GAN-D: Generative Adversarial Networks for Image Deconvolution

Ha Yeon Lee, Jin Myung Kwak, Byunghyun Ban, Seon Jin Na, Se Ra Lee, Heung-Kyu Lee✉
School of Computing and Department of Bio and Brain Engineering
Korea Advanced Institute of Science and Technology
Daejeon, Korea
{hayeon926, kwak.jinmyung, bhban, monokai, skawjdch, heunglee}@kaist.ac.kr

***Abstract*—We propose new generative adversarial networks for generalized image deconvolution, GAN-D. Most of the previous researches concentrate to specific sub-topic of image deconvolution or generative image deconvolution models with a strong assumption. However, our network restores visual data from distorted images applied multiple dominant degradation problems such as noise, blur, saturation, compression without any prior information. As a generator, we leverage convolutional neural networks based ODCNN [12] which perform generalized image deconvolution with a decent performance, and we use VGGNet [11] to distinguish true/fake of an input image as a discriminator. We devise the loss function of the generator of GAN-D which combines mean square error (MSE) of network output and ground-truth images to traditional adversarial loss of GAN. This loss function and the presence of discriminator reinforces the generator to produce more high-quality images than the original model structured with a single convolutional neural network. During experiments with four datasets, we find that our network has higher PSNR/SSIM values and qualitative results than ODCNN.**

Keywords—Generative Adversarial Nets(GAN); Deep Neural Network(DNN); Deconvolution; Deblurring; Desaturation;

I. Introduction

Many images suffer from several types of quality degradation problems caused by a camera noise, image compression, and intensity saturation. These image damaging processes can be modeled as a translation-invariant convolution. In contrast to quality degradation process, a deconvolution process is restoring distorted image produced by translation-invariant convolution back to the original image. Since eliminating degradations is the very important issue in the computer vision, many previous types of research [1-10] were concerned with handling the deconvolution problems. Deblurring [1,2,3], super-resolution [4,5,6], denoising [7,8], removing saturation [9,10] are the topics that are carried out in deconvolution problems. Recently, the deep neural networks have proven to have high accuracy and performance in the computer vision problems such as image classification [11] and image recognition. Thus, many types of research have attempted to solve deconvolution problems using deep neural networks [4,5,6,8].

However, handling all deconvolution issues at the same time can generate unexpected artifacts related to the images. Due to these difficulties, most of the research regarding to deconvolution are narrowed to sub-topic problem [1-10] or design generative models with very strong assumptions. In generative perspective studies; for example, the image priors are assumed to follow a Gaussian Mixture Model [15] or a hyper-Laplacian [16,17] that the real-world images do not follow or a noise model is assumed to follow a Poisson distribution.

Because of the difficulty of configuring all cases of degradation, few researches have attempted to restore visual data without any limitations regardless of the causes of the degradation problem that were applied to the distorted images. Especially, ODCNN [12] demonstrated a well-designed convolutional neural network is appropriate to handle multiple real-world degraded problems without any prior information that other previous method required.

Motivated by achievements of ODCNN, we proposed the advanced generalized deconvolution method, GAN-D by improving ODCNN structure as a novel generative adversarial network. Generative adversarial nets(GAN) is a powerful framework for generating high-quality images from adversarial relationships between a generator and a discriminator. We designed the network structure and appropriate parameters empirically through various experiments. For the generator, we transform ODCNN and for the discriminator, we adapt VGGNet. But due to the memory issue, we modified VGGNet. We also propose an objective function for image deconvolution.

Our main contributions are:

1) Our model address multiple major degradation problems such as blur, saturation, compression, and noise at the same time without any mathematical assumptions or prior knowledge.

2) To our knowledge, the proposed method is the first attempt to design GAN framework for the generalized deconvolution problem. Therefore, we propose the new objective function of GAN combining MSE which is traditional objective function to the original adversarial loss function [14]. Addition to the generator, augmenting discriminator allows to produce more natural high-quality images than the ODCNN.

