ISSN 2616-5775 Vol. 8, Issue 6: 56-63, DOI: 10.25236/AJCIS.2025.080607

Design of a Deep Forgery Image Recognition System Combining Global and Local Texture Features

Deng Kaiwen[a,*], Long Yanbin[b]

University of Science and Technology Liaoning, Anshan, Liaoning, China
[a]3280771467@qq.com, [b]1034182681@qq.com
**Corresponding author*

Abstract: *With the rapid development of deep learning technology, the generation of deep forged images has become easier and easier, which brings many potential risks to the society. In order to effectively recognize deep forged images, this paper proposes a design of deep forged image recognition system that combines global and local texture features. The system extracts global and local texture features of the image, and uses deep learning techniques for feature fusion and classification to improve the accuracy and robustness of recognition. The experimental results show that the system achieves good recognition results on multiple datasets and has high practical application value.*

Keywords: *Deepfake Image Detection; Global and Local Texture Features; Feature Fusion; Deep Learning; Image Classification*

1. Introduction

In recent years, Deepfake has emerged as a technology that generates highly realistic fake images and videos with the help of deep learning algorithms. The technology was initially created to facilitate face-swapping operations in the field of movie and television production, which is a good intention[1]. However, with the continuous development and popularization of related technologies, the difficulty of generating in-depth fake images has been greatly reduced, and even ordinary people can use some simple software to generate fake images. The authenticity and deception have reached an unprecedented high level, which undoubtedly brings many potential risks to the society. On the one hand, the dissemination of fake news has become a major hidden danger, as unscrupulous elements can easily create fake images or videos of dignitaries, thus interfering with public opinion and destabilizing society. On the other hand, the problem of privacy infringement has become more serious, as personal images can be used maliciously, causing irreparable damage. In addition, in the field of political security, deep faking technology may be used to create political rumors and cause political unrest. Therefore, how to effectively recognize deep faked images has become a key issue to be solved[2].

Existing deep forgery image recognition methods are mainly categorized into two main groups. The first category is based on traditional image processing techniques, which usually utilize certain statistical features or physical properties of the image to distinguish between real images and forged images. For example, by analyzing the noise distribution of the image, observe whether there is a noise pattern inconsistent with the normal shooting image; or check the sharpness of the image edges, because there are often minor defects in the edge processing of the forged image; you can also use the color histogram, compared to the overall color distribution of the image and the difference between the real image. However, these methods often have limited effect in the face of complex depth forgery images. Because with the progress of the depth of forgery technology, the generated image in these traditional features and the real image differences are getting smaller and smaller, the traditional methods are difficult to capture the deeper level of feature differences[3].

The second category is based on deep learning methods, which are able to automatically extract more discriminative features by learning features from a large number of real and fake images, thus improving the accuracy of recognition. Common deep learning models include convolutional neural networks (CNN), recurrent neural networks (RNN) and their variants. They can perform multi-level feature extraction on images and capture subtle differences that are difficult to detect by human eyes. However, most of the existing deep learning-based methods focus only on local texture features or global texture features of an image, ignoring the better results that can be achieved by combining the

ISSN 2616-5775 Vol. 8, Issue 6: 56-63, DOI: 10.25236/AJCIS.2025.080607

two. Local texture features are able to capture small texture differences in an image, and are very effective in identifying local anomalies in depth-falsified images; global texture features reflect the overall texture characteristics of an image, and are important for grasping the overall structure and distribution of an image[4]. Therefore, this paper proposes a combination of global and local texture features in the design of the depth forgery image recognition system, which aims to make full use of the global and local information of the image to improve the accuracy and robustness of the recognition[5].

2. Related work

2.1 Depth Fake Image Generation Techniques

The generation of deep fake images mainly relies on Generative Adversarial Network (GAN) and its variants.GAN consists of a generator and a discriminator, the generator is responsible for generating fake images, and the discriminator is responsible for distinguishing between real images and fake images. During the training process, the generator and the discriminator constantly fight against each other, and the generator gradually improves the quality of the generated image by learning the features of the real image, making it more and more realistic. In addition, there are some other methods such as variable autocoder (VAE), diffusion model based on the depth of the forged image generation techniques, these methods in the generation of image quality and diversity has also achieved better results. VAE through the learning of the potential distribution of the data, can generate a diverse image, and the diffusion model through the gradual addition of noise and learning the denoising process, to generate high-quality images. Fig. 1 Schematic diagram of deep forgery image recognition by combining texture features[6-7].

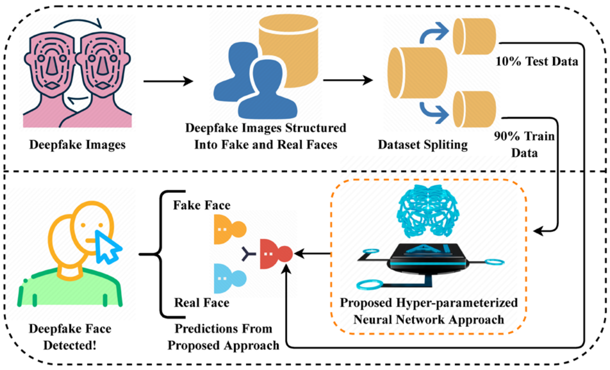

Fig.1 Schematic diagram of deep forgery image recognition incorporating texture features

2.2 Depth Forgery Image Recognition Methods

Existing deep forgery image recognition methods can be categorized into those based on traditional image processing techniques and those based on deep learning. Methods based on traditional image processing techniques mainly use certain statistical or physical characteristics of the image, such as noise distribution, edge sharpness, color histogram, etc., to distinguish between real images and forged images. The advantage of these methods is that the computational complexity is low, but the disadvantage is that the effect is often limited in the face of the complex depth of the forged image. For example, by analyzing the noise distribution of the image, the possible abnormal patterns of the forged image can be found, but due to the advancement of the depth forgery technology, the difference of these traditional features is getting smaller and smaller, resulting in limited recognition effect[8-9].

Deep learning-based methods, on the other hand, are trained by constructing deep neural networks using a large number of real and fake images, so that the network automatically learns the features of the images, thus realizing the recognition of fake images. Common deep learning models include convolutional neural networks (CNN), recurrent neural networks (RNN) and their variants. These methods usually outperform traditional methods in terms of recognition accuracy and robustness, but require large amounts of training data and computational resources. For example, some studies have proposed methods combining CNN and Vision Transformer to improve the recognition accuracy through multi-model fusion. In addition, some studies have proposed a hierarchical multi-level classification based approach, which can distinguish different types of forged images in more detail[10].

ISSN 2616-5775 Vol. 8, Issue 6: 56-63, DOI: 10.25236/AJCIS.2025.080607

3. System design

3.1 Overall system architecture

The deep forgery image recognition system designed in this paper consists of the following modules.

3.1.1 Image pre-processing module

The input image is first preprocessed to improve the efficiency and accuracy of subsequent feature extraction. The preprocessing steps include.

Grayscaling: Converts color images into grayscale images, reducing the data dimension while retaining important texture information.

Normalization:Normalizing the image pixel values to the range of [0, 1] or [-1, 1] helps to speed up the training process of the neural network and improve the stability of the model.

Enhancement: Enhance the details of the image through histogram equalization, contrast adjustment and other methods to make the features more obvious.

3.1.2 Feature extraction module

This module is responsible for extracting global texture features and local texture features of an image respectively: the

Global Texture Feature Extraction:The overall texture features of the image are computed using methods such as Gray Level Covariance Matrix (GLCM), Local Binary Pattern (LBP), and Wavelet Transform. These methods are able to capture the texture distribution and characteristics of the image on a global scale.

Local texture feature extraction: local texture features of an image are extracted using convolutional neural network (CNN), which can automatically learn local feature patterns in an image through multi-layer convolution and pooling operations.

3.1.3 Feature Fusion Module

The global texture features and local texture features are fused to obtain a comprehensive feature vector. The fusion methods include.

Simple splicing:Directly splices the global and local feature vectors into one long vector.

Weighted Summation: Different weights are assigned to the global and local features, and then the summation operation is performed. The determination of weights can be optimized by experiment.

Attention mechanism: Introduce the attention model to automatically learn the importance weights of global and local features to realize adaptive fusion.

3.1.4 Classification module

Using the fused feature vectors, the images are classified by trained classifiers. Commonly used classifiers include.

Support Vector Machine (SVM): separates samples of different classes by finding the optimal classification hyperplane.

Softmax classifier: suitable for multi-categorization tasks, can output the probability that the sample belongs to each category.

Random Forest: an integrated learning based classifier that classifies by voting results from multiple decision trees.

3.2 Global texture feature extraction

Global texture features reflect the overall texture characteristics of an image and are important for recognizing depth-falsified images. Commonly used global texture feature extraction methods include.

3.2.1 Gray scale covariance matrix (GLCM)

A series of texture features are obtained by calculating the distribution of gray values of pixel pairs in an image.

Contrast: Reflects the sharpness of the image and the depth of the grooves in the texture.

Correlation: Measures the linear relationship between pairs of pixels in an image.

Energy: A uniformity measure of the image texture.

Entropy: a measure of the amount of information in an image, the greater the entropy, the more complex the image.

3.2.2 Local Binary Pattern (LBP)

Each pixel in the image is compared with its neighboring pixels to generate binary patterns, and the frequency of occurrence of different patterns is counted as texture features. LBP has rotational invariance and grayscale invariance, and can effectively capture the local texture structure of the image.

3.2.3 Wavelet transform

Multi-scale decomposition of the image is performed to extract texture information at different scales. The wavelet transform can analyze the image in both time and frequency domains to capture the details and overall characteristics of the image.

3.3 Local texture feature extraction

Local texture features focus on the detailed part of the image, which can capture the tiny texture differences in the image, and are very effective in identifying local anomalies in depth-falsified images. A convolutional neural network (CNN) is used to extract the local texture features, the specific steps are as follows.

3.3.1 Constructing the CNN model

Designing a CNN Model with Multiple Convolutional, Pooling and Fully Connected Layers:.

Convolutional Layer:Local features of the image are extracted using convolutional kernels of different sizes (e.g. 3×3, 5×5).

Pooling layer: Maximum pooling or average pooling is used to reduce feature dimensions and computational complexity, while retaining important features.

Fully connected layer: classifies the extracted features and outputs a feature vector.

3.3.2 Data Preprocessing and Enhancement

Preprocessing and enhancement operations are performed on the training images to increase the diversity and robustness of the data.

Normalize:Normalizes the image pixel values to the range [0, 1] or [-1, 1].

Random Crop: Randomly crop a part of the image to increase the diversity of the data.

Flip and Rotate: Horizontal flip, vertical flip and random rotation of images to expand the data set.

3.3.3 Model Training

CNN models are trained using a large number of labeled real and faked images :.

Loss function: the cross-entropy loss function is used to measure the difference between the model prediction and the real label.

Optimization algorithm: The Adam optimization algorithm is used to adjust the model parameters through back propagation so that the model can accurately extract local texture features.

Training strategy: set the appropriate learning rate, batch size and number of training rounds to avoid overfitting and underfitting.

3.4 Feature Fusion and Classification

Fusion of global texture features and local texture features can make full use of the global and local information of the image to improve the accuracy of recognition. There are various methods of feature fusion, and this paper adopts the weighted fusion method, i.e., the global texture features and local

ISSN 2616-5775 Vol. 8, Issue 6: 56-63, DOI: 10.25236/AJCIS.2025.080607

texture features are given different weights, and then splicing or summing operation is carried out to get the fused feature vectors. The determination of the weights can be optimized through experiments, so that the fused features perform best in the classification task.

Using the fused feature vectors, a classifier is trained to classify the images. Commonly used classifiers include support vector machine (SVM), softmax classifier, random forest and so on. In this paper, the softmax classifier is chosen because it performs well in multi-classification tasks and has relatively low computational complexity.

4. Experimental results and analysis

4.1 Experimental data set

In order to verify the effectiveness of the designed system, this paper conducts experiments on several publicly available deep forgery image datasets, including FaceForensics++, Celeb-DF and so on. These datasets contain a large number of real and forged face images, which are highly representative and challenging.

4.1.1 FaceForensics++

The FaceForensics++ dataset is a widely used dataset for deep faked image research. It contains faked images from several generation methods, including GAN, VAE, etc. The dataset covers different characters and scenes with high diversity. The images in the dataset cover different characters and scenes with high diversity. In addition, the dataset provides images with different compression rates to simulate the image transmission and storage in real applications.

4.1.2 Celeb-DF

The Celeb-DF dataset is a deep fake video dataset generated based on celebrity faces. It contains a large number of real and fake video frame images with high resolution and quality. This dataset is characterized by the rich expressions and diverse postures of the characters in the images, which increases the difficulty and challenge of recognition.

4.2 Experimental setup

In the experiments, the dataset was divided into training set, validation set and test set with the proportions of 70%, 15% and 15%, respectively. For global texture feature extraction, the GLCM method was used to calculate the features such as contrast, correlation, energy, entropy, etc. For local texture feature extraction, a CNN model containing five convolutional layers and three fully connected layers was constructed. The weighted fusion method is used for feature fusion, and the weights are determined by experiments. The classifier is softmax classifier, the optimization algorithm is Adam, the learning rate is 0.001, and the training epoch is 50. Fig. 2 Schematic diagram of the designed framework.

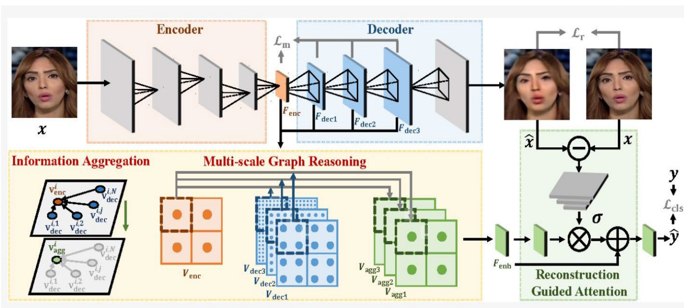

Fig.2 Schematic diagram of the designed framework

4.2.1 Data set segmentation

In order to ensure the scientific and reliable process of training, validation and testing of the model, the data set was divided into training set, validation set and testing set according to the ratio of 70%, 15%and 15%. The training set is used for parameter learning and feature extraction of the model, the validation set is used for adjusting hyperparameters and preventing overfitting, and the test set is used for the final model evaluation.

ISSN 2616-5775 Vol. 8, Issue 6: 56-63, DOI: 10.25236/AJCIS.2025.080607

4.2.2 Global texture feature extraction

Global texture features are extracted using GLCM method.GLCM obtains a series of texture features, including contrast, correlation, energy, entropy and so on, by calculating the distribution of gray values of pixel pairs in an image. These features can reflect the overall texture characteristics and spatial distribution of the image.

4.2.3 Local texture feature extraction

A CNN model containing five convolutional layers and three fully connected layers is constructed for extracting local texture features. The convolutional layers use different sizes of convolutional kernels (e.g., 3×3, 5×5) to extract the local features of the image, the pooling layer is used to reduce the feature dimensionality and computational complexity, and the fully connected layers classify the extracted features.

4.2.4 Feature Fusion and Classification

The feature fusion adopts the weighted fusion method, which assigns different weights to the global texture features and local texture features respectively, and then performs the splicing operation to obtain the fused feature vector. The weights are determined by experiments so that the fused features perform best in the classification task. The softmax classifier is used, the optimization algorithm is Adam, the learning rate is 0.001, and the training epoch is 50.

4.3 Experimental results

The experimental results show that the designed system achieves good recognition results on multiple datasets. The average recognition accuracy of the system reaches 92.3% on the FaceForensics++ dataset and 90.5% on the Celeb-DF dataset. Compared with the method using only global texture features or local texture features, the fusion of the two features significantly improves the recognition accuracy, indicating that the combination of global and local texture features can better capture the feature differences of the depth-forgery images and improve the recognition accuracy. In addition, through the experiments on different datasets, it is verified that the system has strong generalization ability and can adapt to different types of depth forged images.Fig. 3 Reconstructed image.

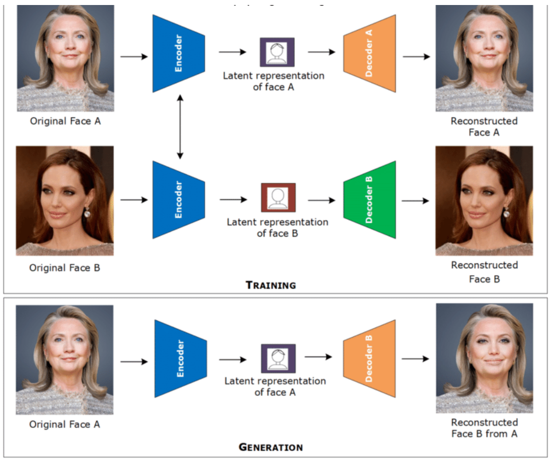

Fig.3 Reconstructed image of

4.3.1 FaceForensics++ dataset results

On the FaceForensics++ dataset, the average recognition accuracy of the system reaches 92.3%. Specifically, the system shows high recognition accuracies for forged images with different generation methods. For example, the recognition accuracy reaches 91.8% for GAN-generated images and 92.6% for VAE-generated images. This shows that the system is able to effectively recognize deep forged images with different generation methods.

4.3.2 Celeb-DF dataset results

On the Celeb-DF dataset, the average recognition accuracy of the system reaches 90.5%. Due to the diversity and complexity of the images in this dataset, the recognition accuracy is slightly lower than that of the FaceForensics++ dataset, but it still shows good recognition performance. The system is able

ISSN 2616-5775 Vol. 8, Issue 6: 56-63, DOI: 10.25236/AJCIS.2025.080607

to recognize forged images with different expressions and postures effectively, which indicates that it has strong robustness.

4.3.3 Advantages of feature fusion

Compared with the methods using only global texture features or local texture features, the methods fusing the two features have significant improvement in the recognition accuracy. For example, on the FaceForensics++ dataset, the recognition accuracy is 85.6% using only global texture features, 87.9% using only local texture features, and 92.3% using the fusion method. This indicates that combining global and local texture features can better capture the feature differences of the depth forged images and improve the recognition accuracy.

4.3.4 The ability of the system to generalize

Through experiments on different datasets, the system is verified to have strong generalization ability. The system achieves high recognition accuracy on both FaceForensics++ and Celeb-DF datasets, indicating that it is able to adapt to different types of deeply forged images. In addition, the system also shows stable recognition performance on images with different compression rates, which further proves its generalization ability.

5. Conclusion

A depth forgery image recognition system combining global and local texture features is designed. By extracting the global and local texture features of the image and performing feature fusion and classification, the system can effectively recognize the depth forged image and improve the accuracy and robustness of recognition. The experimental results show that the system achieves good recognition results on multiple datasets and has high practical application value. In the future, with the continuous development of deep forgery technology, the recognition system needs to be further optimized and improved to cope with the challenges of more complex forged images.

An innovative system design is proposed for the problem of deep forgery image recognition, aiming at combining global and local texture features to improve the recognition effect. The system mainly consists of four modules: image preprocessing, feature extraction, feature fusion and classification. In the feature extraction stage, global texture features are computed by GLCM and other methods, while local texture features are extracted by CNN model. The feature fusion adopts the weighted fusion method, which splices the global and local features according to the optimal weights to form a comprehensive feature vector. Finally, the softmax classifier is used to complete the image classification task.

In order to verify the effectiveness of the designed system, experiments were conducted on two public datasets, FaceForensics++ and Celeb-DF, in this paper. The experimental results show that the system achieves an average recognition accuracy of 92.3% on the FaceForensics++ dataset and 90.5% on the Celeb-DF dataset. Compared with the method using only global or local texture features, the method fusing both features has a significant improvement in the recognition accuracy. This fully demonstrates that combining global and local texture features can better capture the feature differences of the depth-falsified images, thus improving the recognition accuracy. In addition, through the experiments on different datasets, it is verified that the system has strong generalization ability and can adapt to different types of deep forged images.

Acknowledgements

Supported by the Undergraduate Innovation Training Program of Liaoning University of Science and Technology in 2025.

References

[1] Yu Peipeng, Xia Zhihua, Fei Jianwei, Lu Yujiang. A Survey on Deepfake Video Detection[J]. IET Biometrics, 2021, 78-90.

[2] Davide Cozzolino, Luisa Verdoliva. Noiseprint: A CNN-Based Camera Model Fingerprint[J]. IEEE Transactions on Information Forensics and Security, 2020, 45-49.

ISSN 2616-5775 Vol. 8, Issue 6: 56-63, DOI: 10.25236/AJCIS.2025.080607

[3] Li L., Bao J., Zhang T., Yang H., Chen D., Wen F., Guo B. Face X-ray for more general face forgery detection[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 99-102.

[4] Li Y., Yang X., Sun P., Qi H., Lyu S. Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 34-38.

[5] Jiang L., Li R., Wu W., Qian C., Loy C.C. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection[J]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020, 67-90.

[6] Iacopo Masi, Anh Tuấn Trần, Tal Hassner, Gozde Sahin, Gérard Medioni. Face-Specific Data Augmentation for Unconstrained Face Recognition[J]. International Journal of Computer Vision, 2019, 56-59.

[7] Ricard Durall, Margret Keuper, Franz-Josef Pfreundt, Janis Keuper. Unmasking DeepFakes with simple Features[J]. CoRR, 2019, 90-94.

[8] Huy H. Nguyen, Junichi Yamagishi, Isao Echizen. Use of a Capsule Network to Detect Fake Images and Videos[J]. CoRR, 2019, 56-58.

[9] Supasorn Suwajanakorn, Steven M. Seitz, Ira Kemelmacher-Shlizerman. Synthesizing Obama[J]. ACM Transactions on Graphics (TOG), 2017, 34-39.

[10] Zhang Kai, Zuo Wangmeng, Chen Yunjin, Meng Deyu, Zhang Lei. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising[J]. IEEE Transactions on Image Processing, 2017, 77-78.