Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN IMPROVED DEEPFAKE DETECTION METHOD BASED ON CNNS

DAFENG GONG[1], YOGAN JAYA KUMAR[2], ONG SING GOH[3], CHOO YUN HUOY[4], ZI YE[5], WANLE CHI[6]

[1]School of Artificial Intelligence, Wenzhou Polytechnic, China
[1,2,3,4,5,6]Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia

E-mail: [1]289133894@qq.com, [2]yogan@utem.edu.my, [3]osgoh88@gmail.com, [4]huoy@utem.edu.my, [5]yezi1022@gmail.com, [6]358455713@qq.com

ABSTRACTION

Today's image generation technology can generate high-quality face images, and it isn't easy to recognize the authenticity of the generated images through human eyes. This study aims to improve deepfake detection, a face swapping forgery, by absorbing the advantages of deep learning technologies. This study generates a unified and enhanced data set from multiple sources using spatial enhancement technology to solve the problem of poor detection performance on cross-data sets. Taking the advantages of Inception and ResNet networks, new deepfake detection architecture composed of 20 network layers is proposed as the deepfake detection model. To further improve the proposed model, hyperparameter values are optimized. The experiment result shows that the proposed network significantly enhanced over the mainstream methods, such as ResNeXt50, ResNet101, XceptionNet, and VGG19, in terms of accuracy, loss value, AUC, numbers of parameters, and FLOPs. Overall, the methods introduced in this study can help to expand the data set, better detect deepfake contents, and effectively optimize network models.

Keywords: *Face Swapping, Cross Data set, Deepfake Detection, Data Enhancement, Optimized Hyperparameters*

1. INTRODUCTION

In recent years, with the rapid development of science and technology and the rapid improvement of computing power, the theory and technology of artificial intelligence (AI) are becoming more and more mature. Now it has been widely used in the fields of finance, medical treatment, urban service, industrial manufacturing, and life service. Artificial intelligence technology is leading a new round of all-around industrial transformation and promoting the human world into an intelligent era. Machine learning (ML) is the core research field of artificial intelligence. According to the statistics of Deloitte's white paper on global AI development published in 2019[1], 89% of AI patent applications and 40% of AI related patents belong to machine learning. As a hot research direction in machine learning, deep learning (DL) provides a strong technical support for the innovation of computer vision, unmanned driving, natural language processing and speech recognition. However, deep learning technology not only leads a new wave of artificial intelligence, but also poses potential threats to personal privacy data, social stability and national security. Since 2017, voice fraud[2] supported by "deep forgery" or "deepfake" technology has attracted wide attention all over the world, such as "face swapping" video incidents that spoof political figures[3] and public figures[4] have emerged in endlessly[5], causing very bad negative effects[6], and even have indirectly lead to the military coup in Gabon Republic, a country on the west coast of Central Africa[7].

The deep learning model is dependent on the specific data distribution. Through training, it can accurately judge the given serious forgery video data set, but the detection accuracy of the cross-data set will decrease. There are many data sets such as Kaggle, Facebook Detection[8], FaceForensics++[9] and Deepfake-TIMIT[10]. At present, different models are mostly built based on different data sets, and the results are only

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

compared with their own data sets, which has obvious limitations in the cross-data set[11].

The difference in video compression and resolution will also affect the detection result of the model. After training the forged video data set with a low compression rate, the accuracy rate will be significantly reduced when the high compression rate video is detected. Therefore, based on CNN[12], the RGB frame can be transformed into a new domain by combining the noise flow of the video frame or making Fourier transform on the video. The discrete signal can be decomposed into sinusoidal components with different frequencies, and this problem can be solved by analyzing its spectral phase diagram[13].

The current deep learning algorithm has some limitations. For example, we generally understand that increasing the depth of the CNN network can improve the accuracy, but can we learn a good network by simply stacking more layers? The problem is that gradient disappearance and gradient explosion are due to stacking more layers[14]. To solve the two problems, normalization, initialization, and middle layer standardization can be used. However, the accuracy rate saturates and decreases rapidly with the increased network depth. This phenomenon is called a degradation problem[15]. This problem is not caused by overfitting but by adding extra layers into the appropriate depth model, which leads to higher training errors. When the network is too deep, there will be a phenomenon where the training error decreases first and then increases. At this time, it is difficult to train the network, so residual learning is proposed. Residual network (ResNet)[15], which can be stacked to form a deep network, can solve the problem of network optimization, and the classification performance is improved. Therefore, in our study, we will propose a network structure based on ResNeXt[16] to solve better the problem of detecting deepfake.

The determination of hyperparameters in neural networks is a complex problem. At first, the hyperparametric optimization of the machine learning algorithm was completed by human experts. However, the manual search of hyperparametric parameters is inefficient and time-consuming and needs some expert experience to guide. The research on automatic hyperparametric optimization can be traced back to the 1990s[17]. It can automatically find the optimal hyperparametric configuration in the hyperparametric search space of the model without human expert intervention. It is the most fundamental problem in the field of automatic machine learning. Applying the automatic hyperparametric optimization method can effectively improve the performance of machine learning algorithms and improve the repeatability and fairness of scientific research[18]. The two most widely used automatic hyperparametric optimization methods are grid search and random search[19]. Among them, grid search searches for the optimal configuration by traversing all the hyperparametric combinations in the search space. This method can undoubtedly search for the optimal hyperparametric configuration. However, usually, there are a lot of parameters of convolutional neural networks, which leads to consuming of computing resources, and it is challenging to obtain ideal results under limited laboratory conditions. Random search searches for the optimal configuration by randomly sampling in the hyperparametric search space. Compared with grid search, this method can improve search efficiency, especially in the high-dimensional search space. However, it cannot guarantee to obtain the optimal hyperparametric configuration in a small number of sampling times. The search results have great randomness. Therefore, in this study, we will use a univariate approach to get optimized parameter values by sampling a fixed number of parameter values from a specified distribution[20]–[22].

In order to solve the above problems, the main contributions are as follows:
1) To build a unified and enhanced deepfake data set from multiple data resources.
2) To propose a DeepfakeNet model based on CNNs to detect deepfake efficiently.
3) To optimize hyperparameters values of DeepfakeNet.

2. RESEARCH OPERATIONAL FRAMEWORK

There are 5 phases in this operational framework. Phase 1: Preliminary research literature; Phase 2: Build a unified and enhanced data set from multiple data resources with the spatial transformation of images; Phase 3: Propose DeepfakeNet model based on CNNs to detect

deepfake efficiently; Phase 4: Obtain optimized hyperparameters values; Phase 5: Report Writing. The following sections will provide the details on each of these phases, such as Figure 1.

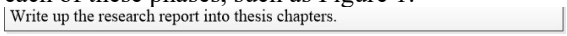

Figure 1: Research Operational Framework

2.1 Preliminary Research Literature

Finding and explaining research problems is one of the critical tasks that must be carried out in any research. This phase mainly introduces the preliminary research related to the study of the typical GANs and deepfake model generation mechanism. It includes background research and a literature review. It starts by identifying a wide range of problem areas until an in-depth investigation is conducted to solve the problems in specific research areas. In this study, we study the problem of deepfake detection. Through the expression of questions, we can decide the questions that need to be answered in the research in a particular way.

2.2 Build a Unified and Enhanced Data Set

In this phase, a unified and enhanced data set will be analyzed and built. As mentioned previously, current models have the problem of insufficient generalization[11]. In order to improve this deficiency, we must analyze and generate an enhanced data set from various data sets to train and test the new model. It involves a selection of data sets to be used for training and evaluation.

This research mainly uses the data sets, such as TIMIT[10], FaceForensics++[23], and Kaggle competitions[24]. For each pair of videos of TIMIT, different training models and fusion techniques are used to generate low-quality (LQ) and high-quality (HQ) videos. The input/output image size of LQ model is 64 × 64, while that of HQ model is 128 × 128.

For the data set of FaceForensics++, the fake face video is realized based on deep-faceswap method of self encoder model, and H.264 codec is used to synthesize videos with a compression ratio of 0 (C0), a compression ratio of 23 (C23) and a compression ratio of 40 (C40), namely, non-compressed fake face video, high quality (HQ) fake face video and low quality (LQ) fake face video. It is worth noting that the FaceForensics+ + data set defines the HQ and LQ facial videos according to the degree of video compression. In contrast the deepfake TIMIT data set represents the HQ and LQ facial videos according to the input/output image size of the training model.

This Kaggle data set includes a video data set with a compressed size of 470GB. The data set comprises 50 files equally divided by the original data set. Each file contains 100 video files and a pair of JSON files, including video number, video tag, data set, the partition of video, source video name, etc., as shown in Table 1.

Table 1: Kaggle JSON File Data Format

No	Attribute	Memo
1	Video_ID	Video file ID
2	Label	Video tags are divided into real / fake
3	Split	The data set is divided into train / test
4	Original	Source video file name

Public verification set: it is composed of 400 videos. When submitting the result data and files to Kaggle, the verification is mainly based on these 400 videos. In order to simplify the experimental calculation, this research primarily uses part of the data set, a total of 1200 video files as the training set and 400 as the test set and the verification set.

The data size and quality of the above three data sources are shown in Table 2.

Table 2: Analysis of 3 Data Sources

	FaceForensics++	TIMIT	Kaggle
Size	4000	1199	119154
Quality	1280 × 720 FHD: Compression Ratio of 0 (C0); HQ: Compression Ratio of 23 (C23); LQ: Compression Ratio of 40 (C40).	HQ:128 × 128; LQ: 64 × 64.	1280 × 720

According to our training and verification needs, we extract part of the data from these data sets and combine the following data enhancement techniques to obtain more effective data sets. These methods include:

1) Stretch size: In a specific scale range, the original image is appropriately enlarged or reduced without affecting the detection of the main content.

2) Rotate angle: Within a specific rotation angle

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

range, rotate the entire image without affecting the main content, and get images with different angles.
3) Change brightness: Images of the same subject, if the brightness is different, are also considered to be the same content.
4) Flip horizontal/vertical: The content of the same subject is the same after it is flipped horizontally or vertically.

After obtaining the original videos from the original data source, each video is split into a series of images. Then the four data enhancement methods mentioned above are used to obtain a unified, more extensive, and more effective data set. The data enhancement process is shown in Figure 2.

Figure 2: Data Enhancement Process

Table 3: Total Data after Enhancement

Operation	Increased Number of images
original images	104,481
scaling	417,924
rotating	1,985,139
changing color	940,329
flipping	104,481
clipping	313,443
total	**3,865,797**

It can be seen from Table 3 that the number of original data sets increased from 104,481 to 3,865,797 after the data enhancement of the five methods mentioned above, with a total expansion of 36 times. Such a large number of images will play a very positive role in training a convolutional neural network.

2.3 Propose DeepfakeNet Architecture

In this study, ResNeXt is used as the basis of feature extraction. Convolution networks mainly comprise convolution layers, full connection layers, pooling layers, and activation functions, but different composition methods will form other network frameworks. The traditional way to improve the model's accuracy is to deepen or widen the network. However, with the increase in the number of hyperparameters (such as the number of channels and filter size), the difficulty of network design and computational overhead will increase. Therefore, the ResNeXt structure proposed in this study can improve the accuracy without increasing the parameter complexity while reducing the number of hyperparameters. This idea adopts the idea of VGG stack and the idea of concept split transform merge at the same time, but it has robust scalability. It can be considered as increasing the accuracy without changing or reducing the complexity of the model.

Figure 3: Network Structure of Face Swapping Detection

As shown in Figure 3, the input face image is preprocessed during feedforward, and the image size is scaled to 256 × 256 × 3. DeepfakeNet processes the face image to get a feature map with the size of 8 × 8 × 2048, and then the feature vector with the size of 1 × 1 × 2048 is obtained after processing. Finally, the prediction probability value is obtained through the full connection layer and Softmax. Accordingly, in this study, we will propose DeepfakeNet network to get the optimal result.

In this phase, DeepfakeNet can simplify the network structure and reduce the parameters and computation amount to achieve a better accuracy. This network structure consists of 5 stages:

Stage 1: This layer is composed of basicblock. The input image is preprocessed. After it passes through the convolution layer, the batch normalization (BN) layer, the ReLU activation function, and then the max-pooling layer, the output image with the shape of (64, 112, 112) is obtained.

Stage 2: This layer is composed of bottlenecks. In each bottleneck, a convolution layer is added between the input and output. The output image with the shape of (256, 56, 56) is obtained.

Stage 3: In each bottleneck, a convolution layer is added between the input and output, and there is also one downsample (generate a smaller image) in this layer and the output image with the shape of (512, 28, 28) is obtained.

Stage 4: The structure of this layer is similar to that of stage 3. But this layer has more channels and a smaller output size with the shape of (1024, 14, 14).

Stage 5: The structure of this layer is similar to that of stage 3. But this layer also has more channels and a smaller output size with the shape of (2048, 7, 7). Finally, the output is "true" or "false".

2.4 Obtain Optimized Hyperparameter Values

We will enhance the proposed DeepfakeNet network by modifying the full connection layer and hyperparameters to perform better. The process of setting hyperparameter values requires expertise and extensive experiments. It is difficult to set appropriate parameter values, especially learning rate, batch size, epochs, dropout, and sample ratio. The deep learning model is full of hyperparameters, so finding the optimal configuration of these parameters in such a high-dimensional space is not a trivial challenge. In subsequent experiments, this research will discuss and optimize five hyperparameters of the learning rate, batch size, epochs, dropout, and sample ratio. The univariate principle is used for multiple parameter variables to find the best value for each hyperparameter.

2.5 Report Writing

Our research operational framework final phase ends with report writing. The report concludes the background and analysis of the research findings and provides the directions for future work.

3. EXPERIMENTAL EVALUATION

In this section, we describe in detail the evaluations that will be carried out for each experimental phase in this research. In the test, for the input video, due to the small time span of the test video, a certain number of frames in the video are randomly selected for prediction, and the results are statistically analyzed to predict the results of the whole video. For the test video, the AUC of each prediction value is calculated. The statistical analysis calculates positive and negative samples' prediction, accuracy, and recall. After continuous experiments, we will get the best network architecture.

In the field of machine learning, there are many indexes used to evaluate a model. Several concepts involved are as follows:

TP (True Positive): the number of positive is truly positive, and the judgment is correct. That is, the positive prediction is positive.

FN (False Negative): the number of positives is a false negative. If it is wrong and negative, it means that the positive is judged as negative.

FP (False Positive): a negative number is wrongly judged as positive.

TN (True Negative): the negative is judged as the number of negatives. If the judgment is correct and negative. That is, the negative is judged as negative.

Accuracy: The accuracy rate refers to how many of the judgments are correct, that is, the positive judgment is positive, and the negative judgment is negative; there are $TP + TN + FN + FP$ in total, so the accuracy rate is:

$$\mathrm{P} = \frac{(TP+TN)}{(TP+TN+FN+FP)} \quad (1)$$

ROC: ROC curve can well describe the generalization performance of the model. AUC is the area under the ROC curve. The larger the AUC value, the better the performance of the model. The calculation equation of AUC is as follows, and *m* is the number of samples:

$$A_{AUC} = \frac{1}{2}\sum_{i=1}^{m}\left(F_{PR}^{(i+1)} - F_{PR}^{(i)}\right) \times \left(T_{PR}^{(i)} - T_{PR}^{(i+1)}\right) \quad (2)$$

FLOPs: FLOPs is the abbreviation of floating point operations, which means floating-point operands and is understood as computation. It can be used to measure the complexity of an algorithm or model. The smaller the value, the better the algorithm or model.

Assuming that the sliding window is used to realize convolution and the nonlinear computational overhead is ignored, one of the FLOPs of convolution kernel is:

$$\mathrm{FLOPs} = 2\mathrm{HW}(C_{in}K^2 + 1)C_{out} \quad (3)$$

Among them, *H, W* and C_{in} are the height, width, and number of channels of the input characteristic image (that is the input image), *K* is the core width and cout is the number of output channels.

Number of parameters: In a neural network, the number of parameters has a great relationship with the amount of calculation. Generally, the fewer parameters, the less calculation, and the faster the model runs. On the contrary, it runs slower.

Journal of Theoretical and Applied Information Technology
15th September 2022. Vol.100. No 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In CNN, each layer has two parameters: weight and deviation. The total number of parameters is the sum of all weights and deviations. Calculation equations are:

$$W_c = K^2 \times C \times N \quad (4)$$

$$B_c = N \quad (5)$$

$$P_c = W_c + B_c \quad (6)$$

W_c is the umber of weights of the Conv Layer. B_c is the umber ofbiases of the Conv Layer. P_c is the umber of parameters of the Conv Layer. K is the size (width) of kernels used in the Conv Layer. N is the number of kernels. C is the number of channels of the input image.

4. EXPERIMENTAL ANALYSIS

We will conduct experiments with the enhanced data set using random hyperparameters values. At the beginning of the experiments, we preset the range of hyperparameters values, which are set based on [25]–[27] and our practical experience, then select a corresponding value randomly, such as Table 4, and then train the model.

Table 4: Preset Random Hyperparameters Values

Params	Value Range	Selected
batch size	[16, 32, 64, 128, 256, 512]	128
epochs	[15, 20, 25, 50, 100, 200, 300]	100
dropout	[0.6, 0.7, 0.75, 0.8, 0.85, 0.9]	0.7
learning rate	[0.000005, 0.00001, 0.00003, 0.00005, 0.00007, 0.0001]	0.00005
Sample ratio	[0.1, 0.25, 0.5, 0.75, 1.0, 2.0, 5.0]	1.0

Figure and Figure 5 show that as the epoch increases, the loss of DeepfakeNet decreases to 9.72%. And the accuracy increases to 93.97%.

Figure 4: DeepfakeNet Epoch-Loss Curve

Figure 5: DeepfakeNet Epoch-Accuracy Curve

Looking back at the curve change trends in Figure 4 and Figure 5, we can see that there is still room for improvement in this model, and better results may be achieved with different hyperparameter values. A univariate approach is used to get optimized hyperparameter values by sampling a fixed number of parameter values from a specified distribution. After hyperparameters values were optimized, we got Figure 6 and Figure 7.

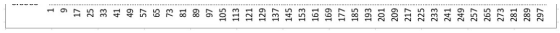

Figure 6: DeepfakeNet Epoch-Loss Curve

Figure 7: DeepfakeNet Epoch-Accuracy Curve

Figure and Figure show that with the epoch increased, the loss value of DeepfakeNet decreases (3.42%), and the accuracy increases (98.69%). Compared with Figure 4 and Figure 5, the improvements in loss value and accuracy are 6.3% and 4.72%, respectively. They demonstrate that the model becomes better. Compared with mainstream methods, DeepfakeNet has better accuracy than ResNeXt50, ResNet101, XceptionNet, and VGG19 (Table 5).

Table 5: Accuracy Comparison

Model	Accuracy (%)
DeepfakeNet	98.69
ResNeXt50	94.36
ResNet101	93.78
XceptionNet	92.03
VGG19	80.22

*Table **Error! No text of specified style in document.**: AUC Comparison*

Model	AUC
DeepfakeNet	0.96
ResNeXt50	0.94
ResNet101	0.93
XceptionNet	0.92
VGG19	0.83

Table 7: Number of Parameters and FLOPs

Model	Params	FLOPs
DeepfakeNet	10.87×10^6	2.05×10^9
XceptionNet	22.8×10^6	3.81×10^9
ResNeXt50	25.08×10^6	4.27×10^9
ResNet101	44.6×10^6	7.85×10^9
VGG19	145.77×10^6	19.67×10^9

Table 6 describes that AUC of DeepfakeNet is the highest. Table also shows that DeepfakeNet achieved better performance, and it increases at least by 46.19% and 52.32% for FLOPs and parameters, respectively.

The experimental results also show that appropriately reducing the network layers and optimizing hyperparameters values for complex deep neural networks can help detect deepfake.

However, we can only do a few experiments due to the limitation of experimental equipment, such as the hardware server performance and the data set's size. In the future, we will continue to improve and optimize the data sets and models to achieve better results.

5. CONCLUSION AND FUTURE WORK

DeepfakeNet can obtain better deepfake detection performance than others. The experiment shows that the method in this study dramatically reduces loss rate, and high accuracy and AUC were obtained in cross-data set detection. The experiments also illustrate that DeepfakeNet has good generality.

This study designs and studies the generated image forensics algorithm based on deep learning and taking the rendered image as the research object. The experimental results verify the effectiveness of the proposed method. The following research prospects will be put forward:

1) Research the latest technology of face swapping and deepfake detection.
The situation of attack and defense confrontation and competitive development of the two technologies determines that the research on them should stand at the forefront of technology, master the latest technology at the first time, formulate strict standard use specifications, study practical and effective prevention algorithms, and reduce the possible risks caused by illegal abuse. By studying the latest forgery generation technology, the corresponding forgery data set is established, and the generation methods of the data set should be diverse and have a considerable amount of data scale.

2) Improve the robustness of the generated image detection model.
The development of the Internet makes information spread quickly in the network. Still, due to the storage cost, most images will go through image post-processing, such as image compression, before they lay on the Internet. Some researchers have noticed these phenomena and tested the detection model. Experimental results show that image post-processing will significantly weaken the detection ability of the image forensics model. Various image post-processing operations need to be considered in the actual detection scene of the generated image, and a more robust detection model can be universal. In the future, the forensics influence of image post-processing can be considered when designing the forensics model.

ACKNOWLEDGMENT

This work was supported by Wenzhou Polytechnic in 2020 (No. WZYYFFP2020005) and Zhejiang Provincial Natural Science Foundation of China (LY20G010007). We would also like to thank Universiti Teknikal Malaysia Melaka (UTeM) for the collaboration.