3) Our proposed method achieved the qualitative and quantitative performance improvement through the structural redesign of the CNN-based network to the GAN framework. Our proposed method has higher PSNR about 4.84dB than the baseline model ODCNN and 5.46dB than total variation(TV) regularization[20].

Section 2 introduces recent representative deconvolution researches, and prior studies of our network components such as GAN which is the main framework of our model, ODCNN, VGG nets. Section 3 describes the architecture of GAN-D. Section 4 shows PSNR and SSIM comparison experiments with ODCNN and TV regularization for the benchmark dataset Set5 [21], Set14 [22], BSD100 [23] and the dataset used in ODCNN. The paper concludes with a discussion and future work in Section4 and brief remarks in Section 5.

II. Related Work

Several prior proposals have addressed image deconvolution problem, and they can be categorized into two types. The first one is solving partial problems such as deblurring, super-resolution, denoising [1-10]. The other one is attempting to solve various degradation problems at once in generative perspective [12]. Since removing various artifacts contained in images is difficult, the most of the approaches focus on getting rid of specific artifacts. VDSR [5] and Beyond a Gaussian Denoiser [8] (BGD) are state of the art in super-resolution which is one of the domain in deconvolution. They have contributed to improving the accuracy and performance of existing models by adding some deep learning techniques. VDSR increases the depth of deep learning model and BGD uses batch normalization to adjust input of layers. Both of them use the deep neural network and train the models using residual learning technique. By using residual learning, they reduce the training time.

In contrast to attempts to solve partial problems, some studies have attempted to solve multiple degradation problems at once. To handle multiple degradation problems, approaches using artificial neural networks tried to obtain sophisticated inverse functions for deconvolution and to expand the pre-existing technology to improve accuracy and performance. ODCNN [12] is the first approach using the deep neural network that successfully performed deconvolution for any degraded images without assuming any prior information. ODCNN reinterprets the pseudo inverse kernel of the classical Wiener filter and designs corresponding a novel neural network structure. Our model also tries to solve the problem like ODCNN, but the main difference between our model and ODCNN is that we use the generative adversarial nets (GAN) [14] to solve the problem and make the network denser.

To improve the performance of the deconvolution, we have built the GAN [14] specialized framework for enhancing the generative model through minimax game in the formulation (2) between the generative model and the discriminative model. The objective of GAN is to get the optimal state of the generative models which the generated distribution is same as the true data distribution described as (1) so the discriminator could not distinguish the difference between the real one and the sample generated by the generator. Goodfellow et al. [14] proves that solving equation (2) match with minimizing the Jensen-Shannon divergence between two distributions theoretically leads to converge to optimal state. In practice, the optimal state is achieved by updating the parameters of the generator and the discriminator alternatively for solving equation (2).

Karen Simonyan et al. [11] evaluate the performance of deep convolutional networks with small filters (3x3) for large-scale image recognition according to increasing depth. They propose best-performing ConvNet models, VGGNets with 16-19 weight layers and show these networks have outperformed the performance on the ImageNet Challenge 2014. VGGNets can be leveraged for the various computer vision problems. Our proposed method apply VGGNet to the discriminator of our network.

III. The GAN-D networks

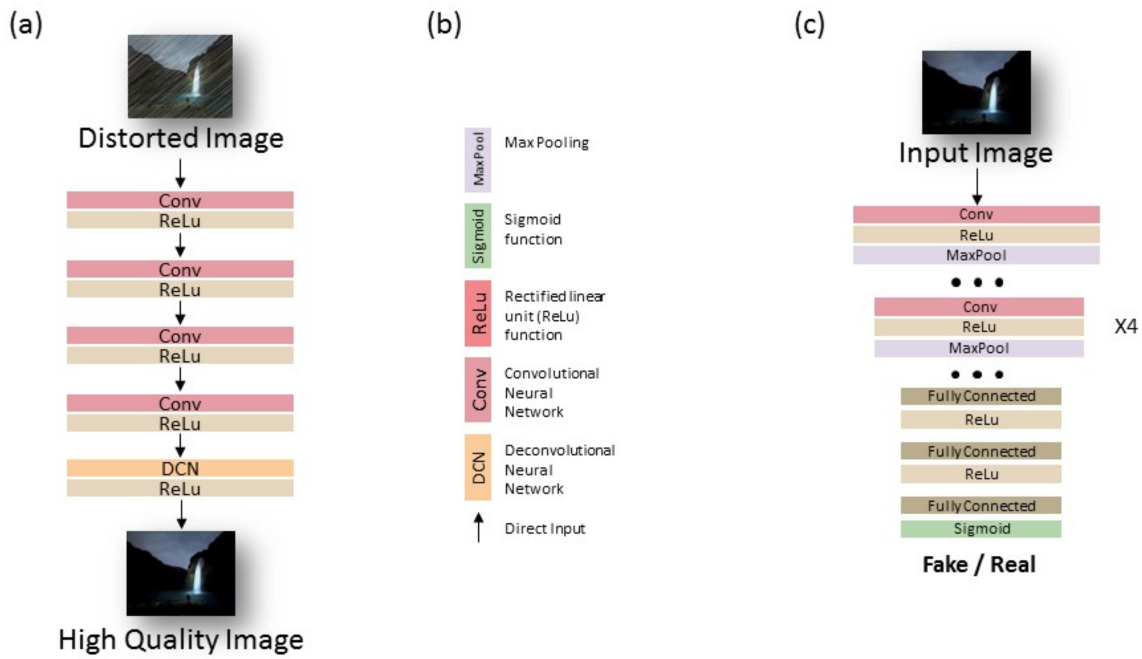
The basic structure of GAN-D networks is similar to that of other generative adversarial networks: a generator produces fake data which is originally far from the real data; a discriminator receives fake or real data to distinguish its originality. Both generator and discriminator are trained alternatively: generator gets better ability to deceive the discriminator as if its output is a real one; discriminator develops its discernment in distinguishing the difference between fake and real data. However, there are two fundamental differences. Firstly, unlike other generative adversarial networks for data augmentation or multiplication which tries to generate real-like data mainly depending on the reconstruction of possible features, our generator is trained to accomplish the specific task, and the discriminator evaluates the quality of work by pass/fail. Secondly, the generator has its objective function for the specific task while the objective function for the whole generative adversarial network also has terms for generator parameters. In other words, the generator itself is trained to be skillful in the specific task while the min-max game also promotes generator's ability as well.

A. Network Architecture

Let $\mathrm{G}(\mathrm{z}, \theta_t)$ be function enacted by the generator network at time-step t. The input distorted image for the generator is z. θ_t is generator parameter at time step t. $\mathrm{D}(\mathrm{x}, \emptyset_t)$ is discriminator function at time-step t. The input for the discriminator to be distinguished as fake/real is y-direction and parameter for discriminator at time step t is $\emptyset_t$. Both θ_t and $\emptyset_t$ are updated by Adam optimizer. [24]

At each time-step t, the generator receives low-quality input image x to return $\mathbf{G}(\mathbf{z}, \boldsymbol{\theta}_t)$ which is also an image. Then the discriminator receives an input image x, which might be high-quality original image of x (real) or output of generator $\mathbf{G}(\mathbf{z}, \boldsymbol{\theta}_t)$ (fake) to return a scalar value between (0, 1): it is the probability for given data to be real data; 1 means 100% sure to be real data and 0 means 100% sure to be fake data.

