An Image Denoising Method Based on Deep Residual GAN

Ziyuan Wang\textsuperscript{1,2}, Lidan Wang\textsuperscript{1,2,3,4,5,*}, Shukai Duan\textsuperscript{1,3,4,5,6} and Yunfei Li\textsuperscript{1,2}

\textsuperscript{1} Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, Chongqing 400715, China
\textsuperscript{2} School of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
\textsuperscript{3} College of Artificial Intelligence, Southwest University, Chongqing 400715, China
\textsuperscript{4} Braininspired Computing & Intelligent Control of Chongqing Key Lab, Chongqing 400715, China
\textsuperscript{5} National & Local Joint Engineering Laboratory of Intelligent Transmission and Control Technology, Chongqing 400715, China
\textsuperscript{6} Chongqing Brain Science Collaborative Innovation Center, Chongqing 400715, China

*Corresponding author’s e-mail: ldwang@swu.edu.cn, wzy416234750@163.com

Abstract. As people come into contact with image data more often, high quality and clear images attract more attention. Many methods have been proposed to deal with image noise problem including deep learning (DL). However most of them is lack of capability when customers want more perceptual details of the image without information loss. In this paper, a deep residual network based on generative adversarial (GAN) network was proposed to complete the image denoising mission. Firstly, a generative-adversarial network structure based on residual blocks was designed. Secondly, a refined loss function was given to train the GAN network. The well designed loss function can help the generated image to be very close to the clear counterpart (ground truth) while enhancing more details in colours and brightness. Finally, extensive experiments show that our network is not only convincing for images denoising, but also effective for other image process tasks, such as image defogging, medical CT denoising etc., presenting impressive and competitive effects.

1. Introduction

With the popularity of electronic products, image data has become increasingly easy for customers to access, thus high quality and clear images stored in mobile phones and other devices get more and more important, therefore the task of turning the noising low-definition images to clear ones is significant, which attracted great attention of researchers in related fields. Based on this reality, many researchers have focused their attention on how to process images to get high quality counterpart. Denoising and high definition reconstruction of low quality digital image has been one of the appealing topics in recent years.

There are tremendous ways to get clear pictures, which are generally divided into two categories. One is from hardware perspective, namely the acquisition of the image, such as well-designed camera lens and related hardware circuits, which works well but takes longer to develop and costs more.
Consequently many researchers turn to the second method: thinking about the problem from the perspective of image processing algorithm. This method is relatively easy to implement, since once the source data is available, researchers can process it by algorithm programming to get clear image with the help of existing powerful computer processing capabilities.

Till now, many algorithms have been proposed for image denoising. Traditional methods like Gaussian filtering, spatial pixel feature denoising algorithm, bilateral filtering etc. can deal with many basic noise removing tasks [1], but they are not able to offer more details if proper parameters are not given. However, it’s not easy to get those parameters by manually searching.

Recently, with the progress of deep learning, many image related works can be done by it. The method based on multilayer perceptron can deal with image noise effectively, but the heavy training parameters limited it’s widely use. The most impressive one is the deep convolutional based networks such as the ideas in [2]. These model achieves clear image by stacking more convolutional layers to extract features. But the problems appear if asking for more specific details. On the one hand, the training time is longer due to the deepening of the network, on the other hand, the image obtained by this method may lost content details when compared to the original clear image labels, presenting unpleasant visual experience. After propose of generative adversarial network (GAN) [3], many researches have made a lot of progress on image generating, image debluring and super-resolution, many of which have achieved satisfactory results.

In this paper, we propose a deep residual network structure based on GAN networks for image denoising. The main part of this method is to design a generating network structure based on residual blocks. This network can generate a high definition image by taking the noised image as input, then the real clear image and the generated clear image are sent to a discrimination network to obtain corresponding scores. The key of the network is to design a loss function to train the GAN network. We define a loss function to make the generated image to be closer to a clear image while retaining more colour and content details. Extensive experiments prove that the trained network are not only effective for low-definition images in the test set, but also effective for other images that are not in the range, such as image defogging, Medical CT image denoising, motion deblurring, etc.

The rest of this paper is organized as the following order: In section II, some related works are briefly discussed; Section III gives the design principles and methods of our network; the IV part defines a modified loss function; the experiment process and the final conclusion are given in V and VI respectively.

2. Related work

Traditional algorithms like mean filtering, Gaussian filtering, and bilateral filtering can perform some basic processing on images, such as edge extraction, sharpening, and image denoising. Although these algorithms can enhance the image quality to a certain extent, meanwhile remove part of the noise, they are not able to perform well if customers need more details, colours and brightness, so that the images can be vivid in a highly photorealistic degree.

Since the advent of the artificial neural networks, many problems can be solved by applying it. For example, for feature extraction process of image classification problems, the input can be mapped into the multilayer perceptron (MLP) and then output a vector containing the features of the input image. Based on the specific applications researchers can determine whether to do classification, regression, or some other processes. In [4], the authors used multi-layer perceptrons to perform image denoising task. In that paper, they considered the problem as a regression problem. They sent the blurred-clear image pairs to the network for training. The error was back-propagated as a loss function, so that the network training converges.

Due to the massive parameters training mode of the MLP network, it takes very long time and consumes a lot of resources to finish the learning and training process while the final result will not be excellent enough. Therefore, the researchers proposed the convolutional neural network with fewer parameters. This kind of network can implement the effective feature extraction of the original data, and can be stacked more layers in this network. Seungjun Nah et al. proposed a neural network based on
deep convolution for deblurring of dynamic scenes [5]. In their research, the authors proposed a method of image synthesis of measured dynamic scenes due to the difficulty of obtaining measured blurred-clear image pairs for training in existing deblurring algorithms and the difficulty of capturing the blurring kernel of local images, etc. The end-to-end image deblurring algorithm based on deep learning is used to study the clear image restoration on the gopro dataset. Relatively prominent results are provided in their experiment.

Convolutional neural networks have many benefits when used to deep model training, but as the network deepens, feature losses gets more serious. Too many layers may cause gradient disappearance or gradient explosion during the learning process. Although the possibility of gradient disappearance can be alleviated through the selection of proper activation functions and some other methods such as dropout, initialization parameters, etc., it is still not possible to set the network layers for too many. To deal with this problem, Kaiming He, Xiangyu Zhang and others proposed a residual network [6]. They added mapping and jump connections on the basis of traditional convolutional networks, and also added a jump connection to each two layers to form a residual block, which can prevent gradient disappearance or explosion during the learning process, meanwhile retain the features of the input data of previous layer.

With propose of generative and adversarial network (GAN) by Goodfellow et al. In 2016, a lot of related works based on which have been done. Some researchers have done image style transfer studies such as cycleGAN proposed by Bair et al. [7], in which they took it as an image-to-image translation problem. Orest Kupyn et al. proposed DeblurGAN for blind motion deblurring [8]. There are also some researchers who have done research about single image super-resolution, such as photo-Realistic Single Image Super-Resolution using a Generative AdversarialNetwork (SRGAN) proposed by Christian Ledig et al.[9][10][11], in which, the author designed a deep residual network based on GAN and proposed a loss function for the network training. In their training results, not only the resolution of the image was raised, but also the content and colour of the original image were enhanced and many details were retained, due to the well-designed loss function.

Different from the previous loss function of deep convolutional networks, their loss function contains the perceptual loss of the image, which consists of content loss and generation loss. The generation loss comes from the difference between the real image and the fake image generated by the generation network. The purpose of the content loss is to solve the problem of structure similarity of high-level abstract features. A well-trained VGG16 is used as a feature extractor to extract the features of the real target image and the newly generated image for comparison, which is regarded as one of the content loss terms.

Our research is mainly inspired by SRGAN [9]. Firstly, we improved the network and added some new content on it. Though the network structure of convolution-residual block-deconvolution architecture is used, our residual network structure is deeper, reaching a depth of 8 layers, and the jump connection is also redefined. Secondly, we redesigned a new loss function to train the network, which includes not only perceptual loss, namely content loss and generative loss, but also a smooth loss function that can effectively prevent the checkerboard artefacts and halo artifacts. Thirdly, the network and related parameters were validated on different dataset through massive experiments.

3. Method

The generative adversarial network (GAN) is a two player network composed of a generator and a discriminator, whose idea is that the generator generates an indistinguishable distribution of fake data, then sends both the real data and the generated data to the discriminator network to be judged. The output will be true when the score is close to 1, and fake when 0. If the network is trained well enough, the final score should be 0.5 for both real and fake data distribution.

