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 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:

\[ L_{GAN} = \min_{G} \max_{D} \mathbb{E}_{x \sim p_{data}} \left[ \log D(x) \right] + \mathbb{E}_{z \sim p_{z}} \left[ \log (1 - D(G(z))) \right]. \]  

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 a real picture. 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".

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.

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 deconvoluted, 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 Discriminator: 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 \( N \times N \) 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:

\[ L_{GAN} = \min_{G} \max_{D} \mathbb{E}_{x \sim p_{data}} \left[ \log D(x) \right] + \mathbb{E}_{z \sim p_{z}} \left[ \log (1 - D(G(z))) \right]. \]
In addition, Image restoration and reconstruction should have clear directional targets. We emphasize this constraint in the form of L1 norm loss:

\[ L_{\text{L1}} = \lambda_{\text{L1}} || X_{\text{out}} - X_{\text{gt}} ||_1 \]  

Based on the above, our total Loss is:

\[ L_{\text{Loss}} = L_{\text{GAN}} + L_{\text{L1}} \]  

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>
<thead>
<tr>
<th>Method</th>
<th>PSNR</th>
<th>MSE</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pconv</td>
<td>13.58</td>
<td>3339.23</td>
<td>0.66</td>
</tr>
<tr>
<td>PIC</td>
<td>13.55</td>
<td>3414.57</td>
<td>0.69</td>
</tr>
<tr>
<td>PatchMatch</td>
<td>13.75</td>
<td>3300.39</td>
<td>0.70</td>
</tr>
<tr>
<td>Our</td>
<td>18.52</td>
<td>1065.15</td>
<td>0.71</td>
</tr>
</tbody>
</table>

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: