurry input images and dataset B be the corresponding sharp images
- Each model is trained using the same set of synthetic blurred facial images and motion blurred facial images and the batch-size of training dataset is varied.
- Repeat: The training is repeated until the output of the model has approximately zero or nil Minimum Mean Squared Error (MMSE) with actual sharp output.
- Adjust the iterations or batch size to obtain the required results.

IV. Results

This paper gives an overview of three different approaches for deblurring of images. All models are trained using motion-blurred and synthetically-blurred facial images. The dataset used is a combination of celebA and YouTubefaces.

PSNR and time to deblur are the key metrics used in the evaluation.MMSE takes into account the pixel values between sharp image and model output and calculates the error between the two. The higher the PSNR value, the better is the model. Time to deblur is a measure to find which model can deblur the image fast. The shortest time to deblur is taken to consideration. SSIM is used for measuring image quality. Its value ranges between -1 and 1. SSIM is used to measure the similarity between the original sharp image and processed deblur output image. Value 1 indicates perfect similarity and value 0 indicates no similarity. The most efficient model is taken based on the trade-off between the PSNR value, Time to deblur, SSIM.

Table II shows the key metric measurement values for three images from the testing/inference dataset and these images are shown in Fig. 6. The key metric measurement values shown are obtained for the corresponding reference images.

Fig 6. Test Images 'a', 'b' , 'c'

Fig 7. Results of DeepDeblur model

Fig 8. Results of DeblurGAN model

The outputs of DeepDeblur, DeblurGAN and DeblurGANv2 are shown in Fig. 7, Fig. 8 and Fig. 9, respectively. The images are arranged in three columns in these figures: the first column has the sharp image, the second column has the synthetically blurred image and the third column shows the output from the corresponding model, i.e., the deblurred images.

Table II. Key Metric Measurements

Sl. No	*Model*	*Image No.*	*PSNR value*	*Time to Deblur (sec)*	*SSIM value*
1	DeepDeblur	a	28.244	3.82	0.804
2	DeepDeblur	b	28.453	3.74	0.647
3	DeepDeblur	c	30.496	3.98	0.765
4	DeblurGAN	a	29.913	4.90	0.901
5	DeblurGAN	b	29.737	3.96	0.758
6	DeblurGAN	c	29.187	6.01	0.796
7	DeblurGANv2	a	30.922	9.20	0.753
8	DeblurGANv2	b	33.341	7.08	0.789
9	DeblurGANv2	c	31.228	5.08	0.752

Fig 9. Results of DeblurGAN-v2 model

From the PSNR graph of DeblurGAN shown inFig. 13, we can seit can be observed that the 100th iteration plot has the highest PSNR value and the corresponding model output is shown in Fig. 10.

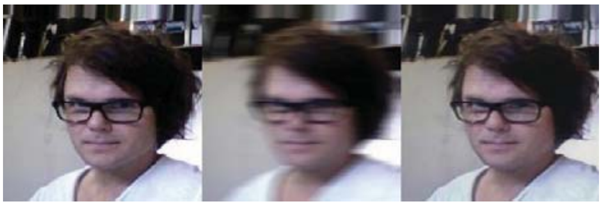

Fig 10. Results of DeblurGAN model - Iteration 100

The pretrained model reference outputs of DeblurGAN and DeblurGANv2 are shown in Fig. 11 and Fig. 12, respectively.

Fig 11. Results of Pretrained DeblurGAN model

Fig 12. Results of Pretrained DeblurGAN-v2 model

Comparison of the results shown in Fig. 8 and Fig. 11 reflects the improvement for the DeblurGAN model whereas Fig. 9 and Fig. 12 reflects the improvement shown in the DeblurGANv2 model. This improvement is solely because these models were pretrained for deblurring facial images alone.

Figure 13 indicate the PSNR values and loss function. PSNR values increases after each iteration whereas the loss function is decreasing for consecutive iterations.

