Original Research Paper | Computer Science | Volume : 5 ¦ Issue : 9 ¦ Sep 2020 ¦ e-ISSN : 2456-5040

PEDESTRIAN IMAGE INPAINTING BASED ON GENERATIVE ADVERSARIAL NETWORK ARCHITECTURE

Tongshun Zhang

School of Computer Science and Software Engineering Tiangong University, Tianjin, 300387, China.

ABSTRACT

The research of image defect repair aims to automatically repair the defect content in the image through the computer. In recent years, the emergence of deep neural network technology has effectively promoted the development of image restoration technology. This article focuses on the subject of pedestrian image restoration. Today, surveillance cameras have been widely deployed in every corner of the city. Therefore, huge pedestrian images can be acquired every second. Therefore, how to automatically analyze and understand its basic content has become an urgent research topic, which has obvious theoretical and practical value. Obviously, the quality of the source image will seriously affect the subsequent stage of understanding. Therefore, this article will discuss how to recover damaged pedestrian images. In order to recover robustly, we adopted an adversarial generation framework.

KEYWORDS: image inpainting; deep learning; generative adversarial networks; neural networks; computer vision.

1. INTRODUCTION:

In the past few years, due to the rapid development of deep machine learning technology, research in the field of computer vision and image processing has made amazing progress on many topics such as image restoration, style transfer, and image quality improvement. In particular, the generative confrontation network has made great achievements in the field of image restoration. Aiming at the theme of pedestrian images, we use generative confrontation networks to achieve the purpose of repairing pedestrian images.

Image restoration is a technology that restores the missing content on the image, restores the lost part of the image, and reconstructs them based on the background information. It refers to the process of filling in missing data in a designated area of visual input. In the digital world, it refers to the application of complex algorithms to replace missing or damaged parts of image data.

The PatchMatch (Barnes C, Shechtman E, Finkelstein A, et al) method fills up the damaged area by calculating and searching for more similar blocks in the image. In addition, deep learning methods have developed rapidly in recent years. Iizuka et al. added a content-aware layer to realize image restoration on the basis of generating a confrontation network (Iizuka S, Simo-Serra E, Ishikawa H). Yu et al. designed a loss function based on global and local missing information, and focused on repairing missing content while maintaining image semantics (Yu J, Lin Z, Yang J, et al.). Liu et al. designed and proposed a method based on partial convolution to repair irregular image holes. Context Encoders (Pathak D, Krahenbuhl P, Donahue J, et al.) infers the damaged area of the image based on unsupervised context pixel prediction.

2. GENERATIVE CONFRONTATION NETWORK:

GAN stands for Adversarial Generative Network. As the name suggests, it is a type of generative model, and its training is in a state of adversarial game. The main structure of GAN includes a generator G (Generator) and a discriminator D (Discriminator). G is a network that generates pictures. It receives a random noise z, and generates pictures through this noise, denoted as G(z). D is a discriminant network to determine whether a picture is "real". Its input parameter is x, x represents a picture, output D(x) represents the probability that x is a real picture, if it is 1, it means 100% is a real picture, and the output is 0, it means it cannot be real picture of. In the training process, the goal of generating network G is to generate real pictures as much as possible to deceive and discriminate network D. The goal of D is to separate the pictures generated by G from the real pictures as much as possible. In this way, G and D constitute a dynamic "game process".

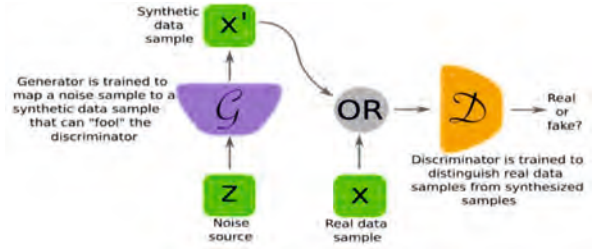

Figure1: Generative confrontation network

3. OUR METHOD:

Figure 2 shows our overall architecture, which consists of a generator and a discriminator. We will use Color Transfer to process pictures into pictures of other tones. We can see that our architecture can generate restored pictures with normal tones.

