Design and Implementation of Occlusion Image Recognition Algorithm Based on Deep Convolution Generative Adversarial Network

Zhijun Liu¹, Feng Cai¹, Xuefeng Pan²*

¹College of Information Science and Engineering, Wuchang Shouyi University, Wuhan, Hubei, China
²College of Computer, Wuhan Qingchuan University, Wuhan, Hubei, China
*E-mail: liuzhijun@wsyu.edu.cn

Abstract. In the image recognition process, the accuracy of image recognition is affected due to partial coverage or light problems. This paper proposes a method to solve the problem of target occlusion based on a deep convolution generation confrontation network. The method actively occludes the feature map after the feature is extracted to generate the confrontation sample, and generates a map and a mask from the input image. At the same time, a method is proposed. This new loss function applies the algorithm to handwriting recognition and trains the network through large-scale sample data sets. Experiments show that this method significantly improves the accuracy of image extraction.

1. Introduction
Handwriting recognition produces large differences in images due to writing. When the handwriting is unclear or other reasons, part of the image information is lost, which greatly affects the accuracy of recognition. This type of problem is summarized into the problem of image occlusion, and Generative Adversarial Networks (GAN) proposes new ideas for occlusion image recognition.

The deep convolutional generation confrontation network DCGAN (Deep Convolutional GAN) [1] realizes the first combination of CNN and GAN, which replacing the multi-layer perceptron model to ensure the stable training of the GAN network. But it can only process low-resolution images, which has greater limitations. Yeh et al. [2] searched for the closest pixel block in the latent space and decoded it to obtain a smooth and complete image. Li et al. [3] used a deep generative model, a code generator to output a fuzzy picture, and two discriminators, a local discriminator and a global discriminator, perfected the detailed information of the generated picture. Yang et al. proposed a high-resolution image inpainting algorithm [4], which uses context encoders for global content constraints, and uses the image blocks extracted by the intermediate feature layer to model the texture constraints of local regions. The visual effect is significant, but the optimization process Complex and slow in processing speed. Yu et al. proposed Deep Fill V1 [5] based on the content-aware generative model, which can extract features with high similarity from long-distance regions to achieve high-quality restoration effects. Zhu et al. [6] introduced Cycle GAN to solve the problem of image-to-image conversion, which has received widespread attention. Sun et al. combined the traditional division method based on radicals and font structure with neural network [7] to obtain more accurate Chinese character information.
The illegible handwriting and missing handwriting in handwriting recognition by Use deep neural convolution to generate a confrontation network to realize the recognition. The algorithm first uses the standard data set samples to train the adversarial network, and then uses the trained model to complete the detection of the occluded target and realize the recognition of the image. The algorithm is robust.

![Characters with occlusion](image)

**Figure 1** Characters with occlusion

### 2. Pretreatment

Character recognition steps include: preprocessing, feature extraction and classification (recognition) process. Suppose S is taken as the input image, expressed as $S=[S]_{uv}$, where $u$ and $v$ are the pixels of the rows and columns of the image when performing character recognition. The feature vector is generated based on the character image, and the generated feature vector can be expressed as $F=[F]_{uv}$.

In the recognition process, the preprocessing mainly filters and resizes the image of the recorded characters. In the preprocessing process, the extraction of the region of interest (ROI) [8] is carried out in the input character image covering the pixel intensity, which represents the existence of the character sequence in the image. After the ROI is extracted, the image is scaled to distinguish between the character image and the unnecessary background of the image. The size of the input character image is not completely minimized in the preprocessing process, on the contrary, the pixel size is reduced.

### 3. Detection model based on generative adversarial network

#### 3.1. Adversarial learning

Suppose the original target detection network is $D(X)$, where $X$ is one of the target candidate frames. The target detector will have two outputs, $D_t$ represents the predicted category, and $D_r$ represents the predicted detection frame position [9]. At the same time, assuming that the space position of the ground truth box of $X$ is $L$ and the ground truth category is $C$, then the initial detector loss is as in formula 1.

$$L_D = L_{\text{softmax}}(D_c(X), C) + [C \notin b_g]L_{\text{bbox}}(D_r(X), L)$$

(1)

Among them, $L_{\text{softmax}}$ represents the Soft Max classification loss, and $L_{\text{bbox}}$ represents the positioning loss of the predicted frame position and the target real frame position.

Assuming that the generator of the anti-occlusion network is denoted as $M$, the feature $X$ obtained from the image $F$ is input into the anti-occlusion network, thereby generating a new anti-occlusion network $M(X)$. This paper uses the loss function defined in formula 2 to train the anti-occlusion network $M$.

$$L_M = 1 - L_{\text{softmax}}(D_c(M(X)), C)$$

(2)

The anti-occlusion network and the original detector are in a relationship of mutual competition and mutual learning. When the detector and the anti-occlusion network learn from each other through competition, and the loss reaches a balance, the entire network can better detect the partially occluded
target.

3.2. Construction of the confrontation network

Confronting the occlusion network to create occlusion for the deep features of the target. After the picture passes through the ROI pooling layer, the convolutional features of each foreground target candidate area can be obtained, and these features can be used as the input of the anti-occlusion network. For the input feature of a certain target, the anti-occlusion network will generate a binary mask, indicating the area of the feature to be occluded, making it difficult for the detector to recognize. The structure of the anti-occlusion network is shown in Figure 2.

Figure 2. Anti-occlusion network structure diagram

In scene character recognition, real pictures have a great influence on the recognition of characters in images [10]. Input the picture into the network, after the multi-layer feature extraction and ROI pooling layer, a feature layer X with a size of d×d×c will be obtained, where d represents the feature size and c represents the number of feature channels. In VGG-16, the value of d is 7 and the value of c is 256. After obtaining this feature layer, the anti-occlusion network will preset a binarization mask M with a size of d×d, and its internal value is 0 or 1. Denote the value of the i-th row and j-th column of the binarization mask M as \( M_{ij} \), and denote the value of the i-th row and j-th channel of the feature layer X as \( X_{ijk} \). If the value of \( M_{ij} \) is 1, then \( X_{ijk} \) remains No change, on the contrary, if the value of \( M_{ij} \) is 0, the value of the corresponding spatial position on all channels of X is cleared to zero, that is, \( X_{ijk} = 0 \).

Figure 3. Structure occlusion sample

The feature layer X of size d×d is input to the anti-occlusion network, and the size of the sliding window used is \( d/n \times d/n \), as shown in Figure 3 (at this time \( n=3 \)), for each sliding window, Clear the values at the corresponding spatial positions on all channels of the feature layer occluded by the window (that is, 0), and generate a new feature vector for the suggested region, and input it to the
classification layer to calculate the loss. After obtaining the loss values of all \( n^2 \) \( \frac{d}{n} \times \frac{d}{n} \) windows, select the sliding window with the highest loss value to obtain the required occlusion sample, and then send the occlusion sample to the classifier for classification.

### 3.3. Loss function

The anti-occlusion network generates a binarization mask on \( n \) candidate frames, and finally generates \( n \) pairs of training samples with occlusion \(((X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n))\). In this paper, the cross-entropy loss shown in formula 3 is used to train the anti-occlusion network:

\[
L_A = \frac{1}{n} \sum_{k} \sum_{i,j} \left[ y_{ij}^k M_{ij}(X^k) + (y_{ij}^k - 1)(M_{ij}(X^k) - 1) \right]
\]

(3)

Among them, \((MijXk)\) represents the output of the feature map \( Xk \) at position \((i, j)\) after passing through the anti-occlusion network. The purpose of processing is to allow the adversarial occlusion network to generate occlusion training samples that can bring high losses to the detection network and improve the recognition ability of the detector. The anti-occlusion network is used to improve the performance of the booster detector during training.

