DCGANs for image super-resolution, denoising and debluring

Qiaojing Yan
Stanford University
Electrical Engineering
qiaojing@stanford.edu

Wei Wang
Stanford University
Electrical Engineering
wwang23@stanford.edu

Abstract

Advance of computational power and big datasets brings the opportunity of using deep learning methods to do image processing. We used deep convolutional generative adversarial networks (DCGAN) to do various image processing tasks such as super-resolution, denoising and deconvolution. DCGAN allows us to use a single architecture to do different image processing tasks and achieve competitive PSNR scores. While the results of DCGAN shows slightly lower PSNR compared to traditional methods, images produced by DCGAN is more appealing when viewed by human. DCGAN can learn from big datasets and automatically add high-frequency details and features to images while traditional methods can't. The generator-discriminator architecture in DCGAN pushes it to generate more realistic and appealing images.

1. Introduction

1.1. Related Work

1.1.1 Image super-resolution

Deep learning methods had been tried on doing super-resolution (SR). Most of them use deep convolutional neural network to construct high-resolution image [6, 1, 7]. Ledig et al. proposed the use of GAN to do the super-resolution. The intuition behind this is that the discriminator allow the GAN model to generate images that looks authentic to human. While this may not mean it will give higher PSNR, the resulting image often appears more clear than other methods.

1.1.2 Image denoising

Deep learning methods had also been tried. The common method is to use stacked sparse denoising auto-encoder architecture to do denoising [11, 12].

1.1.3 Image deconvolution

A lot of researchers had tried to do deconvolution using convolutional neural network. One class of methods is to try to use deep learning to predict the parameter of the blur kernel [14, 10]. For example, Sun et al. used CNN to estimate the orientation and magnitude of motion blur. Another class of methods is to use deep CNN architecture to do deconvolution directly. Xu et al. argued that traditional CNN is not suitable for doing direct deconvolution [13]. Their reasoning is that deconvolution can be transformed into a convolutional operation with a kernal that is non-local. However, traditional CNN can only do convolution with local kernal. They proposed a new CNN architecture that can do convolution with non-local kernel. On the other hand, for specific dataset, traditional CNN had been proven to be useful. For example, Hradi et al. used CNN to do direct text deblurring [5].

1.2. Contribution

We propose the use of deep convolutional generative adversarial network (DCGAN) for both image denoising and image super-resolution. This model gives competitive results compared to non-deep-learning methods and can sometimes perform better.

We also analyzed the performance of the preformance of our model on different tasks and different datasets (human face and scene). For super-resolution, we compare the performance under different sampling rate.

2. Method

2.1. DCGAN

The high-level architecture of the deep learning model is to use a generative adversarial network (DCGAN) proposed by Goodfellow et al. [2]. The idea of DCGAN is to have a generator G, which is trained to generate the desired image from downsampled or noisy input, and a discriminator D, which is trained to discriminate between the original image and the generated image. The generator and discriminator