The loss produced by the discriminator is mainly derived from the score errors generated by the distribution of true and false data when compared to corresponding labels. The loss of the generator network is derived from the loss of the identification part of the generator. The training process of the
network is the process of minimizing and maximizing these two loss functions. The mathematical expression can be written as follows:

\[
\min_G \max_{D \in \mathcal{D}} E[D(x)] - E[D(\tilde{x})]
\]

Where \(x\) represents the distribution of real data, whose discriminative score is expected to be as large as possible. \(G_z\) represents the distribution of the generated data, and the score is expected to be smaller when cast to the same discriminator. Based on the original GAN, we redesigned the generative network, and added residual blocks to the network, then defined a loss function that fits the purpose of our image denoising task. The details are as follows:

3.1. Generator network
Generator network is used for generating indistinguishable images so that the discriminator can not distinguish from the real images. Noised images are used as input to generate images that can be comparable to real high-definition image labels, which was expect to surpass the labels in details and colours.

To this end, we designed a Convolution-Residual-Deconvolution network architecture. During the whole process, the shape of images remained unchanged, and jump connections between the residual networks and convolution parts were added. A connection is also added to the deconvolution part near the output end, which can ensure that all features of our input image can be preserved and enhanced. In this network, we stack the residual networks to 8 layers. The network model of the generator is shown in Figure 1. Experiments show that the residual network can retain more details. You can stack the residual blocks more only if you want. 8 layers can make the training result amazing in our experiment.

3.2. Adversarial network
Adversarial network was designed to give score for the generated image and the real image respectively, which can be regard as a regression problem. If the network is smart enough namely well trained, it will give a score close to 1 for the real label image and 0 for generated image, which indicates that the network has a strong discrimination ability and can effectively distinguish true and fake images. In our experiment, the image labels are clear images, and the fake ones are images obtained by the noised images after passing through the generator. The network structure of the discriminator is relatively simple, which consists of a series of convolutional layers, then connects to the full link layer, and finally was sent to the sigmoid function to normalize the confidence score to a probability between 0 and 1. The structure of the network is shown in Figure 2.
3.3. Training process
The training process of the network is similar to the training of the original GAN. It is divided into training of generative network and adversarial network. This is an alternating process. At the same time, the discriminator part of the generative network and the discrimination network share the same parameters. To prevent the disappearance or explosion of gradients, batch normalization was added to each of convolutional layers and residual blocks for both generative and discriminator network, except for the last layer.

4. Refined loss function
In the original GAN network, the input of the generator is random noise, and the goal is to generate a data distribution consistent with the target distribution, therefore the loss function used there is mainly composed of two parts: the generative loss and the discriminative loss. However, in our image denoising task, in addition to generate a distribution similar to the target image, more details such as texture details and brightness, colours need to be kept and enhanced to some degree, so we designed a loss function with more detail enhancement for the generation loss, feature loss, and local smoothing loss which considers the loss of local similarity of two images, the loss function can be summarized as follows:

$$L_{\text{total}} = \lambda_p L_{\text{pixel}} + \lambda_d L_{\text{disc}} + \lambda_f L_{\text{feature}} + \lambda_s L_{\text{smooth}}$$  \hspace{1cm} (2)

Where $L_{\text{pixel}}$ represents the pixel loss, which is the $L_2$ loss between the generated and ground truth images. $L_{\text{disc}}$ is the loss of generator. $L_{\text{feature}}$ means feature loss coming from VGG16 when taking generated and clear images as input. $L_{\text{smooth}}$ is smooth loss. $\lambda_p, \lambda_d, \lambda_f,$ and $\lambda_s$ are corresponding coefficients, which can be set manually.

5. Experiment
This section introduces the training process and details, and also give some of the training parameters used in our experiment. During the whole training, the ImageNet data was used to train the designed model. 3000 of the ImageNet dataset images were picked out to build our dataset. These images were cropped to 256x256x3 and then Gaussian noise were added to make the noised-clear pairs. 2000 of them were used as training set and the rest 1000 as test. The experiment result is shown in Figure 3. In addition, a self-made data set containing 1000 medical radiation CT images were used to train the model and 100 as test set. In the CT images training process, due to the small amount of data, we used 50 pairs of noised and clear image pairs as validation set, which makes the network more credible, Figure 4 (a) gives the CT denoising experiment result, which shows an impressive perceptual effect. Figure 4 (b) shows our model was robust when applying on image defogging since the image was in the dataset. As shown in Figure 5, the PSNR grows very fast before 5K iteration and PSNR of image shape 256x256 goes faster than other two image shapes. It has slower growth after 16K iterations. The final PSNR of our network reaches at 30.58 and SSIM is 0.9027.

