Blur2Sharp: A GAN-based Model for Document Image Deblurring

Hala Neji $^{1,2,3}$, Mohamed Ben Halima $^2$, Tarek. M. Hamdani $^2$, Javier Nogueras-Iso $^3$, Adel M. Alimi $^2$

$^1$ University of Gabes, Tunisia
E-mail: hala.neji@ieee.org
$^2$ REGIM-Lab, University of Sfax, Tunisia
$^3$ I3A, University of Zaragoza, Spain

Abstract

In this paper, we propose a novel end-to-end model for document deblurring using cycle-consistent adversarial networks. The main objective of this work is to achieve image deblurring without knowledge of the blur kernel. Our method, named ‘Blur2Sharp CycleGAN’, generates a sharp image from a blurry one and shows how Cycle-Consistent Generative Adversarial Networks (CycleGAN) can be used in document deblurring. Using only a blurred image as input, we try to generate the sharp image. Thus, no information about the blur kernel is required. In the evaluation part, we use Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to compare the deblurring images. The experiments demonstrate a clear improvement in visual quality with respect to the state-of-the-art using a dataset of text images.

Keywords: Generative Adversarial Network (GAN), Cycle-Consistent Generative Adversarial Network (CycleGAN), Document deblurring, Blind deconvolution, Motion blur, Out-of-focus blur

1. Introduction

Thanks to mobile technology we are able to capture documents in a simple way and at any moment. Text documents such as bank documents, advertisements, courier receipts, hand-written notes, digitized receipts, public information signboards and information panels captured by portable cameras are very common in our daily lives. Portable cameras offer great conveniences for acquiring and memorizing information as they provide a new alternative for document acquisition in less constrained imaging environments than personal scanners. However, due to the variations in the imaging conditions as well as the target document type, there are many factors that can degrade the images such as image blurring, sharing distortions, geometrical warping or noise pollution. Frequently, the motion blur caused by camera shake and the out-of-focus blur can affect the quality of the images obtained by mobile devices. Although blur may be not very relevant at first sight, it may be the cause of problems in ulterior processing tasks such as text segmentation or Optical Character Recognition (OCR). Thus, traditional scanner-based OCR systems cannot be directly applied on camera-captured images and a new level of processing needs to be addressed.

Fig. 1. Examples of real source images with blur (right) and sharp images (left).
The objective of this work is to study a new model for removing different types of blur from real blurry document images and generate the correspondent sharp images. Fig. 1 illustrates the difference between source blur images (right) and sharp images (right).

Various deblurring techniques have been proposed so far based on blur kernel estimation: blind deconvolution methods, and non-blind deconvolution methods. In non-blind deconvolution methods, we have some knowledge about the blur kernel. In contrast, in blind deconvolution methods no information about the blur kernel is known. Blind deblurring estimates a latent sharp image and a kernel blur, namely Point Spread Function (PSF), from a blurred image. This problem has been deeply studied (see section 2). However, generic methods have not been effective with real world blurred images.

The proposal of this work is to eliminate blur to restore a blurred image that was captured via handheld camera or smartphones. We want to show how mobile phones can replace desktop scanners and ceaselessly improve mobile devices (image quality and computing power). For this purpose, we propose to extend a Cycle-Consistent Generative Adversarial Network (CycleGAN) for translating a blurred input text image into a sharp one. Recent methods based on Generative Adversarial Networks (GAN) for tasks such as image-to-image translation depend on the availability of training examples where the same data is available in both domains. However, CycleGAN is able to learn such pair information without one-to-one mapping between training data in source and target domains. The challenge of this work is to propose a new architecture based on CycleGAN, which we call ‘Blur2Sharp CycleGAN’, for the task of text document deblurring.

The rest of this work is organized as follows. Section 2 reviews the state of the art in blind deconvolution methods. Section 3 describes our suggested system. The evaluation and experimental results are provided in section 4. Finally, section 5 provides some concluding remarks.

2. State of the art

Given its wide range of applications, document image deblurring has attracted considerable attention in recent years, and various approaches have been proposed, overall in the field of blind deconvolution methods.

In blind deconvolution methods, the first task is to estimate the blur kernel. After this, the second part of the method is to restore the final latent image thanks to a non-blind deblurring algorithm. With the assumption that the blur is uniform and spatially invariant, the mathematical formulation of the blurry image can be modeled as $Y = X \ast K + \epsilon$, where $X$ is a latent sharp image, $K$ is an unknown blur kernel, $\ast$ denotes the convolution operator, $\epsilon$ is the additive noise and $Y$ is the blurred observation.