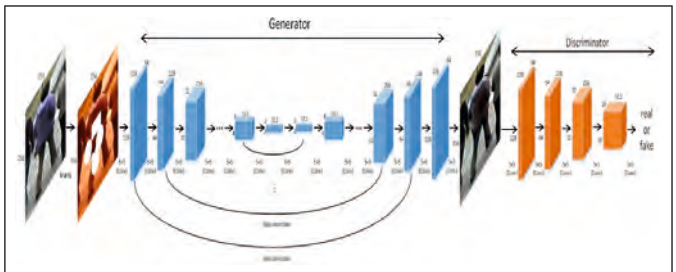

Figure 2: Our Network

3.1 Generator:

Our generator uses U-NET architecture based on Encoder-Decoder structure. Due to the structural similarity between input and output, we consider design generator architecture based on them. The encoder-decoder network structure down-samples the input into low-dimensional embeddings through convolution. After down-sampling to a certain extent, it reaches the bottleneck layer, and then reverses the process, and up-samples the embedding to reconstruct the image by transposed convolution. Finally achieve the repair effect. Each convolutional layer in the generator uses a 5 * 5 convolution kernel with a step size of 2. After convolution in the encoder part of the network, batch normalization and Leaky ReLU activation are performed. After the decoder is deconvolved, batch normalization and ReLU activation are performed. What is not completed is the last layer of the decoder, which uses Tanh nonlinearity to match the input distribution [-1,1].

3.2 Dscriminator:

Our discriminator is completely different from ordinary GAN. The general network discriminator of GAN maps the input to a real number, that is, the probability that the input sample is a real number. PatchGAN maps the input to N*N color blocks, and averages the probability that each color block is a real sample as the final output of the discriminator. We use PatchGAN because it only distinguishes based on the size of the patch, distinguishing whether each patch in the image is right or wrong, so for each block in the input, its acceptance field is very small. For general discriminators, PatchGAN can pay more attention to the details of the image. There are two differences in different tasks. Each convolutional layer uses a 5 * 5 convolution kernel with a step size of 2. After convolution, it is batch normalization and Leaky ReLU activation.

3.3 Loss Function:

The method in this paper takes the competitive generative confrontation as the basic framework, and the loss function must first reflect the competitive confrontation. Our adversarial loss function is defined as:

$$L_{GAN} = \min_G \max_D \mathbb{E}_{\mathbf{x}_{GT}\sim P_{data}}[\log D(\mathbf{X}_{GT})] + \mathbb{E}_{\mathbf{x}_{in}\sim P_{data}}\left[\log\left(1 - D(\mathrm{G}(\mathbf{X}_{in}, z))\right)\right]. \quad (1)$$

In addition, Image restoration and reconstruction should have clear directional targets. We emphasize this constraint in the form of L1 norm loss:

$$L_{\ell_1} = \lambda_{\ell_1} \|\mathbf{X}_{out} - \mathbf{X}_{GT}\|_1, \quad (2)$$

Based on the above, our total Loss is:

$$L_{Loss} = L_{GAN} + L_{\ell_1}. \quad (3)$$

4. EXPERIMENT:

4.1 Dataset:

PETA's data set on human attributes includes 10 subsets, each of which is a pedestrian data set acquired in different environments. We used several subsets of the human attribute data set PETA (Deng Y, Ping L, Chen CL) to train and evaluate our method. We chose a large subset of the PETA data set as our ground truth images, and used color transfer to generate images with poor tones. A total of 4291 pairs of images are used as our training set. We pre-set for five tonal distortions and divide 4291 images into five equal parts, close to the number of each type of cold, warm, green, dark and bright.

4.2 Test:

We have done corresponding tests on a subset of the PETA data set. Figure 3 and Figure 4 show our test results.