REFERENCES

[1] Deloitte, “Global artificial intelligence industry whitepaper,” 2019.
[2] C. Stupp, “Fraudsters Used AI to Mimic CEO’s Voice in Unusual CybercrimeCase,” Aug-2019. [Online]. Available: https://www.wsj.com/articles/fraudsters-useai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402.
[3] N. Kang, “Deepfake: The Good, The Bad and the Ugly,” May-2019. [Online]. Available: https://medium.com/twentybn/deepfake-the-good-the-bad-and-theugly-8b261ecf0f52.
[4] Q. Wong, “Deepfake video of Facebook CEO Mark Zuckerberg posted on Instagram,” Jun-2019. [Online]. Available: https://www.cnet.com/news/deepfakevideo-of-facebook-ceo-mark-zuckerberg-posted-on-instagram/.
[5] JosephFoley, “deepfake examples that terrified and amused the internet,” Sep-2019. [Online]. Available: https://www.creativebloq.com/features/deepfakeexamples.
[6] M. Brundage *et al.*, “The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation,” *arXiv Prepr.*, 2018.
[7] A. Breland, “The Bizarre and Terrifying Case of the ‘Deepfake’ Video that Helped Bring an African Nation to the Brink,” Mar-2019. [Online]. Available: https://www.motherjones.com/politics/2019/0

3/deepfake-gabonali-bongo/.
[8] P. K. Nicholas Dufour, Andrew Gully, "Deepfakes detection dataset," Oct-2019. [Online]. Available: https://deepfakedetectionchallenge.ai/.
[9] R. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper, "Unmasking DeepFakes with simple Features," *arXiv Prepr.*, 2019.
[10] P. Korshunov and S. Marcel, "DeepFakes: a New Threat to Face Recognition? Assessment and Detection," *arXiv Prepr.*, pp. 1–6, 2018.
[11] H. Yongjian, "Deepfake Videos Detection Based on Image Segmentation with Deep Neural Networks," *J. Electron. Inf. Technol.*, vol. 43, no. 1, pp. 162–170, Jan. 2021, doi: 10.11999/JEIT200077.
[12] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-generated images are surprisingly easy to spot... for now," *arXiv Prepr.*, 2019.
[13] R. Liang *et al.*, "A Survey of Audiovisual Deepfake Detection Techniques," *Journal of Cyber Security*. 2020, doi: 10.19363/J.cnki.cn10-1380/tn.2020.02.01.
[14] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, 2015.
[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, doi: 10.1109/CVPR.2016.90.
[16] K. H. S Xie, R Girshick, P Dollár, Z Tu, "Aggregated Residual Transformations for Deep Neural Networks," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1492–1500, 2017.
[17] G. J. R Kohavi, "Autmatic Parameter Selection by Minimizing Estimated Error," *Mach. Learn. Proc.*, 1995, doi: 10.1016/B978-1-55860-377-6.50045-1.
[18] J. V. F Hutter, L Kotthoff, "Automated Machine Learning - Methods, Systems, Challenges," 2019, doi: 10.1007/978-3-030-05318-5.
[19] Y. Bergstra, James, Bengio, "Random Search for Hyper-Parameter Optimization," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 281–305, 2012, doi: 10.1016/j.chemolab.2011.12.002.
[20] C. Jack, "深度学习调参技巧," Oct-2016. [Online]. Available: https://www.zhihu.com/question/25097993/answer/127472322.
[21] L. Zhihua, "积神经网络(CNN)的参数优化方法," Mar-2018. [Online]. Available: https://www.cnblogs.com/bonelee/p/8528863.html.
[22] H. L. Liangping Ding, Zhixiong Zhang, "影响支持向量机模型语步自动识别效果的因素研究," 现代图书情报技术, pp. 16–23, 2019, doi: 10.11925/infotech.2096-3467.2019.0045.
[23] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1–11, 2019.
[24] B. H. Anthony Goldbloom, "DeepFake Detection Challenge," Mar-2020. [Online]. Available: https://www.kaggle.com/c/deepfake-detection-challenge.
[25] J. M. Dmytro Mishkin, Nikolay Sergievskiy, "Systematic evaluation of CNN advances on the ImageNet," *Comput. Vis. image Underst.*, pp. 11–19, Jun. 2016, doi: 10.1016/j.cviu.2017.05.007.
[26] Nanqi, "卷积神经网络参数设置," Oct-2018. [Online]. Available: https://blog.csdn.net/nanqi123/article/details/83537097.
[27] Wqz, "卷积神经网络中的参数设定," Mar-2020. [Online]. Available: https://blog.csdn.net/qq_38684229/article/details/105091254.