IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received August 13, 2019, accepted August 27, 2019, date of publication September 10, 2019,
date of current version September 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940383

An Evaluation of Deep Learning-Based Computer Generated Image Detection Approaches

XUAN NI[1], LINQIANG CHEN[1], LIFENG YUAN[1,2], GUOHUA WU[1], AND YE YAO[1,3]

[1]School of Cyberspace, Hangzhou Dianzi University, Hangzhou 310018, China
[2]Anhui Provincial Key Laboratory of Network and Information Security, Wuhu 240002, China
[3]Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China

Corresponding author: Ye Yao (yaoye@hdu.edu.cn)

This work was supported by the Humanities and Social Sciences Foundation of Ministry of Education of China under Grant 17YJC870021.

ABSTRACT With the rapid development of Computer Graphics, the computer-generated images (CG) are almost as realistic as real photographs(PG) and it is difficult to distinguish between CG and PG accurately with the naked eye. Image is an important carrier for people to get information on a daily basis. However the spread of CG produced for malicious purposes may disrupt social order and even undermine social stability. Therefore, the accurate detection of CG and PG is of great significance. In this paper, we (1) introduce 11 approaches that apply deep learning to the implementations of CG detection, and divide them into 4 categories based on the network structure; (2) give an introduction to the available datasets; (3) design a series of experiments to test the detection performance of each approach,then analyze the experimental results; The experimental results show that most approaches can differentiate CG from PG, while the detection accuracy and efficiency of each model are different. Nevertheless none of these methods is valid when the images tampered by noise. Above all (4) summarize the problems and challenges in this field, and look forward to the trends in future research.

INDEX TERMS Computer-generated images and photographs, deep learning-based classification, digital image forensics, the state of art of detection approaches.

I. INTRODUCTION

Computer-generated image (CG) refers to the image generated by a computer using graphical processing tools. After computerized, CG is difficult to distinguish by the naked eyes. Vision is an important access to the 80% external information. Some CG has counterparts in the real world, but they can be completely different both in terms of content and semantic information. The progress on Computer Graphics has made the authenticity of images suspicious. At the same time, with the popularization of the Internet, courses on graphical processing tools can be seen everywhere on the Internet, and non-professionals can also make CG after studying briefly.

The development of graphical processing tools enables people to have a better experience in movies, games and other fields. Figure 1 shows an example of CG. If you don't make some explanations for this image, people might think it was taken by a camera. Obviously, high-quality CG can fool people's eyes easily. The use of CG with false information in criminal investigations, judicial trials, etc., may hurt innocent people. The rapid development of the Internet enables information to spread around the world overnight. Once widely distributed on the Internet, CG with false information may bring out the disorder.

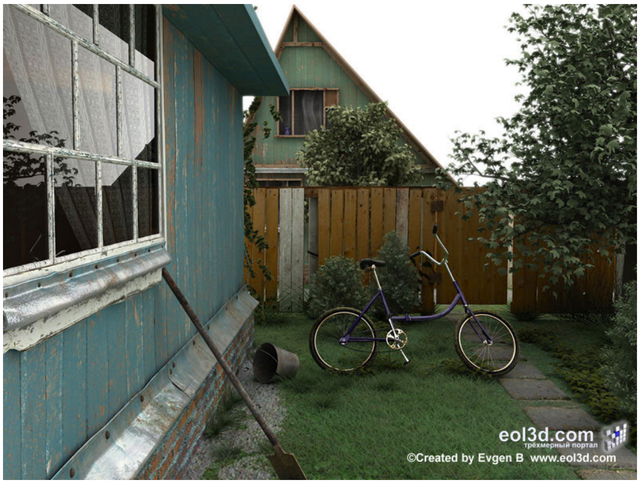


FIGURE 1. A sample of CG from the DSTok dataset. [1].

The associate editor coordinating the review of this manuscript and approving it for publication was Kim-Kwang Raymond Choo.

CG detection is to estimate whether an image is generated by a computer through extracting and analyzing the features of the image. Unlike other image detection tasks, CG and PG have similar appearances, which makes it more difficult to distinguish between them. Affected by the photographic equipment, PG will be injected with unique noise during the generation process. Even PGs that are captured with the same camera model will be slightly different. In addition, the photography habits will also have an impact on the final PG. PG is restricted by many objective factors, such as shooting time, location and climate, while CG is not. The content of CG can be either completely out of touch or consistent with reality. Therefore, compared to PG, the content of CG is more diversified. To improve the visual effect of CG, graphical processing tools simulate illumination, scenes, textures, etc., and inject unique noise into CG during this process. Based on this phenomenon, some traditional CG detection methods extract the statistical features of CG and PG to complete this detection task. However, the accuracy of traditional detection methods decrease as the growth in the quantity of CGs and the progress on Computer Graphics. Besides, the generalization capability of many published studies on CG detection is not ideal. Recently, deep learning technology has achieved great success in the computer vision field [2]–[5]. More and more scholars are trying to use deep learning technology to detect CG and PG.

CG detection is a rapidly developing research subject, and various new detection methods emerge one after another. Due to the wide application of deep learning, the deep learning-based CG detection approach has become a study hotspot in this field. Therefore, this paper mainly introduces the deep learning-based CG detection approach. This paper consists of the following sections. In section 2, we introduce the basic steps of image detection and some traditional detection approaches. Then, we divide the deep learning-based CG detection approaches into four categories according to their network structure and introduce in detail. And the experiment is presented in section 4. In section 5, we summarize the problems and challenges of research in this field and look forward to the future research direction.

II. TRADITIONAL CG DETECTION APPROACHES

Most of CG detection approaches that proposed by researchers follow a fixed process, as shown in Figure 2. Given an original image, what we need to do is to extract image features that can be used for detection. Preprocessing is a necessary operation before feature extraction. For instance, most approaches operate on feature vectors of the same dimension, and as such demand that crop or resize the image to a fixed size. For feature extraction, researchers usually compute statistical features of images and then process them to construct a set of detection features. Finally, feature vectors are used to train classifiers, such as SVM(support vector machine), to obtain the final model.

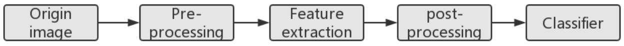


FIGURE 2. Common processing pipeline for CG detection.

Lyu and Farid [6] proposed a detection method based on wavelet decomposition to extract the fourth-order statistical features from each sub-band of the original image: mean value, variance, deviation, and kurtosis. Besides the traditional statistical features of mean and variance used to represent distributions, the deviation is applied to measure the asymmetry of the probability distribution, and the kurtosis is used to measure the degree of difference between the peak and the rest. However, these features are not sufficient for CG detection. To obtain a better detection performance, the author further extracted image features and constructed features such as direction, color band, and scale that could represent the correlation coefficient of images. Finally, a 216-dimensional feature vector is extracted from each image, which can be used to train LDA(Linear Discriminant Analysis) and SVM to complete detection tasks.

Chen *et al.* [7] found that the extracted features were affected by the color space of the image, and proposed an improved wavelet transform detection method. This method extracts 78 features from each color channel of the image and obtains 78*3 dimensional feature vectors. Besides, this method also compares the performances of different color spaces on feature extraction, especially the differences between RGB and HSV color spaces. Experimental results show that the feature extracted from HSV color space contains certain brightness information, and the model trained on it performs better.

Considering the difference between the noise that CG and PG are forced into during the generation, Ng *et al.* [8] proposed a detection approach based on the geometric features of images. In this approach, two different evaluation concepts are put forward for the process of image generation and the content of image respectively. The one is authenticity of the production process, the other is the authenticity of the content. The former one only evaluates the method of generating an image, and the latter only evaluates the content of the image. They are independent of each other. However these two concepts do not suitable for distinguishing between traditional CG and PG, instead, they can be used to separate images that cannot be properly detected. Based on these two concepts, two different methods of image analysis are proposed: (1) describing the geometry by the language of differential geometry; (2) describing the geometry by fractal dimension and partial patch. Once the features are extracted, the SVM is trained to obtain the final detection model.

Dirik *et al.* [9] put forward that CG and PG could be distinguished by tracing the interpolation process of images. The author found that after interpolating the PG with the Bayer filter, the change caused by the interpolation of the same filter is smaller than that caused by the other filter. In addition, this paper also discusses the interference of the lens in the PG acquisition process. Illumination is refracted into the camera through the lens, and different lenses mean different colors, wavelengths, and so on. That is to say, the image taken by different lenses will have a certain color difference with the color channel, while the image with high mutual information does not have this feature.

Peng *et al.* [10] presented that the differences in image generation modes will lead to some differences in the statistical characteristics, visual characteristics and noise characteristics of CG and PG. Firstly, this method extracts the statistical characteristics of the grayscale image in the spatial domain and wavelet domain, such as mean value and variance. Then, the fractal dimension and wavelet band of a gray image are extracted as visual features. Finally, a gaussian high-pass filtering is applied to the image, and enhanced photo response non-uniformity noise (PRNU) is used to extract the physical features of the image.

With the development of traditional detection methods, researchers find that there are many limitations in traditional detection methods:

1) The generalization capability of the traditional CG detection approach is not ideal. Since traditional detection methods need to manually extract and construct detection feature vectors, researchers need to have a certain understanding of the datasets that were used. The extracted features are filtered, combined, and computed to construct the feature vector for detection. These feature vectors can represent the dataset to some extent. Changing the dataset means that the feature vectors no longer match the image features, and the detection accuracy of the model will decrease significantly.
2) It is difficult to make a trade-off between accuracy and efficiency. The detection accuracy of the model is directly affected by the feature vector. Generally, the more semantic information contained in the feature vector, the higher the detection accuracy. It takes a lot of time for researchers to design appropriate feature vectors, which are only applicable to specific datasets. Besides, with the improvement of images quality, the detection accuracy of the model based on manually extracted feature training cannot meet the requirements of scientific research.

The purpose of training a new model is to detect the new incoming image. The limitations of traditional detection methods prompted researchers to find new detection methods. Due to the great success of deep learning technology in the field of computer vision [2]–[5], more and more researchers have begun to try to detect CG through deep learning.

III. DEEP LEARNING-BASED CG DETECTION APPROACHES

In order to get a better visual effect, the texture of CG is smoother than PG, which leads to differences in the statistical features between them. Some traditional detection methods extract features of images for CG detection based on this phenomenon. But with the rapid development of graphical processing tools, the characteristics of CG have been very similar to PG. The traditional detection technology has been unable to accurately distinguish between PG and CG while deep learning technology can obtain more image features through convolution, pooling, and other operations. By using these features to train a neural network, a model with higher detection accuracy can be obtained.

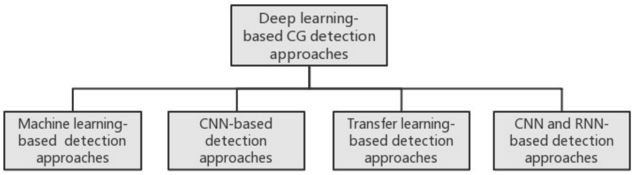

Deep learning-based CG detection approaches
Machine learning-based detection approaches
CNN-based detection approaches
Transfer learning-based detection approaches
CNN and RNN-based detection approaches

FIGURE 3. Classification of existing deep learning-based CG detection approaches.

In recent years, more and more scholars have conducted scientific research in this field and proposed a series of detection approaches. This paper reviews the deep learning-based CG detection approaches. The existing methods are divided into four categories according to their network structure. As shown in Figure 3, the four categories consist of Machine learning-based [11], [12], CNN-based (Convolutional Neural Network) detection approaches [13]–[15], Transfer learning-based detection approaches [16]–[20] and CNN and RNN-based (Recurrent Neural Network) detection approaches [21]. It is a study hotspot in this field to improve or replace the traditional detection method through using deep learning technology to obtain higher detection accuracy.

A. MACHINE LEARNING-BASED DETECTION APPROACHES

Inspired by the powerful feature extraction ability of the CNN, Rahmouni *et al.* [11] proposed a deep learning-based image detection approach. Figure 4 shows the basic steps of Rahmouni's approach: filtering, feature extraction, and classification. This method uses CNN as the feature extractor of the model and combines it with MLP[27] (multi-layer perceptron) to complete the CG detection task. In order to obtain feature vectors with the same dimensions, the original image is cropped into image patches with the same resolution in the pre-processing stage. Then, the author extracts image features through convolution operation and transfers them into the custom pooling layer for aggregation. The acquired features are used to train the MLP to obtain the final detection model. The category of the original image can be predicted by calculating and combining the results of image patches. At the same time, this approach can carry out local CG detection to judge whether the PG contains CG.