### 3.4. The detection algorithm against the network

Enter:
- \( F \): Original image
- \( m \): Occlusion mask
- \( Z \): Occluded image

Output:
- \( X \): Predicted image

1. Define a discriminator \( d \) and a generator \( g \) and set them as training mode.
2. Define the optimizer object
3. Repeat 4-10 when network \( M \) does not converge
4. Repeat 5-8 to update the discriminator
5. Batch samples \( F \) in the training data
6. Generate simulated occlusion image \( Z \leftarrow F \odot (1 - m) \)
7. Obtain the predicted sample \( X' \leftarrow Z + (Z, m) \odot m \)
8. Use the results to update the discriminator parameters
9. Batch samples \( x \) in the training data
10. Update and repair the parameters of the network \( M \) according to the L loss and the adversarial loss \( L_g \) of the discriminator
4. Experiment and result analysis

The experimental data set uses CASIA-HWDB data set and ICDAR2015 data set for experiments. The CASIA-HWDB data set is composed of the characters of 420 writers, and the total number of samples is 1,680,258, the most common handwritten Chinese character recognition data set [11]. ICDAR2015 is a Chinese scene text detection and recognition data set. The data set contains high-definition text images. The image size ranges from 116×36 to 1520×60 [12].

(1) Loss curve of model generator and discriminator

In the generative adversarial network model algorithm proposed in this paper, the loss curves of its generator and discriminator are as follows. From the loss trend graph in the figure, it can be seen that the discriminator shows a trend of rapid decline in iteration, and its detection rate increases rapidly. But after a certain number of iterations, although there is a decline, the trend is not obvious, and the model has stabilized.

(2) Accuracy

Train and test on CASIA-HWDB data set and ICDAR2015 data set, and compare the algorithm in this article with other algorithms. The experimental data is as follows. It can improve the accuracy of handwritten Chinese character recognition.
The accuracy of 9 points in the 3×3 area centered on the Chinese character center point is 96.12%. Using the confrontation generation network, after setting the occlusion area training, the handwritten character detection can be completed quickly. Based on the comparison and analysis results, the detection algorithm proposed in this paper can achieve higher detection accuracy.

5. Conclusion
The algorithm for using active occlusion to generate adversarial samples and generating adversarial networks to recognize handwritten characters is proposed. After generating samples from the original pictures and participating in the training of the adversarial network, the model can effectively improve the ability to recognize characters with accurate information. It has a good effect on character information detection and recognition in special scenes.

Acknowledgments
This paper is supported by the Natural Science Foundation of Hubei Provincial under the Grant No. 2016CFB620.Guiding Project of Science Research plan of Education Department of Hubei Province (B2018349)

References
[1] Tang C, Shi J, Yang Y, et al. Malware Detection Model Based on Deep Convolution Generation Adversarial Network[J]. Journal of Physics: Conference Series, 2021, 1738(1):012110 (5pp).
[2] Yeh R, Chen C, Lim T, et al. Semantic Image Inpainting with Perceptual and Contextual Losses[J]. ar Xiv preprint ar Xiv:1607.07539, 2016.
[3] Li Y, Liu S, Yang J, et al. Generative Face Completion[C]. //IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017: 5892-5900.
[4] Yang C, Lu X, Lin Z, et al. High-Resolution Image Inpainting using Multi-Scale Neural Patch Synthesis[C]. //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4076-4084.
[5] Yu J, Lin Z, Yang J, et al. Generative Image Inpainting with Contextual Attention[C]. //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5505-5514.
[6] Lefkimmiatis S. Universal Denoising Networks: a Novel CNN Architecture for Image Denoising[C]. //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3204-3213.
[7] D. Sun, T. Ren, C. Li, et al. Learning to write stylized chinese characters by reading a handful of examples[C]. Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI , July 13-19, 2018, Stockholm, Sweden., 2018, 920-927
[8] Wang W, Bu F, Lin Z, et al. Learning Methods of Convolutional Neural Network Combined with Image Feature Extraction in Brain Tumor Detection[J]. IEEE Access, 2020, PP(99):1-1.

[9] Mahmoud M, Guo P, Xin X. A Novel Method for Traffic Sign Recognition Based on DCGAN and MLP with PILAE Algorithm[J]. IEEE Access, 2019, PP(99):1-1.

[10] Naiemi F, Ghods V, Khalesi H. A Novel Pipeline Framework for Multi Oriented Scene Text Image Detection and Recognition[J]. Expert Systems with Applications, 2020, 170(3):114549.

[11] Liu B, Xu X, Zhang Y. Offline Handwritten Chinese Text Recognition with Convolutional Neural Networks[J]. 2020.

[12] Liu X, Zhang R, Zhou Y, et al. Scene Text Detection with Feature Pyramid Network and Linking Segments[C]// 2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2020.