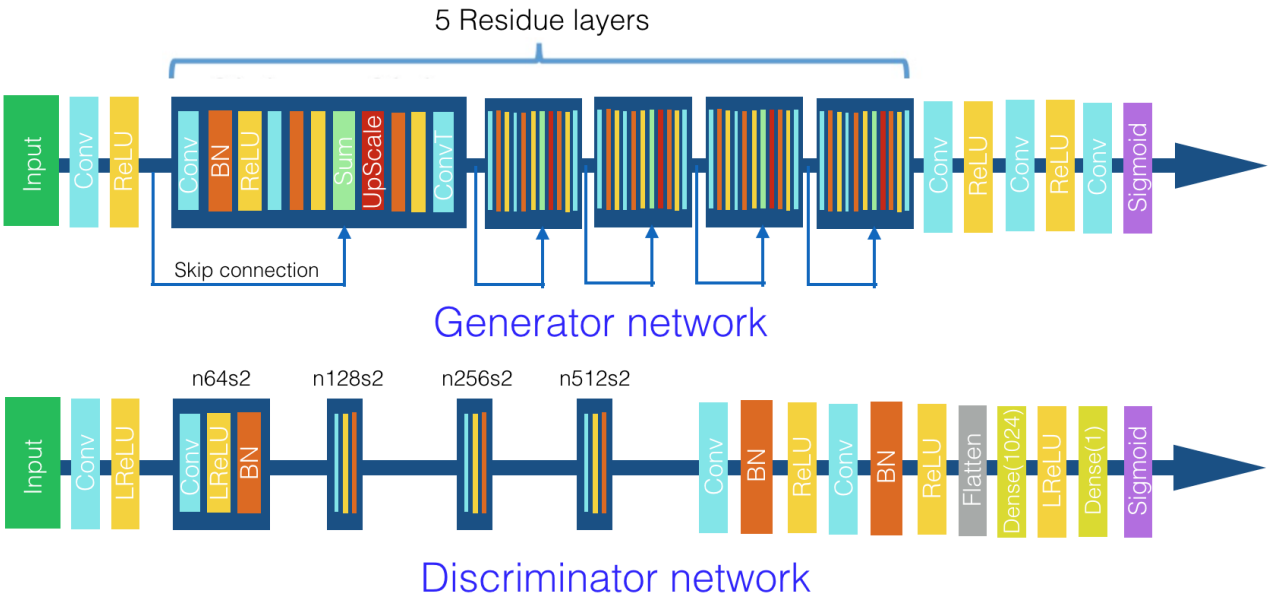

Figure 1: DCGAN image processing architecture.

are adversarial and they are trained together so that after training, generator would be good at generating images that looks authentic. As a simple case, this procedure can be expressed as

$$\min_G \max_D V(G,D) = \mathbb{E}_{I^{generated} \sim p_{train}(I^{original})}[\log D(I^{origin})] + \mathbb{E}_{I^{generated} \sim p_G(I^{input})}[\log(1 - D(G(I^{input}))] \quad (1)$$

On top of this simple model, we made some simple modification to the loss of generator and discriminator inspired by [8]. These will be discussed in section 2.1.2.

2.1.1 Generator and Discriminator

The architecture of the DCGAN is shown in Fig. 1. We used deep convolutional neural networks and deep residue networks [3, 4]. This architecture is based on the work of Ledig et al. [8] and we generalized it to do both super-resolution, denoising and deconvolution. The choice of parameters and sequence of layers are empirical.

In the generator network, we add an optional upscale layer between the ResNets. The upscale layer does a 2x scale-up of the image resolution.

By changing the input, we can let the DCGAN to learn to do different tasks. For example, for super-resolution, we feed in downsampled images and let generator network produce up-scaled images for comparison with the original image. For denoising, we feed in noisy images and for deconvolution, we feed in images blurred by gaussian kernel.

2.1.2 Loss Function

We modified the loss function in the vanilla GAN to better suit our model. The loss function of the generator is

$$l_G = 0.9 * l_{content} + 0.1 * l_{G,adv} \quad (2)$$

Here $l_{content}$ is the content loss between the generated image and the original image and we calculated as the $l1$ norm of their difference:

$$l_{content} = \|I^{generated} - I^{original}\|_1 \quad (3)$$

$l_{G,adv}$ is the adversarial loss and is the same as that in the vanilla GAN

$$l_{G,adv} = \sum_{n=1}^{N} -\log D(G(I^{input})) \quad (4)$$

The loss of the discriminator contains only adversarial loss

$$l_D = l_{D,adv} = \sum_{n=1}^{N} (\log D(G(I^{input})) + \log(1 - D(I^{original}))) \quad (5)$$

3. Experiments

3.1. Data set and evaluation measurements

We perform training and testing on two kinds of datasets:

- Large-scale CelebFaces Attributes (CelebA) Dataset [9] available here[1]. The dataset consists of 202,599 human faces. This dataset represents a narrow knowledge domain of human faces which we hope the DCGAN could learn.
- MIT Places Database [15] available here[2]. The dataset consists of 205 scene categories and 2.5 millions of images in total. The whole dataset's size is 132G. Because of time constraint and computation power available to us, we cannot train on this whole dataset. Instead, we only used the test set of the database for our training. It consists of 41,000 images. This dataset has much more variation that the face dataset and we want to see how DCGAN could perform on complex dataset.