Peng *et al.* [12] holds the ideal that although CG and PG cannot be distinguished by naked eyes, CG cannot imitate natural scenes completely. The texture of CG would be smoother than that of PG. In their scheme, a gaussian low-pass filter is used to remove the high-frequency component of the image, and the multiple linear regression models are used to extract residual images. Then, the author investigates the fitting degree of the regression model. By analyzing

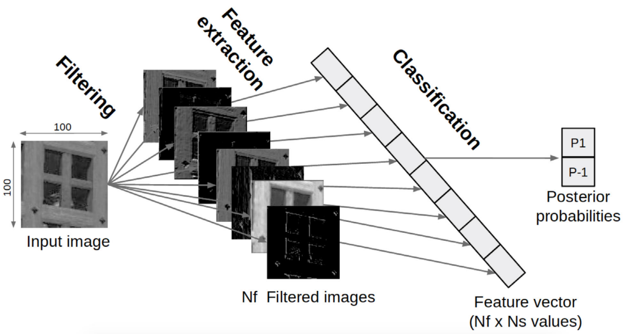


FIGURE 4. The basic steps of Rahmouni's method [11].

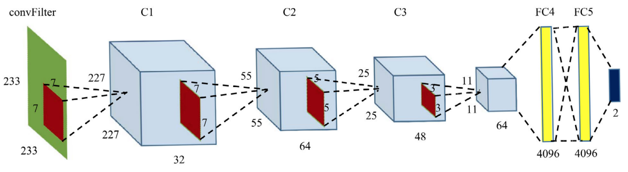


FIGURE 5. Model structure diagram proposed by Quan *et al.* [13].

the difference between the residual images, the features of histogram and multi-fractal spectrum are extracted, and the feature vector of the original image is constructed by combining them with the regression model fitness features. Finally, the extracted features are used to train the LIBSVM to obtain the detection model.

B. CNN-BASED DETECTION APPROACHES

Quan *et al.* [13] presented that the detection accuracy of the model is directly affected by the pattern of image sampling. The network structure is shown in Figure 5. In order to make the neural network learn image information as much as possible, the author proposes to use the MPS(Maximal Poisson-disk Sampling) [22] algorithm for image sampling, and then use the CNN to train the detection model. Even when the number of sampling points is limited, the sampling points given by the MPS algorithm can completely cover the image and retain the original image information as much as possible, which enables the neural network to obtain the most image features. According to the sampling points given by the MPS algorithm, the original image is cropped into image patches with the same size and then transferred to CNN for training. When the detection of the image patches is completed, the category of the original image will be judged by the Local-to-Global strategy. Different from the ordinary CNN, the author added a convFilter layer before the convolutional layer to adjust the size of the input image, so that the network can accommodate images with various sizes. At the same time, visual analysis of convolution kernel and image illumination characteristics is also carried out in the paper. The results show that illumination information is the key to distinguish between CG and PG. This model has strong robustness, and the detection performance is still good after scaling the image or JPEG compression.

Yao *et al.* [14] proposed to use a high-pass filter to remove the low-frequency component of the image and combine it with a CNN for detection. The network structure is shown in Figure 6. The low-frequency component represents image content. Filtering out these signals allows CNN to pay more attention to the residual components and the sensor pattern noise (SPN) introduced by digital cameras. In order to reduce the computational cost of the model, this method firstly clips the RGB image into the same size image patches and converts them to grayscale images. Then, the pre-defined high-pass filter is applied to the image patches and transferred them to CNN for training. Experimental results show that the trained model achieves 100% accuracy even when the image is recompressed.

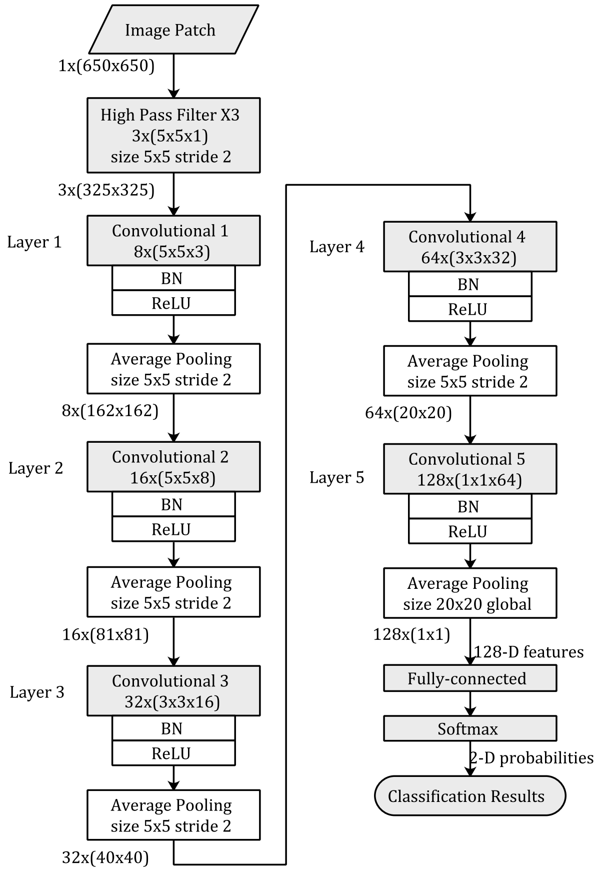


FIGURE 6. Model framework proposed by Yao *et al.* [14].

Compared with the thousand classification task such as ILSVRC (ImageNet Large Scale Visual Recognition Challenge) [23], CG detection is a simple two-class classification task. In order to adapt to the task better, Yu *et al.* [15] proposed a method to remove the pooling layer that can aggregate the features extracted by the convolution layer and reduce the dimension of CNN. Removing the pooling layer can enable the neural network to learn more image features. By simplifying the VGG-net [24], this method designs a new network structure, which includes 6 convolutional layers and 3 fully connected layers, and obtains good performance. Besides, the model can carry out local CG detection to judge whether the PG contains CG.