Chen et al. suggested an effective document image deblurring algorithm based on a Bayesian method that analyzes the local structure of a document image and uses an intensity probability density function as prior for deblurring. However, it is not generally applicable to blurred text images because it depends on text segmentation. Cho et al. proposed another Bayesian method that takes into account more specific properties of text documents. Pan et al. proposed another approach for text deblurring that makes profit of L0 regularized intensity and gradient priors. Nayef et al. suggested a method for document image deblurring that uses sparse representations improved by non local means. Zhang et al. used a gradient histogram preservation strategy for document image deblurring.

More recently, Ljubenovic et al. proposed the use of a dictionary-based prior for class-adapted blind deblurring of document images. Using a large set of motion blurred images with the associated ground-truth blur kernels, Pan et al. proposed a method to learn data fitting functions. Last, Jiang et al. proposed a method based on the two-tone prior for text image deblurring.

Convolutional Neural Networks (CNN) have been also applied to various image enhancement tasks. Focusing on image text image deblurring, Hradiš et al. proposed an end-to-end method to generate sharp images from blurred ones using CNN. Their network, consisting of 15 convolutional
layers, is trained on 3 classes (blur, sharp image and kernel blur), but there are some disadvantages: on the one hand, pairwise images are required to train the networks; on the other hand, the result of deblurring is not appropriate for natural images which have a color background. Pan et al.\textsuperscript{14} presented a genetic approach with two main ideas: modify the prior to assume that the dark channel of natural images is sparse instead of zero, and impose the sparsity for kernel estimation. They used CNN to predict the blur kernel.

In the past few years, Generative Adversarial Networks (GAN) have been used in different image-related applications\textsuperscript{3} such as generating synthetic data, style transfer, super-resolution, denoising, deblurring or text-to-image generation. Using this approach, Xu et al.\textsuperscript{16} proposed a method for jointly deblurring and super-resolving images that are typically degraded by out-of-focus blur. However, they focus on two classes of images: low-resolution blurred face and text images. Based on disentangled representations, Lu et al.\textsuperscript{9} presented an unsupervised method for domain-specific, single-image deblurring. Nimisha et al.\textsuperscript{11} also used GAN for end-to-end deblurring network. Using adversarial loss, the network learns a strong prior on the clean image domain and it maps the blurred image to its clean equivalent.

In summary, the previous approaches cannot be generalized to deal with different scenarios in document text deblurring. Therefore, it is a challenging task to count on a method based on deep learning for building a general image prior that is able to handle different scenarios.

3. Proposed Method

Image deblurring aims to restore a clean image from a blurred one to improve the result of OCR. As mentioned in the introduction, this work explores the feasibility of applying CycleGAN\textsuperscript{18} with a new architecture for the challenging task of cleaning blurred text images. We want to demonstrate that CycleGAN approach is suitable for the simpler task of text image deblurring where the results are good. Our idea is that if we can treat blur and sharpness as a kind of image style, successful image deblurring may be achieved with unpaired image dataset based on CycleGAN.

The main advantage of CycleGAN is that it does not require a pairwise image dataset. To solve the issue of not having a ground-truth document of every input, CycleGAN proposes a reverse mapping function $G_y$ and $G_x$. If a generator function $G_x$ translates a document from domain $X$ to domain $Y$ just changing the appearance, it should be possible to map the image back to domain $X$ with another generator function $G_y$ and reconstruct the initial image: $X \rightarrow G_x(X) \rightarrow G_y(G_x(X)) \approx X$. The functions could be also reversely applied, $Y \rightarrow G_y(Y) \rightarrow G_x(G_y(Y)) \approx Y$, and no ground-truth documents should be needed for deblurring. The model contains also the associated adversarial discriminators $D_y$ and $D_x$. $D_y$ encourages $G_x$ to translate $X$ into outputs indistinguishable from domain $Y$, and vice-versa for $D_x$ and $G_y$. To further regularize the mappings, we introduce two cycle consistency losses that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started. Fig. 2 shows a schema of this model that allows us to train our network with unpaired images to translate from blur to sharp or inversely.

Fig. 3 shows the training process of our proposed Blur2Sharp CycleGAN. It consists of two generators.
and two discriminators. The main goal of generators is to learn the mapping between two image domains. In addition, the two adversarial discriminators aim to distinguish between images and translated images in the same way as they aim to discriminate between the sharp images and the generated images. The following subsections explain the details of the generative and discriminator networks.

