Single Image De-raining Based on a Novel Enhanced Attentive Generative Adversarial Network

Haochen Zhou and Qin Wei
School of Information Engineering, Wuhan University of Technology, Wuhan, Hubei Pro., 430072, China
Email: 763459261@qq.com; qinwei@whut.edu.cn

Abstract. With rapid development of deep learning in artificial intelligence and computer vision, generative adversarial network plays an important role in single image de-raining. Attentive generative adversarial network (AttGAN) has problems about complicated network structure and distortion of background color. Considering the relative complex background in the real image with raindrops, a new single image de-raining algorithm based on enhanced attentive generative adversarial network (EAttGAN) is proposed to retain the original background of the blurred image. In order to restore more complete background and accelerate network training, a generator enhanced by residual scaling and a Markovian discriminator are fused effectively in the network. Compared with AttGAN, experimental results indicate that EAttGAN can not only achieve higher sharpness of a single image, but also take less time in the training process.

1. Introduction
Image de-raining is an important research direction in the field of computer vision. Raindrops would seriously reduce visibility of image and block background information [1]. Therefore, it is of importance to remove raindrops from blurred image and simultaneously restore background information. Along with rapid development of deep learning in computer vision, kinds of networks were proposed to deblur images with raindrops. Yang et al [2] proposed a detection network which could automatically detect areas of raindrops and remove their cumulative effects on these areas. Fu et al [3] completed an attempt to remove rain streaks by a deep detail network. The convolutional neural network (CNN) was applied and proven to be feasible and advantageous in single image de-raining, however, it still has some limitations in the application. First of all, the serial structure and fully supervised learning paradigm of traditional CNN make it stiff and difficult to identify real images with raindrops when trained by synthetic images. Secondly, its lack of sufficient constraints would result in blurred images.

Therefore, generative adversarial network (GAN) was developed to generate sharper samples for more effective identification. Moreover, Qian et al [4] established an attentive GAN (AttGAN), which not only evaluated the local continuity but also enhanced performance of raindrops removal in single image de-raining. But it has a big disadvantage that its outputs are easily to lose the information of background colour. This paper proposes an enhanced attentive generative adversarial network (EAttGAN), which aims to overcome the defeat of GAN and further promote its performance in image de-raining and training speed.

The remainder of this paper is structured as follows. Section 2 describes our proposed method EAttGAN. Section 3 illustrates experiments, results and discussion. Section 4 gives the conclusion.
2. Image De-raining Based on Enhanced Attentive GAN

2.1. General Introduction of GAN

Generative adversarial network has developed rapidly in recent years, which was inspired by the two-character zero-sum game to train the network through confrontation [5]. As shown in Figure 1, it is the structure of GAN, where the generator $G$ is responsible for obtaining the simple distributions, and the discriminator $D$ takes charge of judging whether the input from $G$ is real. Through the game between $G$ and $D$, the data finally generated from $G$ would be closer to real data, and $D$ would lose judgment [6]. This function of GAN makes it possible to repair a variety of blurred images.

2.1.1. Generator

Generator is the core part of GAN, and its training goal is to cause mistakes of discriminator. Its generative network attempts to generate real sample through inputting random noises, and to obtain detailed information from neighboring pixels so as to generate high-quality images.

2.1.2. Discriminator

Discriminator is able to classify real and fake data by generating a scalar ranging from 0 to 1.

2.1.3. Architecture of attentive gan

Attentive GAN is an original network to deal with the problem of raindrops removal recently. Compared with GAN, an attention recurrent network and an autoencoder are added into the generator. This recurrent network with a few blocks can generate visual attention map, and each block consists of five residual networks (ResNets) [7], a long short-term memory unit (LSTM), and a convolutional layer. Behind the recurrent network, a deep autoencoder, which involves 16 convs-relu blocks, is in charge of processing concatenation of input image and final attention map from the attentive-recurrent network. In addition, the discriminator is composed of a few convolutional layers, a fully connected layer and a single neuron. According to this complex structure and attentive network, raindrops could be easily removed from the blurred image.

2.2. Architecture of EAttGAN

EAttGAN has made a great progress on the basis of AttGAN, which is easily to lose the information of background colour with its complex structure. As shown in Figure 2, the architecture of EAttGAN mainly and simply includes two parts: a generator and a discriminator. The generator is able to produce a sharp image without raindrops as much as possible through enhanced ResNets, and the discriminator is capable of making quick decision about whether the image generated by the generator is authentic or not by a Markovian discriminator and simplified structure. Single image with raindrops is transformed into single clear image by training this novel network, in which the generative adversarial loss can be expressed as:

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{R:\text{sharp}} [\log(D(R))] + \mathbb{E}_{I:\text{raindrops}} [\log(1 - G(I))]$$  \hspace{1cm} (1)

Where $I$ is the generated image from generator, $R$ is the sharp natural image.