C. TRANSFER LEARNING-BASED DETECTION APPROACHES

It requires a large number of samples to train the model for a deep learning task. However, sometimes the samples are not enough to train an end-to-end model, in this case, transfer learning [25] is a good solution. The essence of transfer learning is to fine-tune the parameters of the pre-trained model to enable it to perform the specified tasks. If the existing dataset is significantly smaller than that of the pre-trained model and there is some similarity between the two, transfer learning could be an effective approach. Besides, fine-tuning the pre-trained model for the new task can avoid overfitting effectively.

The experiments designed by Gando *et al.* [16] compared the detection capability of the traditional detection method, newly trained model and transfer learning-based model. Firstly, for the traditional detection method, the edges and color intensities of the input image are extracted to construct the feature vector, which is used to train the traditional classifier. In the second step, the author simplified the AlexNet [26] and trained it from scratch. In the end, the author fine-tunes the pre-trained AlexNet model on the same dataset. By comparing the detection accuracy of the three methods, it can be found that the transfer learning-based model is better than the other two methods.

Based on the basic idea of transfer learning, Cui *et al.* [17] proposed a real-time CG detection approach by fine-tuning the pre-trained ResNet-50 [27] model. Figure 7 shows the approach flowchart proposed by Cui *et al.* In the pre-processing stage, the RGB image is converted to grayscale image and its low-frequency components are removed by a high-pass filter. The image is then transferred to ResNet-50 to fine-tune the parameters of the model. The fine-tuned model obtained approximately 98% accuracy on the used dataset. It is worth mentioning that the training time of this method for a single image is 1.02 seconds, which achieving the real-time detection performance.

De Rezende *et al.* [18] used ResNet-50 [27] to extract image features, and the extracted features are used to trained SVM to obtain the detection model. The image preprocessing method of this model is simple. The original image is resized to a fixed size of 224*224 and the average value of the image pixels in the ImageNet dataset is subtracted pixel by pixel. Then, this approach freezes the parameters of the first 49 layers of ResNet-50 to extract image features and replaces the softmax function of the pre-trained model with an SVM with a Radial Basis Function (RBF) kernel. In addition to the detection approach, the author constructed the DSTokExt dataset by extending the DSTok dataset [1] (described in section 4). The DSTokExt dataset contains 8394 CG and 8002 PG. All of which are collected from the network.

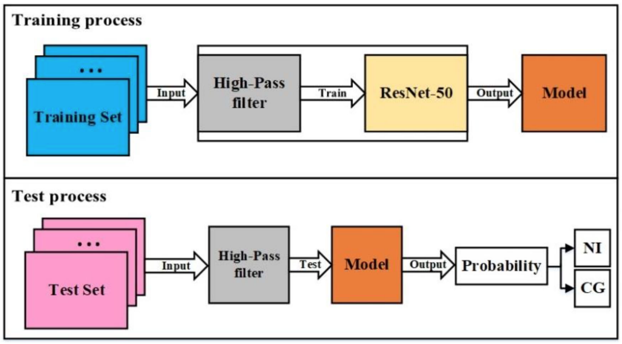


FIGURE 7. The CNN structure proposed by Cui *et al.* [17].

Nguyen *et al.* [19] found that the semantic information of the image would be lost after the layer-by-layer transmission of the neural network. The loss of semantic information makes features tend to be homogenized, which will reduce the detection accuracy of the model. This method uses the VGG-19 [28] model as the feature extractor and the statistical CNN as the feature transformers and classifier. Figure 8 shows the feature extractor, feature transformers, and classifier proposed by Nguyen. They extract the input of the first three pooling layers of VGG-19 as image features, transmitted them to the pre-built feature conversion module to convert image features into statistical features. Only then they trained the MLP to get the detection model.

The detection accuracy of transfer learning depends on the pre-trained model chosen by the researcher. He [20] designed experiments to compare the detection performance of the

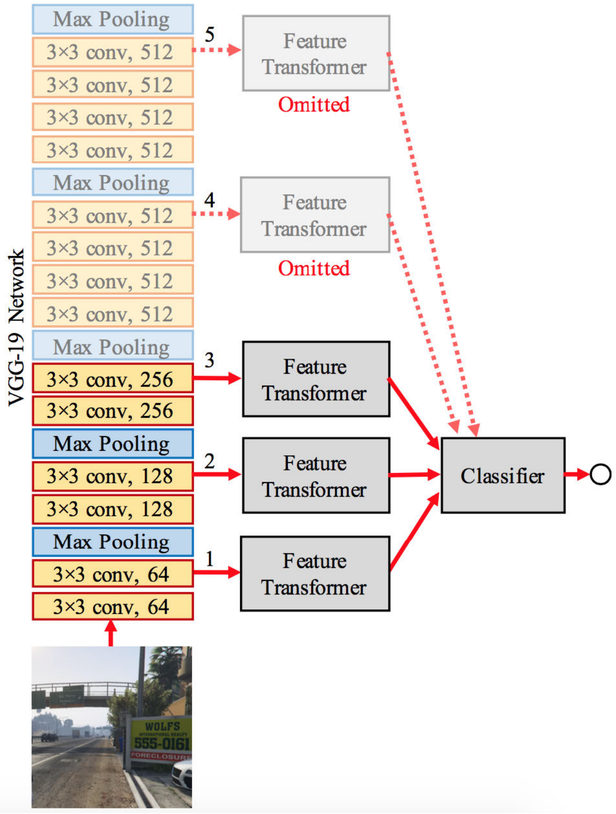


FIGURE 8. The feature extractor, feature transformers, and classifier proposed by Nguyen *et al.* [19].

commonly used VGG-19 [28] model and ResNet-50 [27] model. In their experiments, the two pre-trained models were fine-tuned using the same dataset. The results showed that the detection performance of ResNet-50 was better than that of VGG-19.

D. CNN AND RNN-BASED DETECTION APPROACHES

The Recurrent Neural Network (RNN) is a kind of neural network which is used to process sequential data. However, researchers find that the content, background and other factors of the image have a certain correlation. For example, sky and cloud usually appear together in the image. It is a new research approach to use the RNN to learn the correlation of image features. RNN needs to be combined with CNN to obtain better detection performance, since RNN work not well when it performs image detection alone.