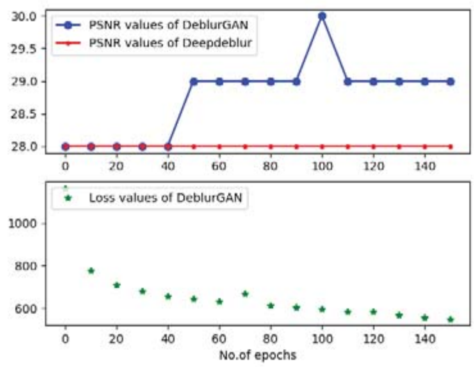

Fig 13. Plots of PSNR and Loss value

V. Conclusion

Previous models used in the field of image deblurring did not work well for facial images. To overcome this we took three models which were working for general purpose images and pre-trained these models for synthetically blurred and motion blurred facial images. We found that these models work well for facial images also.

For subjective raters, the model which gives us the best visual quality image output can be considered as the most efficient model. However, depending on the requirements all the three models are being preferred in various scenarios. Based on the evaluating the models on key metrics PSNR and Time to deblur single images and from the comparative study of the model output as shown in Figures 7, 8 and 9 we can conclude that even though the PSNR value for DeblurGANv2 model is the highest, the best trade off between PSNR, SSIM, Time to deblur and visual quality is seen in DeblurGAN model.

References

[1] Grigorios G. Chrysos, Paolo Favaro, and Stefanos Zafeiriou, "Motion deblurring of faces", International Journal of Computer Vision, 127(6-7), pp.801–823, 2019.

[2] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin,and Jiri Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks", In 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June, 2018.

[3] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang, " Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better", In 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, South Korea, pp. 8877-8886, 2019..

[4] Lingxiao Wang, Yali Li, and Shengjin Wang, "Deepdeblur: Fast one-stepblurry face images restoration", Computing Research Repository(CoRR), abs/1711.09515, 2017.

[5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio, "Generative adversarial nets", Annual Conference on Neural Information Processing Systems 2014, Montreal, Quebec, Canada, December, 2014.

[6] Tsung-Yi Lin, Piotr Dollár, Ross B. Girshick, Kaiming He, Bharath Hariharan, and Serge J. Belongie, "Feature pyramid networks for object detection", IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July, 2017.

[7] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich, "Going deeper with convolutions", IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June, 2015.

[8] J. Luo, W. Zhang, F. Xiang and J. Su, "Intention obfuscated adversarial deceptive path recommendation for UGV patrol maneuver", 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China, , pp. 206-211, 2019.

[9] Lior Wolf, Tal Hassner and Itay Maoz, "Face recognition in unconstrained videos with matched background similarity", IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, USA, 2011.

[10] Ziwei Lui, Ping Luo, Xiaogang Wang, Xiaoou Tang, "Deep learning face attributes in the wild", Proceedings of International Conference on Computer Vision (ICCV), Venice, Italy, December, 2017.

[11] D. Hond, L. Spacek , "Distinctive descriptions for face processing", Proceedings of the 8th British Machine Vision Conference BMVC97, Colchester, England, pp. 320-329, September 1997.

[12] Gary B. Huang, Manu Ramesh, Tamara Berg, and Erik Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments", University of Massachusetts, Amherst, Technical Report 07-49, October, 2007.

[13] G. G. Chrysos and S. Zafeiriou, "Deep face deblurring", 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, pp. 2015-2024, 2017.

[14] Stuart J. Russell and Peter Norvig, Foundations of ML and AI, Artificial intelligence - a modern approach, Prentice Hall, 1995.

[15] For DeblurGAN Pretrained model [Online]. URL https://github.com/RaphaelMeudec/deblur-gan

[16] For DeblurGAN-v2 Pretrained model [Online]. URL: https://github.com/VITA-Group/DeblurGANv2

[17] Skymind. [Online]. URL: https://towardsdatascience.com/a-simple-guide-to-the-versions-of-the-inception-network-7fc52b863202