**Section 3. AI (Intelligent Systems. Machine learning, big data.).**

**UDC 004.075**

**Xu Jiashu , Sergii Stirenko.**

# FACIAL EXPRESSION RECOGNITION SYSTEM BASED ON GAN NETWORK DATA AUGMENTATION

In recent years, facial expression recognition based on deep learning networks has developed. However, the deep model needs enough labelled data for training. Currently, the amount of data in most facial expression recognition databases is very limited. In order to solve the problem of insufficient data sets and unbalanced data sets, this paper designs a GAN-based method to generate more images to expand the training set and solve the problem of unbalanced data sets. In experiments, CNN's mainstream framework was used to verify the effect of enhanced data. The classification accuracy of the trained model on the original FER2013 data set is 62.17%, the classification accuracy after traditional data enhancement is 66.89%, and the classification accuracy after GAN data enhancement is 69.47%. The accuracy of the original CK + dataset is 90.15%, the accuracy of the traditional data after augment is 93.6%, and the accuracy of the GAN data after augment is 95.14%. Finally, the trained model is used in the design of OpenCV-based expression recognition UI system.

**Keywords:** Facial expression recognition, Deep learning, GAN, CNN.

Fig.: 4. Tabl.: 4. Bibl.: 7.

**Problem statement.** Deep neural networks are inseparable from a large amount of data, but for the facial expression dataset, the standard facial expression database CK + has only 327 labelled video sequences [1], and the earlier database JAFFE only has 213 facial expression pictures [2]. Another common problem with facial expression databases: data imbalance. The amount of happy data in the RAF database is 17 times that of fear, and there are only 18 contempt expression data in the CK + database, in the FER2013 the number of happy images dataset is 16.5 times the amount of disgust images. Therefore, image generation technology is needed to solve the problem of data imbalance while expanding the data set. In order to solve this problem, this paper proposes a data augmentation method based on GAN.

**Design of Network Model Based on GAN.** In the original generative adversarial network [3] model, because the input is a random noise signal, we cannot specify what kind of image this generative adversarial network model generates. And Mirza [4] proposed a strategy to add conditional models to the generative adversarial network. The main method is to add label information conditions to the generative model and discriminative model as a condition to restrict the direction of the image

generated by the generative model. For example, in this experiment, in addition to the random noise signal, we also put the original expression picture into the generator.

The design of the generator network takes into account the problem of image conversion. The input and output facial expressions of the generator network are different, but the basic distribution structure of the two is the same. Therefore, the distribution structure of the input samples should be similar to the distribution structure of the output samples. Let the generator use a way to bypass the information bottleneck, we added skip connections according to the "U-net" network structure [5]. Specifically, skip connections are added between the *i-th* layer and the *n-i* layer, and n is the total number of layers in the network. The generator model structure is shown in ***Fig. 1***. The generator network is divided into an encoder network and a decoder network, there are skip connections between the encoder network and the decoder network. On the left is the input image x. After the image is processed by the encoder, a representation of the features of the face image is obtained, denoted as $f(x)$, and then the face image features *f (x),* the facial expression C input into the decoder to generate that expression image *G(x,y)*, condition y represents the encoded information of the facial expression C represented by one-hot vector.

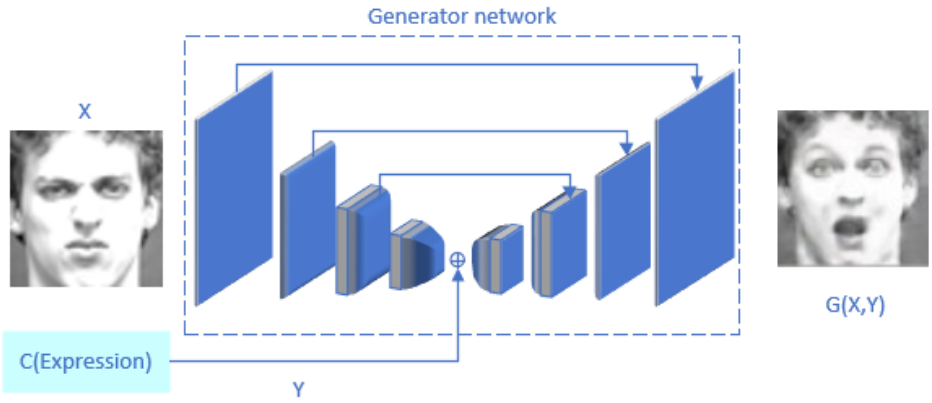


***Fig. 1*** Generator model structure

There are two discriminator networks as shown in ***Fig. 2***. The first discriminator network is trained by confronting the generator to determine the mapping relationship between different facial expressions. The input of the discriminator in the figure is x' real facial expression image, which is from the same person with the facial expression image x, G (x) is the facial expression image generated by the generation network, and C is specific facial expression condition information. The second discriminator network treats the data obtained from the uniform sampling as positive samples prior(f)and the facial image features processed by the encoder as negative samples f(x). The confrontation training between the two can make the facial image features meet the uniform distribution, thereby improving the quality of the generated image.

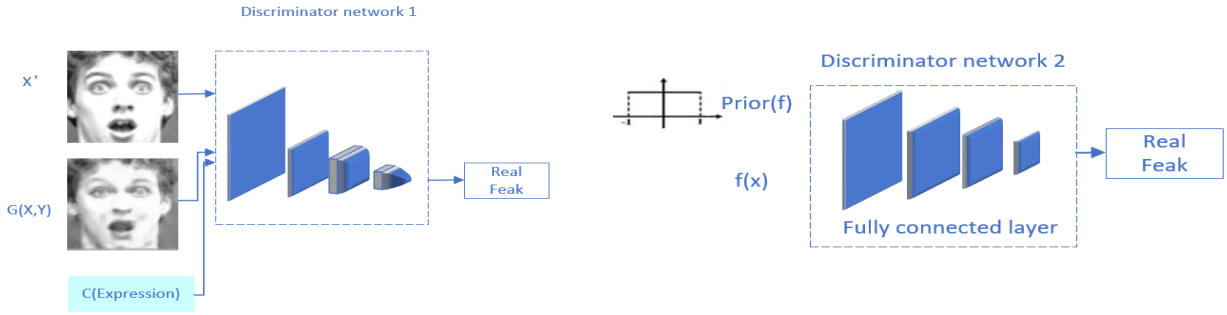


***Fig. 2*** Discriminator network

**Determination of the objective function.** In the experiment of facial expression data augmentation, Due to the imbalance of the facial expression data set, different facial expression images can be regarded as different facial image styles, so the facial expression images need to be converted from one domain to another. Therefore, condition information y is introduced, condition y represents the encoded information of the facial expression $C$. Where x represents the pre-processed real facial image, G (x, y) is the facial image with facial expression $C$ generated by the generator. Adding the condition y here to guide the generator to convert the facial expression image from one domain to another, which effectively solves the problem of imbalance in facial expression data. The real image satisfies the $P_d\ (x)$ distribution, and the discriminator uses the distribution information of the image to determine the authenticity of the image. Through the adversarial training between the generator network G and the discriminator network $D_1$, a model for face image conversion in another expression can be obtained. The training process satisfies Equation (1):

$$\min_{G}\max_{D_1} V(D_1,G) = E_{x,y\sim P_{d(x,y)}}[\log D_1(x,y)] + E_{x,y\sim P_{d(x,y)}}[\log(1-D_1(G(x,y),y))] \#(1)$$

Assuming that $prior(f)$ satisfies a certain prior distribution, $f^*\sim prior(f)$ represents the process of random sampling from the prior distribution. The samples sampled from the prior distribution are regarded as positive samples, and the face image features $f\ (x)$ processed by the encoder in the generator are regarded as negative samples. Through the adversarial training between the generator G and the discriminator $D_2$, a face image that satisfies uniform distribution and high quality can be obtained. The training process satisfies Equation (2):

$$\min_{G_{enc}}\max_{D_2} V(D_2,G_{enc}) = E_{f^*\sim prior(f)}[\log D_2(f^*)] + E_{x\sim P_{d(x)}}\left[\log\left(1-D_2\left(G_{enc}(x)\right)\right)\right] \#(2)$$

**Data augmentation.** In the experiment, we first matched different expressions made by the same sample in the database, As shown in the first picture of ***Fig. 3***. The CK+ database uses 7 expressions to build an expression library, including anger, contempt, disgust, fear, happy, sadness, and surprise. The total number of samples is 981, and they are divided into 3 groups (training, verification, and testing) according to

the ratio of 3: 1: 1. The different expression images of the same person are in the same group. GAN-based data augmentation methods are more diverse, this method does not augment the images in one domain but generates images of other domains from the images of one domain. When a face image is changed from one expression to another, the real changes are actually the key areas such as the mouth, nose, and eyes, and most of the image content has not changed much. An example of the generated sample is shown in the second picture of ***Fig. 3***.


***Fig. 3*** Original image in CK + & Image generated by improved CGAN

Here we use the traditional method and the method of this article to expand the original data set, and then test the effect of data augmentation in the experiment. Table 4 shows the data volume of the CK + training set after applying two data augmentation schemes. Augment the FER2013 dataset in the same way, the number of images after augment is shown in Table 5.