The image contains a large amount of information. The detection efficiency of the model will be very low, if the image is transmitted to the RNN directly. He *et al.* [21] proposed an approach combining a dual-path CNN with an RNN, and its network structure is shown in Figures 9 and 10. This method first converts the original RGB image into the YCbCr image and extracts the color and texture information of the input image by using the Schmid filter, then transmits them to dual-path CNN for feature extraction. Finally, the feature is transferred to the DAG-RNN (Directed acyclic graph-RNN) [29] for training. Compare to other detection approaches, the network structure of this approach is unique. The dual-path CNN can obtain more image information without adding the layer of CNN, which reduces the probability of model overfitting. Meanwhile, the DAG-RNN is used to learn the correlation information between image contents, which enhances the robustness of the model.

IV. EXPERIMENTS

The experiment in this paper can be divided into six subsections. Details of the experimental environments and the programming language are provided in section 4.1. In section 4.2, the datasets of CG and PG are introduced. Then, we illustrate the evaluation criteria of the model. From section 4.4 to 4.6, we design three comparative experiments to evaluate the detection performance of each approach.

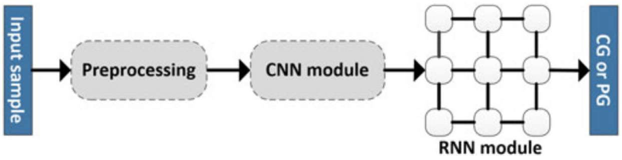


FIGURE 9. Network model framework proposed by He *et al.* [21].

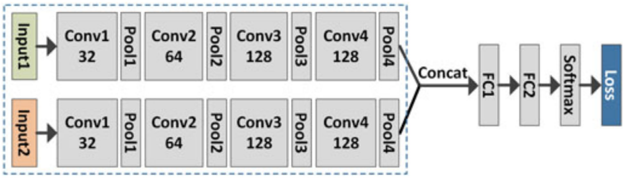


FIGURE 10. Structure diagram of dual-path CNN proposed by He *et al.* [21].

TABLE 1. Available CG and PG dataset.

Dataset	Number of PG	Number of CG	Resolution
Columbia [30]	1600	1600	276*421 to 1398*1404
Raise [31]	8156	N/A	3008*2000 to 4928*3264
Level-Design Reference	N/A	63368	About 1680*1050
DSTok [1]	4850	4850	609*603 to 3507*2737
He's dataset [21]	6800	6800	266*199 to 2048*3200

A. SYSTEM CONFIGURATION

Python[1] and MATLAB[2] include many useful image processing functions, and we will choose between the two according to the actual needs in the pre-processing stage. The neural network framework used in the experiment includes: Tensorflow-GPU 1.4.0,[3] Torch7,[4] Keras 2.2.4[5] and Caffe.[6]

In this paper, the detection performance of 7 deep learning-based detection approaches [11], [13], [14], [16], [18], [19], [21] are compared. The source code of some approaches [11][7][8][9] are already on Github. The source code of [21] is provided by the author. The source code of the other four approaches [14], [16], [19] is written by ourselves according to the original method, and are available on Github.[10]

All the experiments were conduct on a Xeon e5-2620 v4 2.10GHz with 64GB of RAM. The GPU used in the experiments was GeForce GTX1080ti.

B. DATASET

1) AVAILABLE DATASETS

Researchers usually pay more attention to the detection performance of models and approaches themselves, while ignoring the important role of the dataset in deep learning. In fact, the dataset is playing an increasingly important role in deep learning. If the dataset is large enough, then the model trained based on it has a lower probability of over-fitting and has a stronger robustness. As shown in Table 1, this paper collects available datasets for CG detection and introduces the number of PG, the number of CG, and the image resolution in each dataset.

Columbia dataset [30] is the first public dataset used for CG detection. This dataset was produced by Ng *et al.*, from Columbia University. It contains 4 subsets with a total of 3,200 images: (1) 800 photorealistic computer graphics

[1] https://www.python.org/
[2] https://www.mathworks.com/products/matlab.html
[3] https://www.tensorflow.org/
[4] http://torch.ch/
[5] https://keras.io/
[6] https://caffe.berkeleyvision.org/
[7] https://github.com/NicoRahm/CGvsPhoto,quan2018distinguishing
[8] https://github.com/weizequan/NIvsCG,de2018exposing
[9] https://github.com/bazinho/CG
[10] https://github.com/Nx2018/Computer-Generated-image-dataction

from the Internet; (2) 800 photographic images from the personal; (3) 800 photographic images from Google image search; (4) 800 photographed photorealistic computer graphics. This dataset has a small number of images and contains images of heterogeneous sources, which leads to difficulties in detection.

The Raise dataset [31] was produced by Dang-Nguyen *et al.* of the University of Cagliari, Italy. This dataset contains 8156 original high-resolution PG. Dang-Nguyen *et al.* hope that this dataset can provide researchers with a common benchmark for comparison. Now, many scholars select some images from RAISE to construct the PG set.

The Level-Design Reference Database[11] contains 63,368 CG, which based on the open-source software Piwigo. The images in this dataset are all real-time screenshots of 3D games. Since the dataset contains a large number of images, and the content and style of the images in this dataset are significantly different, researchers can choose some high-quality images among them to build CG sets.

DSTok dataset [1] was constructed by Tokuda *et al.*, which contained 4850 PG and CG respectively. All images of this dataset were collected from the Internet, where PG contains indoor and outdoor landscapes taken by various devices, and CG collects realistic photos as much as possible. The dataset is comprehensive in content and has a large number of images, which is an important dataset for CG detection research.

He's dataset [21] contains 6800 PG and 6800 CG. CGs are collected from the network and generated by 3DS Max, Maya and more than 50 kinds of rendering software. PGs are captured by different types of cameras in various scenes, with rich content and resolution ranging from 266*199 to 2048*3200.

In addition to the above datasets, Rahmouni *et al.* [11] selected 3600 images from the Raise dataset and the Level-Design Reference Database to construct their dataset. Besides, some researchers [12], [16], [18] also build CG set by collecting film screenshots, game images and 3D models, and use cameras, mobile phones and other tools to capture images directly to construct PG set.

2) SELECTED DATASET

As mentioned above, there are few datasets available in the field of CG detection. To make sure the experiment is fair, we did not construct new dataset by ourselves but used datasets commonly used in this field. We take the DSTok [1] dataset of Tokuda *et al.* as the main dataset, together with the datasets of He *et al.* [21] and Rahmouni *et al.* [11], to construct a series of experiments based on these three datasets. In the benchmark, we used the DSTok dataset and He's dataset to test the detection performance of each model. In the model generalization capability test section, we used the DSTok dataset to train the model, and then tested them on the dataset of He *et al.* [21] and Rahmouni *et al.* [11] respectively. In the final experiment, we tested the robustness of the models trained based on the DSTok dataset by injecting noise into the test images.

TABLE 2. Benchmark results.

Approach	DSTok dataset	He's dataset	Framework
Rahmouni et al. [11]	75.49%	76.37%	Tensorflow
Quan et al. [13]	93.74%	93.99%	Keras
Yao et al. [14]	98.35%	93.47%	Caffe
Gando et al. [16]	85.50%	89.49%	Tensorflow
De Rezende et al. [18]	95.02%	94.94%	Keras
Nguyen et al. [19]	91.60%	99.27%	Keras
He et al. [21]	91.58%	93.13%	Torch7

C. MODEL EVALUATION CRITERIA

The evaluation criteria of the model vary from field to field. For example, researchers are more concerned with CG masquerading as PG in criminal investigations, while in child abuse cases in the United States, they hope to improve the detection accuracy of PG. Considering the different standards in different fields, this paper uses the average accuracy of CG and PG to evaluate the detection ability of the model. At the same time, we believe that models with detection accuracy below 70% cannot distinguish between CG and PG.

D. BENCHMARK

Since the datasets used by each approach are not uniform, the accuracy given in the original paper can only be used as a reference for experiments. In order to evaluate the detection performance of each model, it is necessary to benchmark the model using a uniform dataset. In this experiment, seven approaches are compared. The image preprocessing method and network structure of each approach are consistent with that of the original paper. Table 2 shows the detection accuracy of each approach on the DSTok dataset and He's dataset, as well as the deep learning framework used to implement the approach. It should be noted that we removed some images in He's dataset that could not meet the preprocessing requirements of some methods due to the small resolution. However, since the number of deleted images is extremely few, there is no impact on the experimental results.

It can be seen that the accuracy of the method proposed by Rahmouni has decreased significantly. This approach is an improvement of the traditional detection method. Although convolution operation can extract more features, the performance of the model decreases significantly when the model used to detect the dataset with comprehensive contents and high quality. Gando's approach doesn't work well too. The approaches of Quan et al. [13], De Rezende et al. [18], and He et al. [21] achieved good detection accuracy and stable performance in both datasets. Yao's method achieves the highest detection accuracy in the DSTok dataset, while Nguyen's method achieves the highest detection accuracy in He's dataset.

[11] http://level-design.org/referencedb/ (accessed on 20 July 2019)

TABLE 3. Experimental results of model generalization capability test.

Approach	Rahmouni's dataset	He's dataset
Rahmouni et al. [11]	60.85%	63.67%
Quan et al. [13]]	56.43%	75.71%
Yao et al. [14]	78.37%	62.75%
Gando et al. [16]	67.48%	75.06%
De Rezende et al. [18]	73.00%	85.25%
Nguyen et al. [19]	36.81%	71.44%
He et al. [21]	56.78%	78.42%

E. GENERALIZATION CAPABILITY TEST

In this experiment, the generalization ability of the model is evaluated by changing the test set to simulate the situation where the model detects unknown data in the real-world application. The generalization capability of the model is evaluated according to the detection performance. We used the DSTok dataset to train the model and then tested it on Rahmouni's dataset and He's dataset. Table 3 shows the results of this experiment.

Limited by the number and content of images in the dataset, the generalization capability of deep learning models has been unsatisfactory. The experimental results are in line with expectations, the detection accuracy of each model decreases to different degrees after replacing the test set. All the models except Yao's achieve higher accuracy on He's dataset than on Rahmouni's dataset. This is because He's dataset is closer to the DSTok dataset in image style and content. However, Yao's approach uses a high-pass filter to remove the content information of images, which makes the model's attention attracted by the residual noise.

F. ROBUSTNESS TEST

Robustness is an important indicator of the model evaluation. Attackers may use various methods to tamper with an image to pass the detection in real-world applications. This experiment simulates this situation by injecting noise into the image. Specifically, we did not make any changes to the training image but tamper with the test image. That is to say, instead of training a new model, we use the model trained on the DSTok dataset in the benchmark. We inject different types of noise into the test images of the DSTok dataset. By comparing the detection accuracy of the model before and after noise injection, we can see the influence of noise on the detection performance of the model and evaluate the model's robustness.

To save on computational costs, we use simple Salt-and-Pepper noise and Gaussian noise to process the test image. PG is still PG after noise injection, the same for CG. Furthermore, considering that the input image size of each approach is different, if all images are injected with the same number of noise points, the approach with a larger input image size will have a significant advantage. Therefore, to ensure the fairness of this experiment, we injected the same Signal-to-Noise Ratio (SNR) noise into the test image.

For Salt-and-Pepper noise, we set three different sizes of SNR: 0.99; 0.95; 0.9. We show CG with Salt-and-Pepper noise from Figure 11 to Figure 14. The visual effect of the noise-bearing image is obviously lower than that of the original image, but the noise did not destroy the image content. That is to say, the semantic information of the noise-bearing image is the same as that of the original image, and the noise only reduces the quality of the original image.

TABLE 4. Comparison of detection accuracy of models at different SNR.

Approach	SNR = 0.99	SNR = 0.95	SNR = 0.9
Rahmouni et al. [11]	52.59%	51.36%	50.73%
Quan et al. [13]	50.27%	50.02%	49.99%
Yao et al. [14]	47.96%	45.44%	50.00%
Gando et al. [16]	79.01%	70.52%	64.53%
De Rezende et al. [18]	92.19%	86.63%	80.55%
Nguyen et al. [19]	76.17%	60.93%	56.46%
He et al. [21]	50.18%	50.01%	50.05%

TABLE 5. Comparison of detection accuracy of models at different SD.