3.1. Generative network

As shown in Fig. 4, both generators consist of three parts: an encoder, a transformer and a decoder. This figure summarizes the architecture of these three parts, which has shown good performance for image deblurring.

**Encoder** The input of the network is [256, 256, 3]. There are 3 layers to extract the features from an image by one stride each time. In the first layer each filter is 7x7 and the stride is 1x1. The activation function is ReLU. The shape of the output is [256, 256, 64] (by padding). Then, the output of the first layer is passed to the following layers. The hyperparameters of the second layer are: 128 filters where each filter is 3x3 with stride 2; the activation function is ReLU; and the output is [128, 128, 128]. The hyperparameters of the third layer are: 256 filters where each filter is 3x3 with stride 2; the activation function is ReLU; and the output is [64, 64, 256].

**Transformer** The transformer aims to transform feature vectors of an image from one domain to another domain. The input is [64, 64, 256]. We started building our model from scratch. We started with 9 layers of ResNet blocks. However, as we did not obtained good results, we tried to change the model structure to improve the result. After testing different possibilities, we chose the hyperparameters and structures that best suited for the deblurring task. We have used 12 layers of ResNet blocks. All filters have 3x3 size, and the stride is 1. Therefore, the output of our model transformer is [64, 64, 256].

**Decoder** The decoding works in the opposite way to the encoder. We reconstruct the image from the feature vectors by applying three deconvolution layers which use the reversed parameters of the encoding step. Therefore, the output of our model decoder is [256,256,3].
3.2. **Discriminative network**

The main goal of discriminator is to extract features. It contains 5 convolutional layers. The hyperparameters of all the layers are the following: the total filters are \([64,128,256,5,1]\); the stride is \([2,2,2,2,1]\); and the activation function is LeakyReLU (slope=0.2) for all the layers except for the last layer with LeastSquareGAN. All the filters have 4\times4 size. The principal role of discriminator is to decide whether these features belong to one particular category or not by the last layer, which is a convolutional layer with output 1\times1.

3.3. **Loss function**

There are two components of the CycleGAN objective function: an adversarial loss and a cycle consistency loss. Both are essential to obtaining good results. We train the models using Least Square GAN loss. To achieve our goal, the loss function must satisfy that the discriminator \(X\) should be trained for images as close as possible to domain \(X\), and fake images close to 0. Thus, discriminator \(X\) would minimize \((\text{Discriminator}_X(x) - 1)^2\). In the same way, it should be able to distinguish original and generated images. The discriminator must return 0 for images generated by the generator \(X\). Discriminator \(X\) would minimize \((\text{Discriminator}_X(\text{Generator}_Y \rightarrow X(y)))^2\).

On the other hand, generator \(X\) should be able to deceive the discriminator \(Y\) about the authenticity of its generated images. This can be done if the recommendation by discriminator \(Y\) for the generated images is as close to 1 as possible. Therefore, generator \(X\) would like to minimize \((\text{Discriminator}_Y(\text{Generator}_X \rightarrow Y(x)) - 1)^2\). And the most important cycle loss captures that we are able to get the image back using another generator and thus the difference between the original image and the cyclic image, which is \([G_y(G_x(x)) - x] + [G_x(G_y(y)) - y]\), should be as small as possible.

4. **Experiments**

The following subsections describe the experiments that have been performed using the proposed ‘Blur2Sharp CycleGAN’ method: the corpus of documents for training and testing together with the implementation details (see section 4.1), and the results that have been obtained (see section 4.2).

4.1. **Datasets and implementation**

With respect to the unpaired dataset used for the experiments, we are using the dataset proposed by Hradiš et al.\(^4\). It consists of images with both defocus and motion blur generated by the random walk. We randomly cropped 2000 blurred image patches with 256\times256 size and 2001 cleaned image patches 256\times256 size from the dataset for training.

For the testing phase, depicted in Fig. 5, we used the test dataset proposed by Hradiš et al.\(^4\), which includes 100 pair of images (blurred images and their corresponding sharp versions) to the evaluate the quality of image restoration.
We have implemented our Blur2Sharp CycleGAN on document deblurring in Python with the help of Keras and running the code on top of Google Tensorflow framework. In addition, it must be noted that all the computation works were run on an Ubuntu server with NVIDIA Quadro P6000 GPUs.