2.2.1. Generator of enhanced attentive gan

The attention recurrent network in original AttGAN is the primary part to be modified in EAttGAN, residual scaling [8, 9] is applied to enhance the performance of ResNets through residual reduction before its accumulation, which is helpful to correct the improper initial weight. Besides, it is also useful to keep network stable and avoid magnifying the
magnitudes of input signals in ResNets. For each ResNet in EAttGAN, all residuals obtained from the last convolution layer are multiplied by 0.2 for lower weight.

\[ L_G(\{A_t\}, M) = \sum_{t=1}^{N} \theta^{N-t} L_{MSE}(0.2A_t, M) \]  

\[ 1 \quad (2) \]

First of all, the input image is concatenated with an initial attention map and multiplied by an initial weight of 0.2 when \( t=1 \) [4]. Moreover, this process would be repeated \( N \) times in the attention recurrent network, in which \( N \) is set to 4 and \( \theta \) is set to 0.8. And a smaller initial weight 0.2 is used in ResNets to help the generative network jump out from the bad local minimum, while the original weight used in AttGAN is set as 0.5.

2.2.2. Discriminator of enhanced attentive gan In addition to advance the performance of generator, a Markovian discriminator [10] is fused in EAttGAN in order to improve the discriminative network. Using a Markovian discriminator to operate image patches [11] could keep fixed high resolution and high details, produce sharper results than standard global discriminators, ensure the authenticity of the restored background, and ultimately produce better restorative effects.

Compared with AttGAN, it replaces redundant convolutional layers with a Markovian discriminator in EAttGAN, and just the last three convolutional layers are retained to extract features. By this way, Markovian discriminator could accelerate the training speed and decrease the complexity of this network. The loss function of discriminator is expressed as below:

\[ L_D(O, R) = -\log(D(R)) - \log(1 - D(O)) + \beta L_{MSE}(D_{MSE}(R), 0) \]  

\[ 2 \quad (3) \]

Where \( R \) is referred to as a sample extracted from the sharp image, \( O \) is the output image of the autoencoder. Since a Markovian discriminator is used, the requirement of attention map are reduced,

Figure 2. Architecture of EAttGAN.
and the calculation of $MSE$ between output image of autoencoder and attention map is omitted in its loss function compared to AttGAN. The last item in the loss function is able to reduce unnecessary feature learning effectively when there is no specific region necessary to focus on, by which computation of discriminator is simplified furtherly.

3. Experiment and Results

3.1. Experiment and Dataset
In order to verify the improvement of our EAttGAN in single image de-raining, four different networks are constructed, trained and tested, which are AttGAN, the enhanced network V1 (only generator enhanced), the enhanced network V2 (only discriminator enhanced), and EAttGAN (both generator and discriminator enhanced). All experiments are realized by GPU NVIDIA RTX2060, and experimental dataset is same with that used in ref [4]. The entire dataset contains approximately 1,000 pairs of images. Each pair is composed of one sharp image and one blurred by raindrops that both have the same background. Before networks are trained, almost 100 pairs of image dataset are selected randomly for test. Furthermore, the single training time and the quality of the generated image would be recorded as key factors to assess the performance of four different networks.

3.2. Experimental Results

3.2.1. De-raining image. Figure 3 displays results of three different images with raindrops. The first column is input image, the second one is result of AttGAN, and the third one is that of EAttGAN. Results indicate that EAttGAN is not only able to remove raindrops from blurred image, but also gets better performance in restoration of background and color information.

Figure 3 From left to right are input images, results from AttGAN and EAttGAN respectively
3.2.2. Performance among different networks. In general, the peak signal to noise ratio (PSNR) [12] and the structural similarity (SSIM) [13] are two vital factors to evaluate the quality of deblurred image. Tables 1 and 2 give those factors among four different networks shown as below.

**Table 1. PSNR of four different networks.**

| Image No. | AttGAN | V1     | V2     | EAttGAN |
|-----------|--------|--------|--------|---------|
| 1         | 21.06  | 22.13  | 20.90  | 21.25   |
| 2         | 24.45  | 26.02  | 24.09  | 24.71   |
| 3         | 21.89  | 22.15  | 21.24  | 21.97   |
| 4         | 24.29  | 25.77  | 23.98  | 24.50   |

Due to higher PSNR generally means better effect of deblurring on images, it is obvious that V1 has the highest PSNR, and EAttGAN is the one behind the first, and V2 has the lowest value in processing different images. These results apparently indicate that a smaller initial weight in generator could obtain a higher PSNR, so that V1 and EAttGAN obtain more information of background and generate higher quality of images without raindrops.