Approach	SD = 10	SD = 30	SD = 50
Rahmouni et al. [11]	52.19%	50.00%	50.25%
Quan et al. [13]	52.95%	49.66%	48.91%
Yao et al. [14]	44.31%	41.23%	50.00%
Gando et al. [16]	75.00%	65.08%	57.50%
De Rezende et al. [18]	96.63%	78.00%	67.44%
Nguyen et al. [19]	57.09%	54.86%	51.21%
He et al. [21]	72.38%	57.47%	54.41%

After injecting noise into all test images, we detected them using the model trained in the benchmark. The experimental results are shown in Table 4. It can be seen that the models of Rahmouni et al. [11], Quan et al. [13], Yao et al. [14], and He et al. [21] no longer have the ability to distinguish between CG and PG when the SNR is 0.99. These four models are all trained from scratch based on CNN. The small number of training set images leads to the poor robustness of the model directly. The other three models with good resistance to noise are all built based on transfer learning, and they use the pre-trained models on ImageNet. Overall, De Rezende's approach [18] performed best in this experiment. However, the accuracy of the model decreases as the noise proportion increases. A model with only 80.55% detection accuracy cannot provide services for forensics and other fields obviously.

In the following experiment, Gaussian noise with fixed SNR is injected into the image. Considering that Gaussian noise has little disturbance to the image, we increase the proportion of the noise. Meanwhile, we set the mean value of Gaussian noise to zero and took the Standard Deviation (SD) as the variable. Specifically, the SNR of the noise is 0.7, the mean is 0, and the SD is 10, 20, and 30 respectively.

Table 5 shows the results of this experiment. Unexpectedly, when the SD of the noise is 10, the detection accuracy of De Rezende's model [18] is improved to 96.63%. By analyzing the network structure of the approach, we believe that this phenomenon is caused by the preprocessing mode of this method. This approach subtracts the mean RGB value of the ImageNet dataset for each pixel in the preprocessing stage. Similarly, He's model [21] also has an unreasonable situation when the SD is 10. We think this is caused by two reasons: 1. The dual-path CNN reduces the influence of noise when extracting the color and texture information of image; 2.

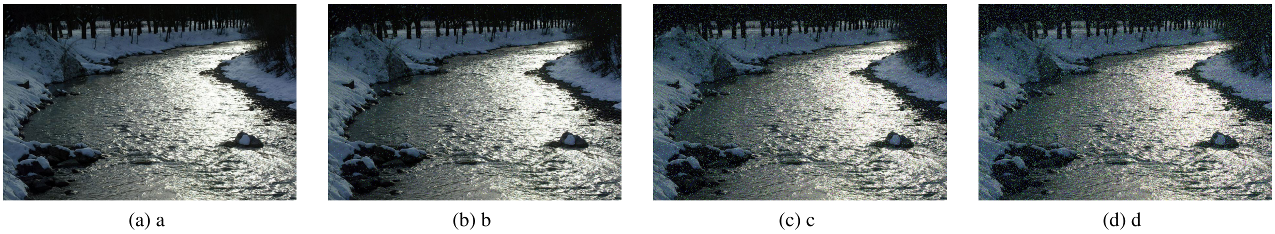

(a) a (b) b (c) c (d) d

FIGURE 11. **PG with Salt-and-Pepper noise. From left to right, SNR = 1 (original image), SNR = 0.99, SNR = 0.95, SNR = 0.9, respectively.**

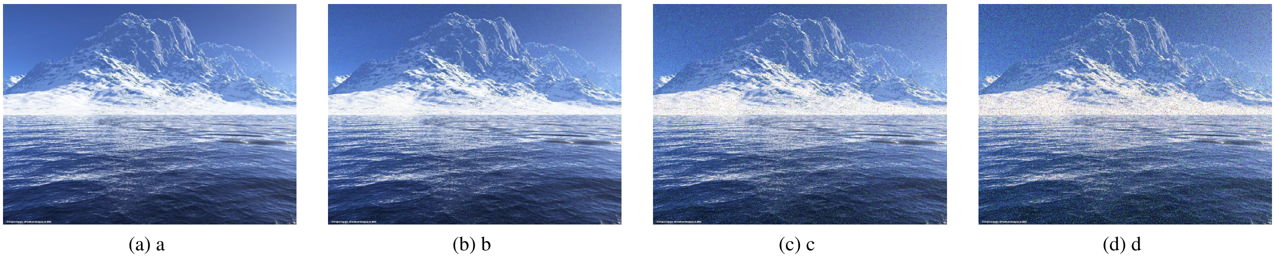

(a) a (b) b (c) c (d) d

FIGURE 12. **CG with Salt-and-Pepper noise. From left to right, SNR = 1 (original image), SNR = 0.99, SNR = 0.95, SNR = 0.9, respectively.**

TABLE 6. **Comparison of CG detection approaches based on deep learning.**

Category	Index	Extractor	Advantage	Limitation
Machine learning-based detection approaches	Rahmouni [11]	CNN	Local CG detection	Low accuracy
	Peng [12]	Multiple Linear Regressions	Fast training	Manually extract features for specific datasets only
CNN-based detection approaches	Quan [13]	CNN	Accommodates different sizes of input	Over-cutting of low-size images
	Yao [14]	Hpf+CNN	Focus more on image generation	Ignore the impact of image content
	Yu [15]	CNN	Local CG detection	Poor robustness
Transfer learning-based detection approaches	Gando [16]	AlexNet	High detection accuracy and fast training	There are not many pre-trained models
	Cui [17]	Hpf+ResNet-50		
	De Rezende [18]	ResNet-50		
	Nguyen [19]	VGG-19		
	He [20]	ResNet-50		
CNN and RNN-based detection approaches	He [21]	CNN	More attention to image content	High hardware requirements

The size of the input image is large (96 * 96 * 15), which makes the noise fluctuations less under the same SD.

V. DISCUSSION

A. COMPARISON

As mentioned above, this paper introduces 11 types of deep learning-based CG detection approaches, and divides them into 4 categories based on the network structure (as shown in Figure 2), then introduces the basic steps and principles of each approach. Table 6 shows the comparative analysis of the 11 deep learning-based CG detection approaches. This table compares the feature extractors of each method and gives the advantages and disadvantages of each method.