4.2. Results

Fig. 6 shows two examples of the application of our ‘Blur2Sharp CycleGAN’ method on the test dataset. Although the visual result of sharp images seems satisfactory, it is difficult to distinguish visually a real image from a synthetic one when the differences between them are small. Therefore, we have compared quantitatively our method using two metrics: Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR).

The equation for SSIM is:

\[ I(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\mu_x^2 + \mu_y^2 + c_1}(\sigma_x^2 + \sigma_y^2 + c_2) \]

where \( I \) is the Structural Similarity Index; \((x, y)\) are coordinates indicating a nearby \( N \times N \) window; \( \sigma_x \), \( \sigma_y \) are the variances of intensities in \( x \), \( y \) directions; \( \sigma_{xy} \) is the covariance; and \( \mu_x \), \( \mu_y \) are the average intensities in \( x \), \( y \) directions.

The equation for PSNR is:

\[ \text{PSNR}(I_{ori}, I_{deblur}) = 10 \log_{10} \frac{255^2}{MSE} \]

where \( MSE \) stands for the Mean Square Error; \( I_{ori} \) is the original image; and \( I_{deblur} \) is the deblurred image. \( MSE \) is computed as follows:

\[ MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{ori}(i, j) - I_{deblur}(i, j)]^2 \]

where \( m, n \) is the size of the image.

Based on the two metrics defined above, we have compared the results obtained with our proposed Blur2Sharp CycleGAN architecture and the state-of-the-art methods using the test dataset previously described in section 4.1: the 100 images extracted from the dataset proposed by Hradiš et al.\(^4\).

The results are shown in Table 1. Our PSNR is equal to 32.52db and according to SSIM, we achieve 0.7689 on average using CycleGAN. Therefore, we can conclude that our ‘Blur2Sharp CycleGAN’ has achieved a comparable SSIM and PSNR with the advantage of not having to use pairwise images for the training phase.

We have also compared the quality of the OCR output obtained with the test dataset. After applying Tesseract software to obtain OCR from the sharp images in the test dataset (used as reference text) and the images returned by the methods, we have computed the average cosine similarity between the character frequency vectors derived from each pair of corresponding OCR output files. We have considered only the detection of letters and digits within the range of printable characters in ASCII encoding. As shown in table 2, our method performs better than the methods of Pan et al.\(^{13, 14}\) and a bit worse than the method of Hradiš et al.\(^4\). It is logical that the method of Hradiš et al. performs better because it is a non-blind deconvolution method. It must also be noticed that the obtained OCR is not perfect, even

Table 1. Quantitative comparison with state-of-the-art methods on the proposed image deblurring dataset

| Method          | PSNR | SSIM  |
|-----------------|------|-------|
| Pan et al.\(^{13}\) | 32.24| 0.6627|
| Hradiš et al.\(^4\) | 33.40| 0.9076|
| Pan et al.\(^{14}\) | 32.59| 0.6708|
| Xiu et al.\(^{10}\) | 20.12| 0.8970|
| Blur2Sharp CycleGAN | 32.52| 0.7689|

Based on the two metrics defined above, we have compared the results obtained with our proposed Blur2Sharp CycleGAN architecture and the state of the art methods using the test dataset previously described in section 4.1: the 100 images extracted from the dataset proposed by Hradiš et al.\(^4\). The results are shown in Table 1. Our PSNR is equal to 32.52db and according to SSIM, we achieve 0.7689 on average using CycleGAN. Therefore, we can conclude that our ‘Blur2Sharp CycleGAN’ has achieved a comparable SSIM and PSNR with the advantage of not having to use pairwise images for the training phase.
for the sharp images in the testing dataset, because we are using cropped images with incomplete words and lines.

Table 2. Quantitative comparison of OCR output

|                  | Average cosine similarity |
|------------------|---------------------------|
| Pan et al.¹³      | 0.679                     |
| Pan et al.¹⁴      | 0.645                     |
| Hradiš et al.⁴    | 0.897                     |
| Blur2Sharp CycleGAN | 0.872                   |

In addition, the results are good in terms of visual comparison with respect to the state of the art. For instance, Fig. 8 compares the results obtained with different methods on an input image that belongs to the text image deblurring dataset proposed by Hradiš et al.⁴. It can be observed that our Blur2Sharp CycleGAN method generates a sharp image with much clearer characters.

5. Conclusions

We presented a novel model that uses Cycle-consistent adversarial networks for document deblurring. Our proposed ‘Blur2Sharp CycleGAN’ architecture adjusts CycleGAN to the task of text image deblurring. We can both deblur images and blur sharp images because CycleGAN has the property of cycle-consistency. It is worth noting that since we use unpaired images as training dataset, we do not need the ground-truth sharp images. Based on our prior knowledge, successful image deblurring can be achieved with an unpaired image dataset using CycleGAN.