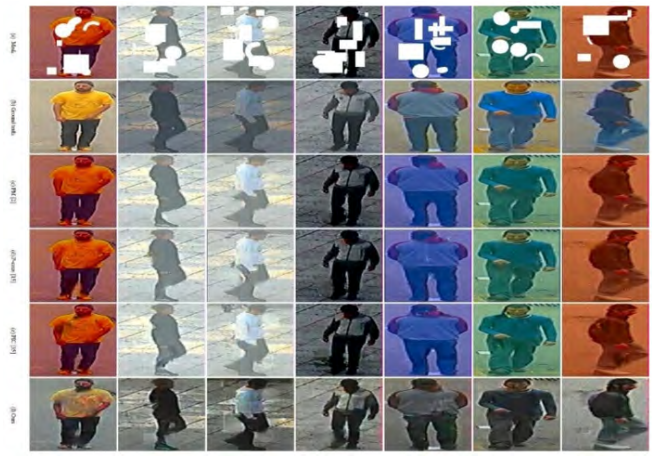

Figure 3: Experiments on SARC3D

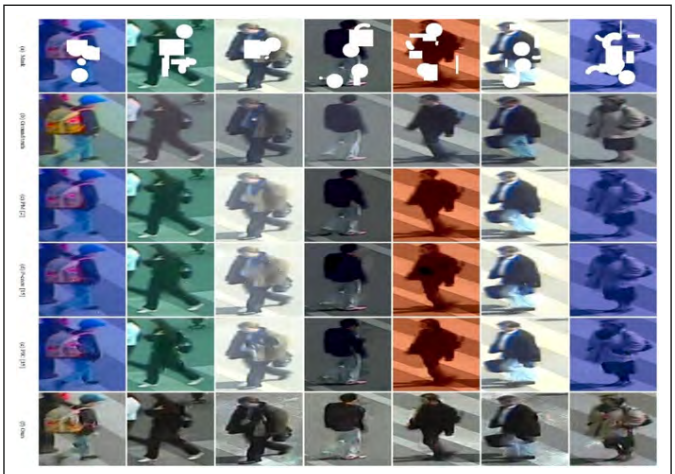

Figure 4: Experiments on PRID

4.3 Evaluation:

First of all, in the comparison experiment setting, we chose PatchMatch, Partial convolution (Guilin Liu, Fitsum A. Reda, Kevin J. Shih, et al.) and PIC (Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai.) as our comparative experiments. Table 1 shows our comparison results. Our results are the best in terms of indicators, because other methods do not take into account the problem of color correction. In the above experiment, we have demonstrated five tonal distortions. Since most of the existing repair methods are only used for image repair, and our method is to correct the color tone when repairing the image, we are not only addressing the problem of image repair, but also similar to image translation. Therefore, there are certain difficulties in comparing methods.

Table 1: Evaluations

Method	PSNR	MSE	SSIM
Pconv	13.58	3339.23	0.66
PIC	13.55	3414.57	0.69
PatchMatch	13.75	3300.39	0.70
Our	18.52	1065.15	0.71

5. CONCLUSIONS:

This paper presents a method of pedestrian image restoration. At the same time, the problem of color correction is taken into account. In order to achieve this goal, we designed a new GANs network structure. Our network structure contains a generator and a discriminator. When processing the picture, the five tonal distortion problems in the real world are considered. Finally, we can get better results in tone correction and restoration.

REFERENCES:

I. Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch[C]// Acm Siggraph. ACM, 2009.

II. Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. ACM Transactions on Graphics, 2017, 36(4): 107.

III. Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, United States, 2018: 5505-5514.

IV. Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: feature learning by inpainting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, America, 2016: 2536-2544.

V. Deng Y, Ping L, Chen CL, Member, IEEE (2015) Learning to recognize pedestrian attribute. Computer Science

VI. Liu G, Reda FA, Shih KJ, Wang TC, Tao A, Catanzaro B (2018) Image inpainting for irregular holes using partial convolutions. In: European Conference on Computer Vision

VII. Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. Image inpainting for irregular holes using partial convolutions. In European Conference on Computer Vision, pages 85–100, 2018.

VIII. Chuanxia Zheng, Tat-Jen Cham, and Jianfei Cai. Pluralistic image completion. In IEEE Conference on Computer Vision and Pattern Recognition, 2019.