The network structure of generator is as same as [12]. It has four layers of convolutional layers with a deconvolutional neural network layer. The first layer has 1 x 121 sized kernel and the second layer has 121 x 1 sized kernel. This two layers corresponds 121 x 121 sized single kernel but can have one



[Figure 1] **(a) Generator**. The generator takes the distorted/blurred image as input. After some convolutional neural networks without pooling, the representation becomes high-level abstract features. Deconvolutional neural network layer reconstructs a high-quality image from the features by combining them. The generator itself showed nice performance in image deconvolution problem. [12] **(b) Notations**. The solid arrow means information stream. For example, arrows between layers mean forward propagation. **(c) Discriminator**. Discriminator takes RGB image as input. After some Conv-ReLu-MaxPool layers followed by three layers of fully connected neural networks, it returns scalar value, whose range is limited between 0 and 1 by sigmoid function. It is shortened form of VGGNet. [13]

additional ReLu activation function. The third layer has 16 x 16 sized 38 kernels, followed by 1 x 1 sized 512 kernels of the convolutional network. The last layer is the deconvolutional neural network, which has 8 x 8 sized single kernel to build up a single RGB image. One thing different from the Xu, Li's work and ours is zero padding: Xu, Li's DCNN contracts originally 184 x 184 sized image into 56 x 56 sized smaller image because of convolutional neural network's kernel stride problem; we applied zero padding techniques outside the figure so that the output of generator can maintain original resolution.

As we considered discrimination of fake/real image as image classification problem, we applied VGGNet [11] structure as discriminator which showed outstanding performance in image classification at 2014 ILSVRC. However, the structure is shortened into five layers of Conv-ReLu-MaxPool and three layers of fully connected networks to reduce memory consumption. Each layer has 2 x 2 sized small kernel with both x, y-direction stride 2. The CNN layers have 64, 128, 256, 512, 512 kernels in order. The FCN layers have 4096, 4096, 100 neurons and the last layer has single neuron whose activation function is sigmoid. Each CNN layer is followed by 2x2 max pooling operation with strides in x, y-direction size 2.

B. Objective function

The optimal state of the generative adversarial network is when the generator output data seems like to be so realistic that the discriminator cannot tell the originality at all. In a mathematical word, this optimal stage is described as:

$$p_{data} = p_G \quad (1).$$

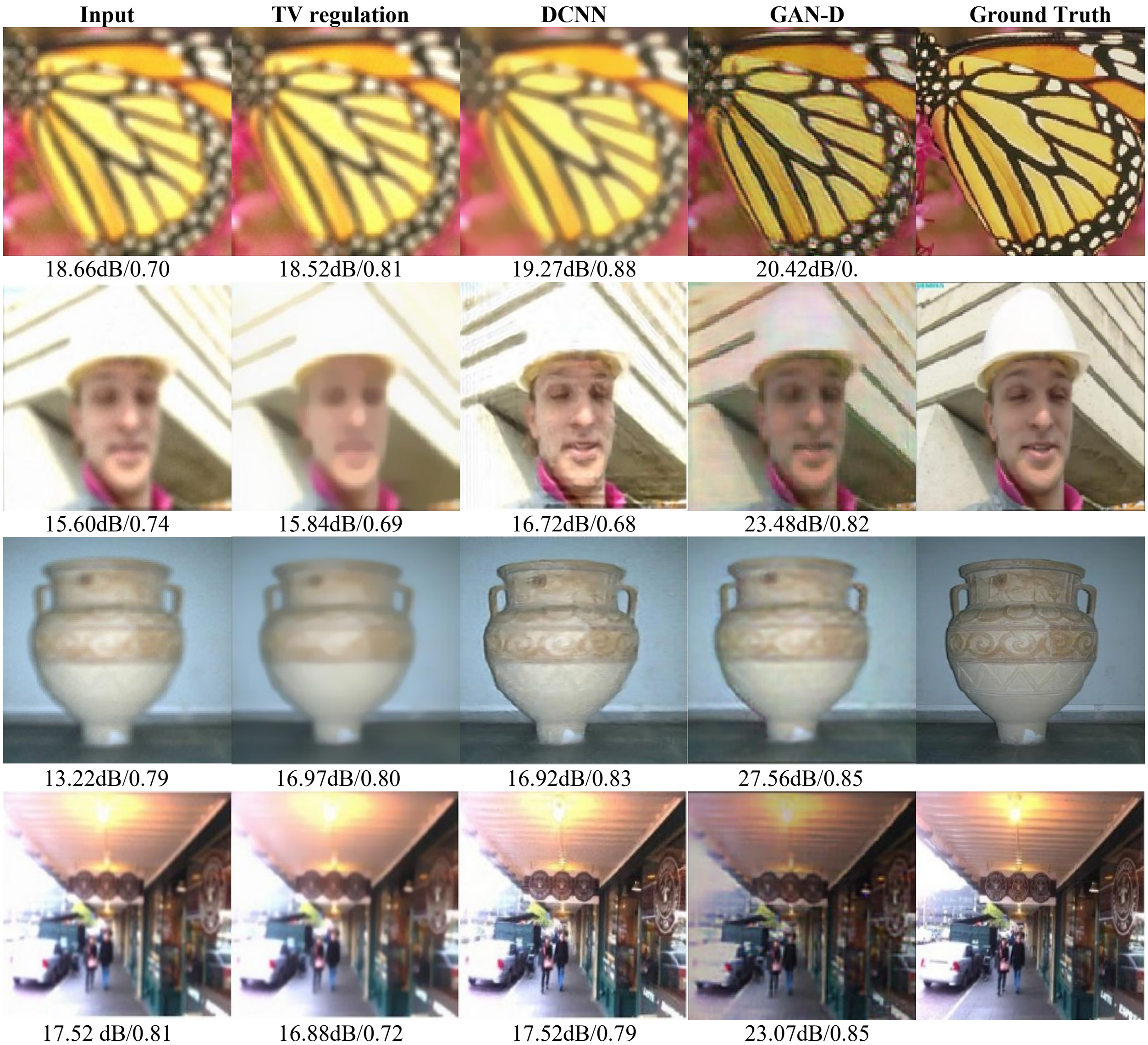
p_{data} is the distribution of real data while p_G is the distribution of generator output. This convergence the state is acquired by updating parameters for conventional GAN objective function:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \quad (2). [14]$$

Where $p_z(z)$ is input variables. As our model is not for data augmentation or replication but for specific task, image deconvolution, we modified the objective function for image deconvolution problem:

$$\begin{cases} \mathcal{L}_D = -\log D(x) - \log(1 - D(G(z))) \\ \mathcal{L}_G = MSE(G(z), x) \end{cases} \quad (3).$$

$\mathcal{L}_D$ and $\mathcal{L}_G$ are loss functions for the discriminator and generator. MSE means mean squared error. x is original high-quality image and z is distorted image, generated by applying filter on x.



[Figure 2] From left to right: total variation regulation, deep convolution neural network for deconvolution, and generative adversarial network for deconvolution, original high resolution image. Corresponding PSNR and SSIM are placed below each image. From top to bottom: each example from Set5, Set14, BSD100, and DCNN dataset.

Generator loss itself means supervised learning of the generator to modify degraded image into high-quality image. As probability distribution is bound (0, 1), the loss function of discriminator is maximized when difference between two probability distribution $D(x)$ and $D\big(G(z)\big)$ is 0 and is minimized when the difference is 1 (maximum); minimizing this function make the discriminator to tell the difference between ground truth high-quality image and generator output in better accuracy.

Reducing generator loss on optimal state leads the probability distribution of $D(x)$ and $D\big(G(z)\big)$ to be similar enough so the discriminator cannot tell the difference between them.

While traditional image deconvolution method considers minimizing error of the network output and ground-truth high-resolution image and traditional generative adversarial network modify internal variable in a way to maximize difference between probability distribution $D(x)$ and $D\big(G(z)\big)$ expecting the generator to produce more realistic image, our model

[TABLE1] Comparison of TV regularization, DCNN, GAN-D, and the original HR on benchmark and DCNN data. Highest PSNR(dB) and SSIM in bold.

Set5	Input	TV regularization	ODCNN	GAN-D	Ground Truth
PSNR	17.15	16.978	18.14	**24.392**	∞
SSIM	0.69	0.77	**0.838**	0.81	1
Set14					
PSNR	16.70	16.05	16.90	**21.67**	∞
SSIM	0.73	0.65	0.72	**0.77**	1
BSD100					
PSNR	16.99	17.38	17.48	**25.41**	∞
SSIM	0.70	0.73	0.74	**0.81**	1
ODCNN dataset					
PSNR	18.58	18.19	18.51	**18.91**	∞
SSIM	**0.86**	0.79	0.78	0.79	1
All					
PSNR	17.35	17.14	17.76	**22.60**	∞
SSIM	0.74	0.74	0.77	**0.80**	1

combined both objectives at the same time; generator directly updates its internal variable in a way to reduce error between x and $G(z)$ first and then the GAN based approach updates internal variable in a way to produce more realistic data.

IV. EXPERIMENTS

A. Data and similarity measure

We perform experiments on three benchmark datasets Set5[21], Set14[19], and BSD100[23]. Additional to the public datasets, we also conduct experiments on ODCNN dataset available at the project webpage [25]. All experiments are performed with the images following the degradation specification defined by Xu et al. [12]. They combined 4 different types of degradation to lower the quality of the images. For the experiments, we proceeded the 4 types of degradation into 4 steps. We first blurred the images using circular averaging filter with the diameter length of 7. Second, we clipped the intensity of the images by multiplying 1.3. Third, we added Gaussian noise with mean and variance assigned to 0 and 0.0005. Lastly, we compressed the images to remain only 70% quality of the images.

For fair comparison, we resized all images of train and validation datasets to 128x128 and measured peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), which represent quantitative scales of the quality of the images. To validate the our PSNR and SSIM results, we reference the PSNR and SSIM results of TV regularization and ODCNN. The source codes of TV regularization and ODCNN were obtained from MathWorks and the project website [25].

B. Training details and parameters

To evaluate our model, we implemented with Python 3.x and TensorFlow 1.2 library. We also activated CUDA 8.0 as the backend of TensorFlow to accelerate training. We used 4 GTX 1080 with 8 GB memory devices independently for training the model.

For the training dataset we have collected 2500 images downloaded from the Google search engine and sampled around two million patches from it. All the images are downgraded to low quality images stated above. The generator GAN-D accepts the 128x128 low quality images and returns the deconvoluted images which are same width and height as the input. Our input values range in between [0, 255]. Since we only use ReLu as the activation function in both networks, discriminator and generator, we did not scale the range of the image datasets.

For optimization, we use Adam [24] with $\beta_1 = 0.9$ with a learning rate of 10^{-5} and updated the learning rate when validation loss dropped to certain range of loss. We alternatively trained the generator and discriminator.

C. Performance of the model

We compared the performance of GAN-D to TV regularization and ODCNN. TABLE1 shows the summary of the quantitative results of each approach. On average, our GAN-D has highest PSNR/SSIM among other two approaches in all three benchmark datasets and ODCNN dataset. The examples in Figure 2 depict that our results surpass the visual restoration of the features and colors of degraded images than the TV regularization and ODCNN. Our model reproduces the low quality image without creating irrelevant artifact like ringing artifact. and recovers the intensity back to the original saturation as expected.

V. Discussion and Future work

Unlike previous work focusing on particular type of degradation problems, our model handles multiple degradation problem such as blur, saturation, compression, and noise altogether. We have achieved high quality in reconstructing the distorted image compared to the original ODCNN model and traditional method. Our model is capable of capturing key features and colors of the images. It works well for complex degradations removal. We found that GAN drives the reconstruction towards the original image which produces perceptually more convincing results than CNN based deconvolution.