In addition, it must be noted that our model for image document deblurring obtains comparable results with respect to the current state of the art. Moreover, the obtained PSNR values are at the same level as the best methods found in the research literature. This demonstrates that the idea of treating blur and sharpness as a kind of image style actually works. In addition, a by-product of generating fake blurring images using CycleGAN is also provided. Last, it must be noted that our model significantly improved the speed of the deblurring process thanks to GPU acceleration.

As future work we want to test the feasibility of ‘Blur2Sharp CycleGAN’ for the problem of document denoising. This is specially relevant for the processing of historical documents in the first time of printing offices (e.g., incunabula documents in 15th and 16th centuries) and the correct application of Optical Character Recognition and other text analysis processes. The documents in this period present a wide range of noise derived from the use of hand-made typefaces, varied concentration of ink, and the logical degradation of paper.

Acknowledgements

The research leading to these results has been partially supported by the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48 and the Regional Government of Aragon (Spain, project T59_20R).

References

1. X. Chen, X. He, J. Yang, and Q. Wu. An effective document image deblurring algorithm. In CVPR 2011, pages 369–376. IEEE, 2011.
2. H. Cho, J. Wang, and S. Lee. Text image deblurring using text-specific properties. In European Conference on Computer Vision, pages 524–537. Springer, 2012.
3. B. Fu, L. Wang, R. Wang, S. Fu, F. Liu, and X. Liu. Underwater Image Restoration and Enhancement via Residual Two-Fold Attention Networks. International Journal of Computational Intelligence Systems, 14(1):84–95, 2021.
4. M. Hradiš, J. Kotera, P. Zemcik, and F. Šroubek. Convolutional neural networks for direct text deblurring. In Proceedings of BMVC, volume 10, page 2, 2015.
5. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 1125–1134, 2017.
6. X. Jiang, H. Yao, and S. Zhao. Text image deblurring via two-tone prior. Neurocomputing, 242:1–14, 2017.
7. Y. Liu, S. Liu, C. Li, and D. Yang. Compressed Sensing Image Reconstruction Based on Convolutional Neural Network. International Journal of Computational Intelligence Systems, 12(2):873–880, 2019.
8. M. Ljubenovic, L. Zhuang, and M. A. Figueiredo. Class-adapted blind deblurring of document images. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), volume 1, pages 721–726. IEEE, 2017.
Fig. 8. Example of the outputs obtained after applying different deblurring methods.
9. B. Lu, J.-C. Chen, and R. Chellappa. Unsupervised domain-specific deblurring via disentangled representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 10225–10234, 2019.

10. N. Nayef, P. Gomez-Krämer, and J.-M. Ogier. Deblurring of document images based on sparse representations enhanced by non-local means. In 2014 22nd International Conference on Pattern Recognition, pages 4441–4446. IEEE, 2014.

11. M. T. Nimisha, K. Sunil, and A. Rajagopalan. Unsupervised class-specific deblurring. In Proceedings of the European Conference on Computer Vision (ECCV), pages 353–369, 2018.

12. J. Pan, J. Dong, Y.-W. Tai, Z. Su, and M.-H. Yang. Learning discriminative data fitting functions for blind image deblurring. In Proceedings of the IEEE International Conference on Computer Vision, pages 1068–1076, 2017.

13. J. Pan, Z. Hu, Z. Su, and M.-H. Yang. Deblurring text images via L0-regularized intensity and gradient prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2901–2908, 2014.

14. J. Pan, D. Sun, H. Pfister, and M.-H. Yang. Blind image deblurring using dark channel prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1628–1636, 2016.

15. W. Wei, J. Yongbin, L. Yanhong, L. Ji, W. Xion, and Z. Tong. An advanced deep residual dense network (DRDN) approach for image super-resolution. International Journal of Computational Intelligence Systems, 12(2):1592–1601, 2019.

16. X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M.-H. Yang. Learning to super-resolve blurry face and text images. In Proceedings of the IEEE International Conference on Computer Vision, pages 251–260, 2017.

17. M. Zhang, C. Desrosiers, C. Zhang, and M. Cheriet. Effective document image deblurring via gradient histogram preservation. In 2015 IEEE International Conference on Image Processing (ICIP), pages 779–783. IEEE, 2015.

18. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 2223–2232, 2017.