We use PSNR to measure the similarities between the output of our methods with the original image.

3.2. Code and training details

Our code used TensorFlow library r0.12 and the code was adapted from the architecture written by David Garcia [3]. The training was done on AWS with GRID K520 GPU. We split our dataset into training set, dev set and test set. Hyper-parameters were trained with the training set and final evaluation was done on the test set. Each task took about 3 hours to train.

3.3. Experimental results

3.3.1 Super-resolution

First, we explored the performance of DCGAN model on single frame image super-resolution (SR). The performance of conventional bicubic interpolation is also included as a comparison. Fig. 2 shows the results of applying SR on human faces and on natural scenes test set, respectively. Table 1 shows the measured PSNR. We can see that DCGAN achieves slightly lower mean PSNR on the test set than the conventional bicubic method. However, if examining closely into the results of the DCGAN based super-resolution images (2c), we can see that, even with some distortion, DCGAN provides finer details on the resulting images, which actually agrees more with human recognizing conventions. The distortion comes from the compensation of common characteristics of human faces extracted by DCGAN model during training phase. In addition, DCGAN model provides a lower standard deviation on the test set. The reason is that many conventional image processing algorithms such as super-resolution, denoising, or deconvolution, only works well on images with certain characteristics. Therefore, various priors are developed and fed into the algorithms, which can boost the performance if known in advance. However it is usually difficult to extract those features blindly. On the other hand, DCGAN is able to minimize the loss function uniformly and doesn't make assumptions on the inputs, and hence providing lower STD. We can consistently observe similar effects on natural scenes as well as in the following discussions.

Table 1: Single frame super resolution result PSNR (2x).

	PSNR Mean (dB)	PSNR STD (dB)
Human Face bicubic	26.5124	2.0854
Human Face DCGAN	24.7346	1.4364
Natural Scene bicubic	23.4309	3.0286
Natural Scene DCGAN	21.7034	2.0999

Second, we tested the performance with 4x sub-sampling factor to measure the DCGAN based super-resolution tolerance. Fig. 3 and table 2 show the results on test set and PSNR, respectively. Since our original image is of size 64×64, the sub-sampled image with a 4x sub-sampling ratio is only of size 16×16 and most of the details are lost. In this case, the bicubic interpolation result is entirely blurry. But the DCGAN based super resolution result still manifests eyes, nose, mouth, etc. and complete with abundant facial details. On the other hand, we can see that the completed information is not highly accurate compared with the original image. This is maybe due to the fact that the sub-sampled image has too few pixel (information) to start with.

In [8] it is discovered that sub-sampling factor of 4 still works great under DCGAN framework for pixel-rich images. They used higher resolution inputs as dataset. Since the training process is computational intensive and limited by the hardware resources, we were unable to performing similar measurements for this type of input. In addition, it is interesting to notice that the natural scene super-resolution results seemingly look better than human faces in terms of visual appealing, this may not be that DCGAN works better for natural scenes but may be the fact that human brain is developed to recognize human faces more sophisticated.

3.3.2 Denoising

Next, the performance of DCGAN on image denoising is evaluated. The results of two conventional image denoising algorithm, median filter and non-local means (NLM), are

[1] http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html
[2] http://places.csail.mit.edu
[3] https://github.com/david-gpu/srez

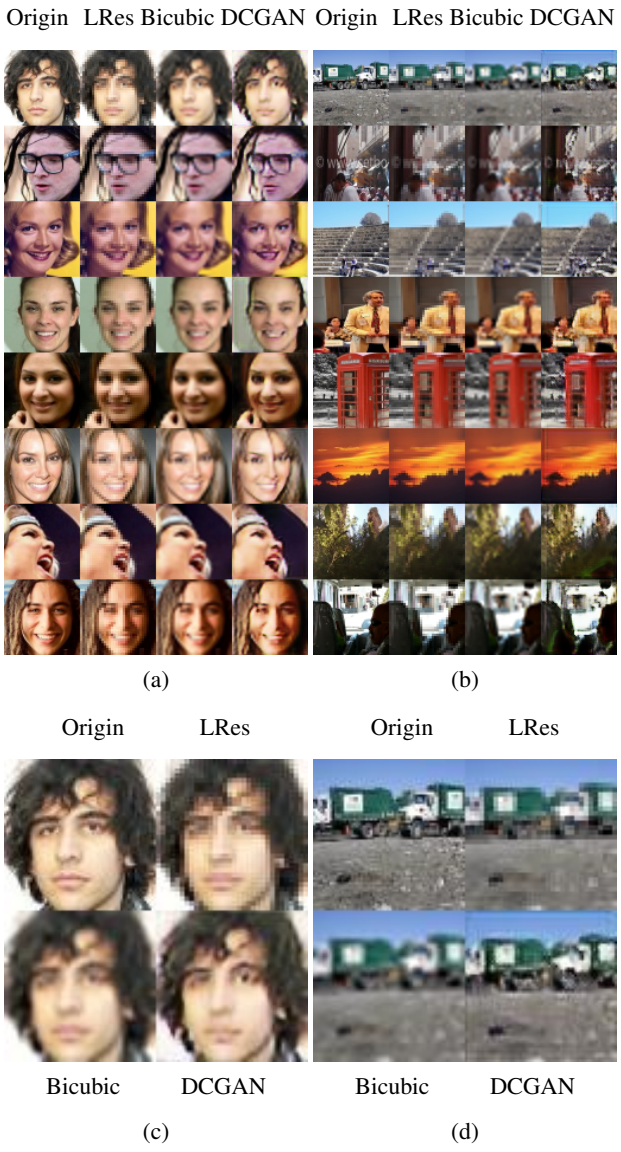

Figure 2: SR results, sub-sampling ratio = 2, (a): human face test result, from left to right are original image, sub-sampled image, bicubic interpolation result, and DCGAN based SR result, respectively, (b): natural scene test result (in same order), (c): one enlarged human sample result from left to right, top to bottom are original image, sub-sampled image, bicubic interpolation result, and DCGAN based SR result, respectively, (d): one enlarged natural scene sample result (in same order).