*Table 4*

**CK + training data set & augmented data set**

| **№** | **CK + training set** | | | | | | |
|---|---|---|---|---|---|---|---|
| | **anger** | **contempt** | **disgust** | **fear** | **happy** | **sad** | **surprise** |
| Original training set | 81 | 33 | 106 | 45 | 124 | 50 | 149 |
| Augmented by traditional method | 486 | 198 | 636 | 270 | 744 | 300 | 894 |
| Augmented method in this article | 504 | 440 | 540 | 456 | 564 | 464 | 596 |

*Table 5*

**Fer2013 training data set & augmented data set**

| № | FER2013 training set | | | | | | |
|---|---|---|---|---|---|---|---|
| | anger | disgust | fear | happy | sad | surprise | normal |
| Original training set | 2972 | 328 | 3073 | 5393 | 3646 | 2401 | 3719 |
| Augmented by traditional method | 5944 | 656 | 6145 | 10787 | 7292 | 4802 | 7438 |
| Augmented method in this article | 5393 | 5393 | 5393 | 5393 | 5393 | 5393 | 5393 |

**Experimental results and analysis.** The hardware environment of this experiment is as follows: Windows 10 operating system, i5-8250CPU, GPU acceleration using Google Colab experimental platform. The machine learning framework uses TensorFlow + Keras. In order to further verify the effectiveness of the data augmentation method in this paper, the training sets in Tables 1 and 2 were trained, the recognition rate of the verification set was recorded, and the recognition rate was calculated on the test set. This article chose Xception [6] for evaluation, which is another improvement of the Inception v3 proposed by Google after Inception. It mainly uses depthwise separable convolution to replace the original convolution operation in Inception v3 [7]. The experimental results are shown in Table 6. The data augmentation method in this paper can effectively improve the recognition accuracy of the model. In the Ck + dataset experiment, the recognition rate is improved by about 4% compared with the unaugment model. In the FER2013 dataset experiment, the recognition rate is improved by about 7% compared with the unaugment model. Both experimental results show that our method is superior to traditional data augmentation methods.

*Table 6*

**Data augmentation experiment results comparison (recognition rate)**

| ***Method*** | ***dataset*** | ***Validation set*** | ***Test set*** |
|---|---|---|---|
| No data augmentation | CK+ | 96.36% | 90.15% |
| Traditional augmentation | CK+ | 97.02% | 93.60% |
| GAN augmentation | CK+ | 97.58% | 95.14% |
| No data augmentation | FER2013 | 63.11% | 62.17% |
| Traditional augmentation | FER2013 | 67.70% | 66.89% |
| GAN augmentation | FER2013 | 74.89% | 69.47% |

**Implementation of Facial Expression Recognition System.** The expression recognition model after data augmented based on GAN used here. In fact, the purpose of deep learning is to apply, so the OpenCV-based UI interface is designed with PyQt5. This system has two modes, you can use the real-time camera to recognize facial expressions, or you can use local imported pictures to recognize. The UI interface of the facial expression recognition system is shown in the Fig. ***4***.

**Conclusion**. This paper uses the improved GAN to generate paired images under other expressions from face images under one expression, and expand the dataset from the semantic level. The experiments on the CK + and FER2013 databases show that the method in this paper can effectively solve the overfitting problem of small sample databases and improve the model recognition rate. Compared with the traditional method, the method in this paper makes the model pay more attention to semantic features and can improve the generalization ability of the model to a certain

extent. At the same time, the method in this paper can also be used as a reference for solving imbalanced data sets.

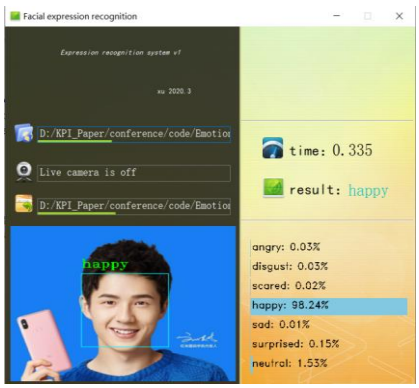


***Fig. 4*** UI interface of the facial expression recognition system.

## References

1. Lucey P, Cohn J F, Kanade T, et al.The extended Cohn Kanade dataset (CK) : a complete dataset for action unit and emotion- specified expression.Computer Vision & Pattern Recognition Workshops, 2010.

2. Lyons M J, Akamatsu S, Kamachi M, et al.The Japanese female facial expression (JAFFE) database. 3rd International Conference on Automatic Face and Gesture Recognition,pp.14-16,1998.

3. Goodfellow I. J., Pouget-Abadie J., Mirza M, et.al. Generative Adversarial Nets. NIPS, 2014.

4. Mehdi Mirza, Simon Osindero. Conditional Generative Adversarial Nets. NIPS, 2014. arXiv:1411.1784.

5. Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015 arXiv:1505.04597.

6. François Chollet, Xception: Deep Learning with Depthwise Separable Convolutions. Google, 2017.

7. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, et al. Rethinking the Inception Architecture for Computer Vision. 2015, arXiv:1512.00567.

## Autors

**Jiashu Xu** – master student, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: 392517729x@gmail.com

**Sergii Stirenko** (supervisor) – professor, Department of Computer Engineering, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.

E-mail: sergii.stirenko@gmail.com