**Table 2. SSIM of four different networks.**

| Image NO. | AttGAN | V1     | V2     | EAttGAN |
|-----------|--------|--------|--------|---------|
| 1         | 0.891  | 0.886  | 0.896  | 0.892   |
| 2         | 0.879  | 0.874  | 0.885  | 0.881   |
| 3         | 0.862  | 0.858  | 0.869  | 0.864   |
| 4         | 0.894  | 0.890  | 0.907  | 0.903   |

Higher SSIM indicates higher similarity between results and ground truth. In Table 2, V2 (discriminator enhanced) has better results than V1 (generator enhanced), AttGAN and even EAttGAN. In fact, EAttGAN has a slight improvement in SSIM while results of EAttGAN are closed to those of AttGAN. Thereby, it is difficult to achieve significant advance in SSIM because of the proximity of output images and real scenes [14].

The training time is also an essential factor to evaluate the structure of networks. Table 3 gives the single training time among four different networks shown as below.

**Table 3. Single training time of four different networks.**

| Batch | AttGAN | V1     | V2     | EAttGAN |
|-------|--------|--------|--------|---------|
| 1     | 0.77±0.05s | 0.79±0.05s | 0.45±0.05s | 0.48±0.05s |

During single trainings, V2 and EAttGAN take less time than V1 and AttGAN. Because V2 and EAttGAN have less convolutional layers in their discriminators, which is definitely a critical time-consuming layer in the network. EAttGAN replaces conventional discriminator with Markovian discriminator, which enhances efficiency of training visibly through the results.

Compared with AttGAN, the proposed EAttGAN achieves very competitive performance, the training speed is faster, and both PSNR and SSIM of EAttGAN outperform those of AttGAN. The PSNR of EAttGAN is increased significantly while the SSIM of EAttGAN has a slight rise. Besides, background color of sharp output image is more complete through EAttGAN in comparison with AttGAN.
4. Conclusion
In this paper, a single image de-raining method based on EAttGAN is proposed, which is powerful and efficient for raindrops removal and background restoration in a single image. An enhanced generator and a Markovian discriminator are combined effectively in this method, which ensures the authenticity and completeness of the restored background and achieves better performance in deblurring simultaneously. Besides, the structure of Markovian discriminator is optimized in order to improve training speed. The experimental results in different improved networks indicate that the EAttGAN has comprehensively better performance than AttGAN in single image de-raining.

5. Acknowledgments
This research is supported by National Natural Science Foundation of China (Grant No.51675389).

6. References
[1] Robby T. Tan 2008 Visibility in bad weather from a single image IEEE Conf. on Computer Vision and Pattern Recognition pp 2347–2354
[2] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo and Shuicheng Yan 2017 Deep joint rain detection and removal from a single image IEEE Conf. on Computer Vision and Pattern Recognition pp 1685–1694
[3] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding and John Paisley 2017 Removing rain from single images via a deep detail network IEEE Conf. on Computer Vision and Pattern Recognition pp 1715–1723
[4] Rui Qian, Robby T. Tan, Wenhan Yang, Jiajun Su and Jiaying Liu 2018 Attentive generative adversarial network for raindrop removal from a single image IEEE Conf. on Computer Vision and Pattern Recognition pp 2482–2491
[5] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio 2014 Generative adversarial networks Advances in Neural Information Proc. Systems pp 2672–2680
[6] Tim Salimans, I.J. Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford and Xi Chen 2016 Improved techniques for training generative adversarial networks Advances in Neural Information Proc. Systems pp 2234–2242
[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun 2017 Deep residual learning for image recognition IEEE Conf. on Computer Vision and Pattern Recognition pp 770–778
[8] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao and Xiaoou Tang 2018 ESRGAN: enhanced super-resolution generative adversarial networks European Conf. on Computer Vision pp 63–79
[9] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke and Alex Alemi 2016 Inception-v4, inception-resnet and the impact of residual connections on learning AAAI Conf. on Artificial Intelligence pp 4278–4284
[10] Chuan Li and Michael Wand. 2016 Precomputed real-time texture synthesis with markovian generative adversarial networks. European Conf. on Computer Vision pp 702–716
[11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou and Alexei A. Efros 2017 Image-to-image translation with conditional adversarial networks IEEE Conf. on Computer Vision and Pattern Recognition pp 5967–5976
[12] Q. Huynh-Thu and M. Ghanbari 2008 Scope of validity of psnr in image/video quality assessment Electronics Letters 44(13):800–801
[13] Zhou Wang and Alan C. Bovik 2002 A universal image quality index. IEEE Trans Image Signal Proc. Lett. 19(3):81–84
[14] Ruoteng Li, Loong-Fah Cheong and Robby T. Tan 2019 Heavy rain image restoration: integrating physics model and conditional adversarial learning IEEE Conf. on Computer Vision and Pattern Recognition pp 1633–1642