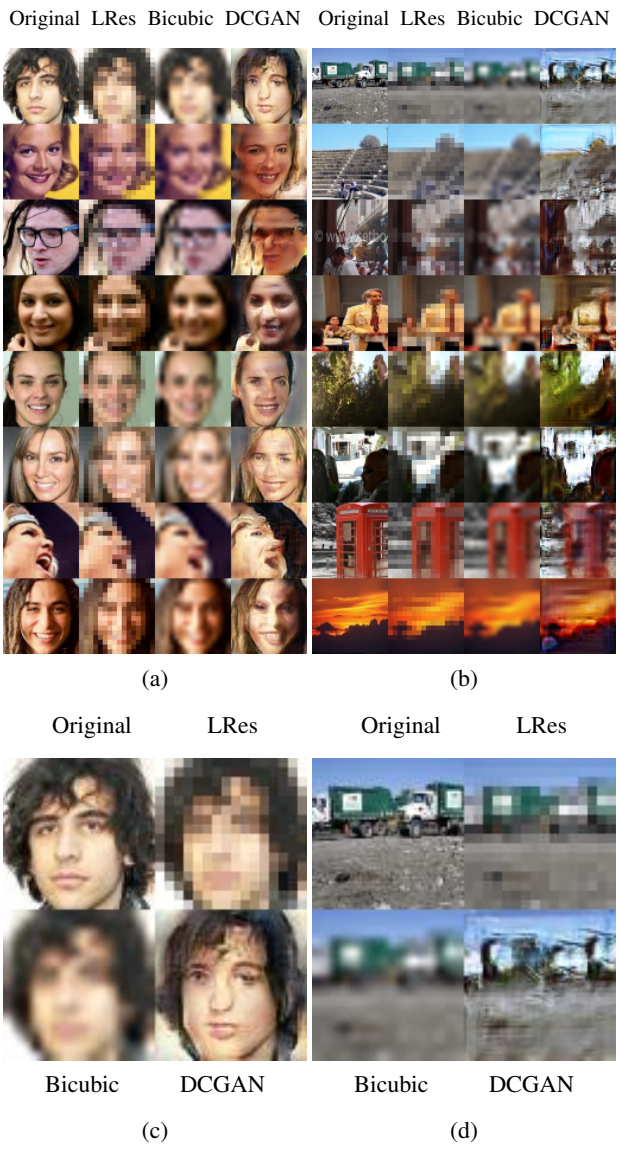

Figure 3: SR results, sub-sampling ratio = 4, (a): human face test result, from left to right are original image, sub-sampled image, bicubic interpolation result, and DCGAN based SR result, respectively, (b): natural scene test result (in same order), (c): one enlarged human sample result from left to right, top to bottom are original image, sub-sampled image, bicubic interpolation result, and DCGAN based SR result, respectively, (d): one enlarged natural scene sample result (in same order).

also included for comparison. Fig. 4 and 5 show the experimental results on human face and natural scene test sets, respectively. Table 3 shows the PSNR measurements for the denoising results. We can see that DCGAN based image denoising achieves similar PSNR as the NLM method, both outperform the median filter method. When checking closely on the resulting image, such as the eye areas from human face set or the grass image from natural scene

Table 2: Single frame super resolution result PSNR (4x).

	PSNR Mean (dB)	PSNR STD (dB)
Human Face bicubic	21.3604	1.5173
Human Face DCGAN	17.1314	1.8369
Natural Scene bicubic	20.2359	2.5090
Natural Scene DCGAN	16.6750	1.1949

set, DCGAN based denoising algorithm retains more details than NLM since NLM still performing average on local blocks with similar structures.