However, the loss function of GAN-D is MSE-based pixel wise averaging solution. Our model is limited in producing finer edge details of the images. Even though we have high PSNR and SSIM outcome, the image quality is not human friendly. As the future work, we suggest in modifying pixel-wise loss function to high abstracted features loss function that considers texture details of the image. Additional experiment is needed for measuring the quality of the images other than PSNR and SSIM.

VI. Conclusion

We have described a generative adversarial network GAN-D that outperformed the ODCNN on public benchmark datasets. For evaluation, we compared GAN-D and ODCNN with widely used PSNR and SSIM measures. We have discussed some limitation in deconvolution and introduced GAN-D. Since GAN has strong characteristic of reconstructing the image towards the natural image among other networks, we built the deconvolution model based on GAN. We have confirmed visually and quantitatively that GAN-D reconstructions produced more high quality deconvolution results compared to the reference methods.

Acknowledgment

This work was supported by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No.2015-0-00233), Managerial Technology Development and Digital Contents Security of 3D Printing based on Micro Licensing Technology)

References

[1] Krishnan, Dilip, and Rob Fergus. "Fast image deconvolution using hyper-Laplacian priors." Advances in Neural Information Processing Systems(NIPS). 2009.

[2] Schmidt, Uwe, et al. "Discriminative non-blind deblurring." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2013.

[3] Pan, Jinshan, et al. "Blind image deblurring using dark channel prior." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2016.

[4] Dong, Chao, et al. "Image super-resolution using deep convolutional networks." IEEE transactions on pattern analysis and machine intelligence(TPAMI) 38.2 (2016): 295-307.

[5] Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2016.

[6] Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network." Computer Vision and Pattern Recognition (CVPR), 2017

[7] Elad, Michael, and Michal Aharon. "Image denoising via sparse and redundant representations over learned dictionaries." IEEE Transactions on Image processing (TIP) 15.12 (2006): 3736-3745.

[8] Zhang, Kai, et al. "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising." IEEE Transactions on Image Processing (TIP) (2017).

[9] Cho, Sunghyun, Jue Wang, and Seungyong Lee. "Handling outliers in non-blind image deconvolution." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[10] Whyte, Oliver, Josef Sivic, and Andrew Zisserman. "Deblurring shaken and partially saturated images." International journal of computer vision 110.2 (2014): 185-201.

[11] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems (NIPS). 2012.

[12] Xu, Li, et al. "Deep convolutional neural network for image deconvolution."Advances in Neural Information Processing Systems (NIPS). 2014.

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR), 2015.

[14] Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in neural information processing systems*. 2014.

[15] Zoran, Daniel, and Yair Weiss. "From learning models of natural image patches to whole image restoration." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[16] Krishnan, Dilip, and Rob Fergus. "Fast image deconvolution using hyper-Laplacian priors." Advances in Neural Information Processing Systems. 2009.

[17] Levin, Anat, et al. "Image and depth from a conventional camera with a coded aperture." ACM transactions on graphics (TOG) 26.3 (2007):70.

[18] Bevilacqua, Marco, et al. "Low-complexity single-image super-resolution based on nonnegative neighbor embedding." (2012): 135-1.

[19] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In Curves and Surfaces, pages 711–730. Springer, 2012

[20] Chambolle, Antonin. "An algorithm for total variation minimization and applications." Journal of Mathematical imaging and vision 20.1 (2004): 89-97.

[21] Bevilacqua, Marco, et al. "Low-complexity single-image super-resolution based on nonnegative neighbor embedding." (2012): 135-1.

[22] Zeyde, Roman, Michael Elad, and Matan Protter. "On single image scale-up using sparse-representations." International conference on curves and surfaces. Springer, Berlin, Heidelberg, 2010.

[23] Martin, David, et al. "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics." Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on. Vol. 2. IEEE, 2001.

[24] Kingma, Diederik, and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).