Before the ImageNet dataset were sent to the network to train, we cropped them to a shape of 256x256x3 so that the whole training process would not take too long, while in the training of the CT dataset the shape size of the original medical image was retained to 384x384x1. The noised images of the CT dataset are obtained by adding Gaussian noise with various parameters to the high-definition images. Our model training is completed on the NVIDIA GTX1080Ti GPU based on tensorflow framework. The training batch is 5 and the total number of training iterations is 20K. Some of the parameters were set before train: $\lambda_p = 0.7$, $\lambda_d = 1.0$, $\lambda_f = 1.0$, and $\lambda_s = 0.0003$. The convolution kernel size of the first layer of convolution layers in generator is 9x9, the stride is 1, the second layer is set to 5x5, and the third layer 3x3 with strides of 1, respectively. The residual blocks are stacked to a total number of 8 layers, and the deconvolution in the generation network is used twice. In the residual blocks, the summation of the previous and present output are sent to next layer as input, which ensures that the main information is retained and can further be enhanced.
The structure of the adversarial network is relatively simple. It is mainly a stack of convolutional layers, and a full connection to 512 neurons then to one node, and finally was normalized between 0 and 1 with sigmoid function, which represents the probability that the output is a real or fake image.

Figure 3. Compared with the ground truth image, our result gives more details and brightness.

Figure 4. (a). The model applying on CT image Denoising. (b). Application for image defogging.

Figure 5. The PSNR curve of three different image shapes with respect to training iterations.

6. Conclusion
In this paper a DRGAN model is presented for image denoising, which achieves good perceptual quality with respect both overall content and specific details. The trained model performs amazing effect not only on single image denoising task, but also on other aspect such as defogging. We stacked more residual blocks up to 8 layers considering its feature maintaining ability. The modified loss function ensures that the model can keep the main feature without the loss of perceptual details. The experiment on the CT dataset shows that our model can also process grayscale images, presenting stronger texture and brightness.

Acknowledgment
The work was supported by National Key R&D Program of China (Grant Nos. 2018YFB1306600), National Natural Science Foundation of China (Grant Nos.61571372, 61672436, 61601376), Fundamental Science and Advanced Technology Research Foundation of Chongqing (cstc2017jcyjBX0050, cstc2016jcyjA0547), Fundamental Research Funds for the Central Universities (Grant Nos.XDJK2016A001, XDJK2017A005)

References
[1] Bhawna Goyal, Ayush Dogra, and Sunil Agrawal, B.S. (2019) Image Denoising Review: From classical to state-of-the-art approaches. Information fusion. vol. 55, pp. 220-244.
[2] Jain, V. and Seung, S. (2009) Natural image denoising with convolutional networks. In Advances in Neural Information Processing Systems. pp. 769-776.

[3] Goodfellow, Ian, et al. (2014) Generative adversarial nets. Advances in Neural Information Processing Systems.

[4] Burger, Harold C., Christian J. Schuler, and Stefan Harmeling. (2012) Image denoising: Can plain Neural Networks compete with BM3D? Computer Vision and Pattern Recognition. IEEE Conference on.

[5] Seungjun Nah, Tae Hyun Kim, Kyoung Mu Lee. (2016) Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. Computer vision and pattern recognition.

[6] He K.M., Zhang X.Y., Ren S.Q. Sun J. (2015) Deep Residual Learning for Image Recognition, Computer vision and pattern recognition.

[7] Zhu, J. Y., Park, T., Isola, P., Efros. A. (2017) Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.

[8] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych. (2018) DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. Conference on Computer Vision and Pattern Recognition.

[9] Wang Y. D., Armstrong, RT., Mostaghimi, P. (2019) Enhancing Resolution of Digital Rock Images with Super Resolution.

[10] Ledig C, Theis L, Huszar F, et al. (2016) Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.

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