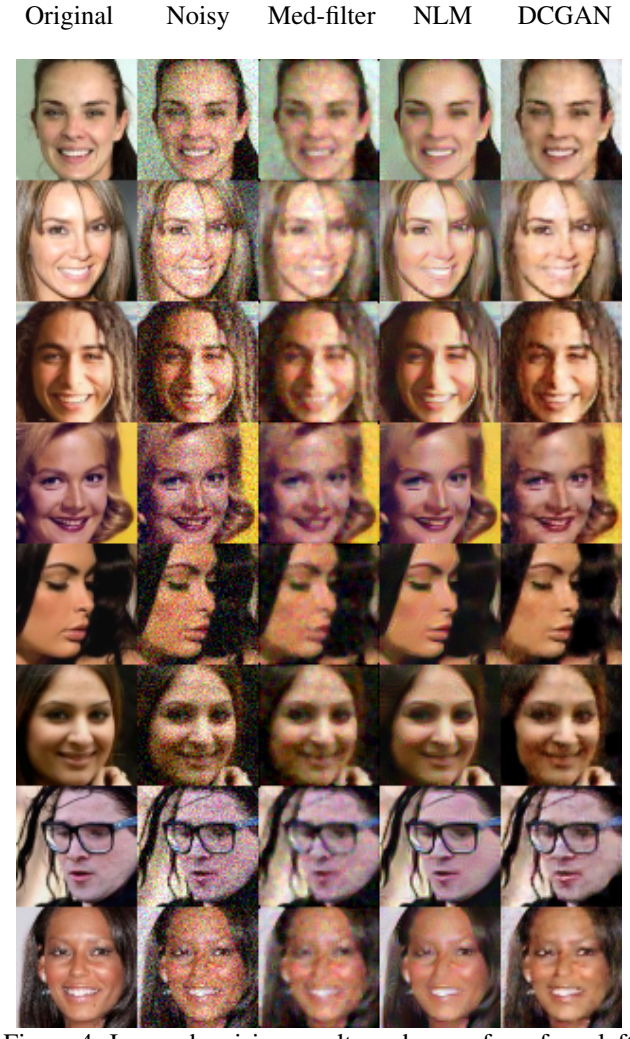

Figure 4: Image denoising results on human face, from left to right are original image, noisy image, median filter result, NLM result, and DCGAN based denoising result, respectively.

Figure 5: Image denoising results on natural scene, from left to right are original image, noisy image, median filter result, NLM result, and DCGAN based denoising result, respectively.

Table 3: Image denoising result PSNR.

	PSNR Mean (dB)	PSNR STD (dB)
Human Face Median	23.5595	1.2906
Human Face NLM	26.7011	0.9317
Human Face DCGAN	26.2448	0.9219
Natural Scene median	20.9344	1.2277
Natural Scene NLM	24.5626	1.6664
Natural Scene DCGAN	23.1454	0.7862

3.3.3 Deconvolution

Last, we evaluated the performance of DCGAN based image deconvolution. A 9×9 Gaussian kernel is used to

blur the image and an additive white Gaussian noise with a standard variance of 0.003 is added to the blurry image. Wiener filter and alternating direction method of multipliers (ADMM) with a sparse gradient prior algorithms are also included for comparisons. Fig. 6 and 7 show the experimental results on human face and natural scene test sets, respectively. Table 4 shows the PSNR measurements for the deconvolution results. We can observe that Wiener filter and ADMM algorithm give similar results, outperforming DCGAN based deconvolution method in terms of PSNR for both test sets. On the hand, DCGAN again recognizes facial details fairly well and deblurred the image with plenty of common facial characteristics. For the natural scene data set, we can see there is a relative large visual difference between the original image and DCGAN deblurred image. The reasons might be twofold. First, the blurry inputs lost too many details for DCGAN to fill in. Second, natural images have relatively complex structures and not so many common features as human face. The performance on natural scenes might be further improved by using training set with similar structures as the blurry image, which can be realized with a pre-filtering of the training set with locality sensitive hashing algorithm. Nevertheless, it is interesting to see that DCGAN is able to give a fairly reasonable deconvolution performance on human faces.

Table 4: Image deconvolution result PSNR.

	PSNR Mean (dB)	PSNR STD (dB)
Human Face Wiener	23.9268	1.7919
Human Face ADMM	22.2161	1.7101
Human Face DCGAN	18.5269	1.1820
Natural Scene Wiener	20.7702	1.5653
Natural Scene ADMM	19.4910	1.2663
Natural Scene DCGAN	16.3362	1.2368

Figure 6: Image deconvolution results on human faces, from left to right are original image, blurry image, Wiener filter result, ADMM result, and DCGAN based deconvolution result, respectively.

4. Discussion and future work

Compared to traditional image processing methods, DCGAN allows us to use a single architecture framework to achieve different objectives. We only need to modify the pre-processing phase and feed in different inputs to train the DCGAN.

In DCGAN, the competition between the generator and the discriminator push the generator to produce images that look more appealing. Because DCGAN can learn from big datasets, it can use trained features to produce images from inputs that lack certain information. For example, with extremely low-resolution human face images as input, DCGAN can complete facial details and produce human faces that look authentic.

For future work, one way to improve the results of DCGAN is to do training set categorization. Currently our work uses a mixed training image dataset with faces of different sexes, races and postures. The super-resolution result could potentially be improved with a characteristic specific training data. For example, when performing SR on a smiley face (or a profile) image, it would be advantageous to use training data set composed of such smiley faces (or profiles) so that the CNN engine could capture more categorized features.

Also, currently our work only discussed the proposed model on image SR and denoising separately. However, for real applications, we often have to deal with noisy low resolution images. With conventional interpolation and denois-

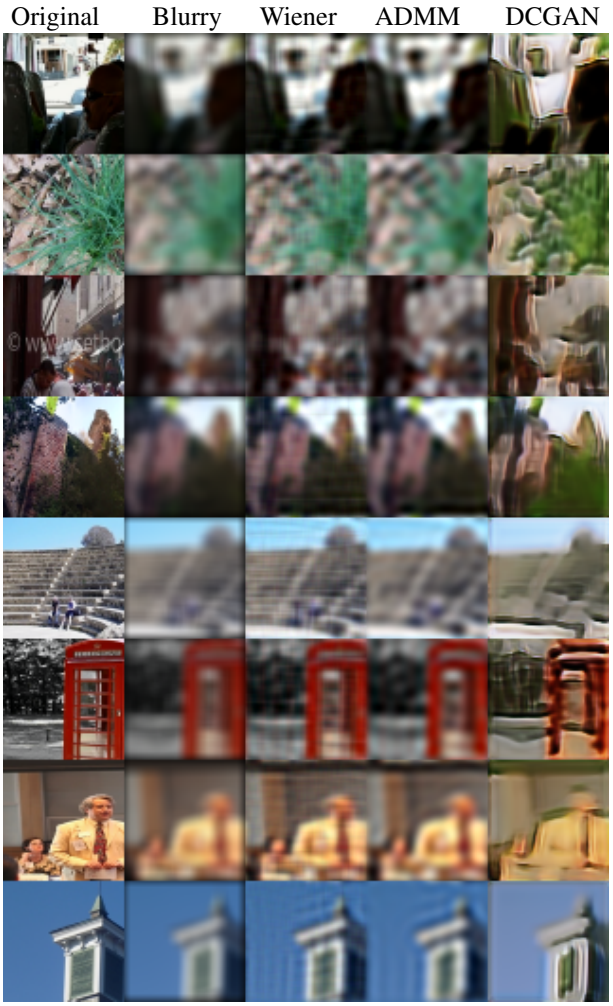

Figure 7: Image deconvolution results on natural scene, from left to right are original image, blurry image, Wiener filter result, ADMM result, and DCGAN based deconvolution result, respectively.

ing, both image processing methods would interfere with each other. Therefore, it might be great incentives to further investigate the combined SR and denoising effects on degraded images.

5. Conclusion

We proposed using DCGAN as a uniform architecture to perform image processing tasks and successfully tested for super-resolution, denoising and deconvolution. For super-resolution and denoising, the DCGAN gives competitive PSNR scores and can generate images that are more appealing compared to conventional methods. For deconvolution, DCGAN can give good results on human face dataset but it is not suitable to use for more complex dataset such as natural scenes.

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer, 2016.

[5] M. Hradiš, J. Kotera, P. Zemcík, and F. Šroubek. Convolutional neural networks for direct text deblurring. In *Proceedings of BMVC*, volume 10, 2015.

[6] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.

[7] J. Kim, J. Kwon Lee, and K. Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1645, 2016.

[8] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

[9] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[10] J. Sun, W. Cao, Z. Xu, and J. Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 769–777, 2015.

[11] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.

[12] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In *Advances in Neural Information Processing Systems*, pages 341–349, 2012.

[13] L. Xu, J. S. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems*, pages 1790–1798, 2014.

[14] R. Yan and L. Shao. Blind image blur estimation via deep learning. *IEEE Transactions on Image Processing*, 25(4):1910–1921, 2016.

[15] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*, pages 487–495, 2014.