As can be seen from Table 6, most approaches claim to be able to distinguish CG and PG, but the detection accuracy and efficiency of each model are different, among them:

1) In the CNN-based detection approach, the model of Quan *et al.* [13] and Yao *et al.* [14] work well. However, training a model from scratch requires a lot of computing resources. At the same time, due to the limitation of the existing dataset, the generalization capability and robustness of the model are weak.
2) The five models trained based on transfer learning work well and do not require excessive computing resources. Although transfer learning is an effective training method, there are not many pre-trained models available to researchers at present. Furthermore, because the pre-trained model's dataset (usually the ImageNet dataset) is much larger than the fine-tuned training set, the scalability of the model is poor.
3) In the approach combining CNN and RNN, He et al. [21] introduced RNN into the field of CG detection for the first time and obtained a good detection performance. In the future, researchers may find new methods to combine CNN with RNN for CG detection.

B. CONCLUSION

The experiment in this paper consists of Benchmark, Generalization capability test and Robustness test. In the benchmark, we used the same dataset to evaluate the detection capability of each model. Next, we changed the test images to obtain the generalization capability of each model. Finally, we add noise to images to test the robustness of each model.

Through these three experiments, we have a general understanding of the detection accuracy, generalization capability, and robustness of each model. Most models have good detection capabilities and can distinguish between CG and PG when it is trained and tested on the same dataset. However, after changing the test dataset, the detection accuracy of the model decreases significantly. Generalization capability is an important index to evaluate a model. Whether the model can perform better on a new dataset determines the application prospect of the model directly. At present, due to the content and quantity of training images, it is difficult to improve the generalization capability of the model. Meanwhile, the model cannot resist the noise attack. As long as the image is injected with noise, the model cannot distinguish whether it is PG or CG accurately. If the attacker designs a new tampering method according to the weaknesses of each model, which can guarantee the visual effect of the image and avoid the detection of the model, the model will not work.

C. CHALLENGES

This paper introduces 11 existing deep learning-based CG detection methods, divides them into 4 categories, and introduces the steps, model structure and application limitations of each method. A comparative analysis of these methods is presented in Table 6.

With the development of graphical processing tools and the explosive growth of image number, the Machine learning-based detection approach has been unable to meet the research requirements. The CNN-based detection approach and Transfer learning-based detection approach extracts image features by convolution, pooling, and other operations, and achieves better detection accuracy. In the next few years, new methods may be found to combine CNN with RNN for further research.

At present, many scientific research achievements have been published in the field of CG detection, but the deep learning-based CG detection is still in its starting stage. Challenges in the research of deep learning-based CG detection mainly include:

1) The field of CG detection lacks large datasets for deep learning training. Currently, the number of images in the publicly available CG detection datasets cannot meet the requirements of deep learning training. In the CG detection task, the CG used for training must be similar to the PG, which brings difficulties to dataset construction. Due to the lack of training samples, researchers had to cut images multiple times, which increased the likelihood of model overfitting. Establishing a CG detection database and publicly sharing it can facilitate the development of deep learning-based CG detection methods.
2) Most of deep learning-based CG detection models lack of generalization capability. It can be seen from our experiment that the detection accuracy of each model decreases obviously after changing the test set. The generation tools of each dataset are different, and the content and style of the image are also different. The limitations of the dataset greatly reduce the generalization capability of the trained model. How to improve the generalization capability of deep learning-based CG detection approaches is the primary problem that researchers need to solve.
3) Most of the research achievements published so far are focused on images that have not been tampered, and the researchers have not yet adjusted the network structure of the model to deal with various attacks. It can be seen from the experiments in this paper that the detection accuracy of the model will be reduced by simply adding noise to the dataset. An attacker can exploit this vulnerability to evade detection. In the next few years, how to improve the robustness of the model will become the emphasis and difficulty in this field.

D. FUTURE DIRECTIONS

The rapid development of the network makes CG easily spread across the Internet. More and more attention has been paid to the authenticity detection of image content. The use of CG seriously damaged the authenticity of the image. If the CG cannot be detected accurately, it may may bring out the disorder. The deep learning-based CG detection approach will be a research hotspot in the field of CG detection. Future directions in this field include:

1) CNN and RNN are further combined to detect the correlation between image contents. At present, the combination of the two frameworks lacks sufficient approaches and theoretical models. Using CNN for feature extraction and RNN for feature correlation analysis can improve the robustness and generalization capability of detection models.
2) Introduce the Game Theory into the GC detection field. At present, Currently, researchers do not consider the concept of game theory in the model construction phase. Attackers can easily evade detection by modifying images. In future research, researchers need to conduct offensive and defensive games with attackers to build models that can resist various attacks.

In the next few years, with the application and popularization of images and the requirements of image security, CG detection will attract more attention. In the filed of deep learning-based CG detection, the theoretical approaches, image processing methods, network structure, and standard datasets will be improved gradually, and the model with good detection performance is expected to be applied in practice.

REFERENCES

[1] E. Tokuda, H. Pedrini, and A. Rocha, "Computer generated images vs. digital photographs: A synergetic feature and classifier combination approach," *J. Vis. Commun. Image Represent.*, vol. 24, no. 8, pp. 1276–1292, 2013.

[2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Cognit. Modeling*, vol. 5, no. 3, p. 1, 1988.

[4] M. Chen, V. Sedighi, M. Boroumand, and J. Fridrich, "JPEG-phase-aware convolutional neural network for steganalysis of JPEG images," in *Proc. 5th ACM Workshop Inf. Hiding Multimedia Secur.*, 2017, pp. 75–84.

[5] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *Proc. 4th ACM Workshop Inf. Hiding Multimedia Secur.*, 2016, pp. 5–10.

[6] S. Lyu and H. Farid, "How realistic is photorealistic?" *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 845–850, Feb. 2005.

[7] W. Chen, Y. Q. Shi, and G. Xuan, "Identifying computer graphics using HSV color model and statistical moments of characteristic functions," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 1123–1126.

[8] T.-T. Ng, S.-F. Chang, J. Hsu, L. Xie, and M.-P. Tsui, "Physics-motivated features for distinguishing photographic images and computer graphics," in *Proc. ACM 13th Annu. Int. Conf. Multimedia*, 2005, pp. 239–248.

[9] A. E. Dirik, H. T. Sencar, and N. Memon, "Source camera identification based on sensor dust characteristics," in *Proc. IEEE Workshop Signal Process. Appl. Public Secur. Forensics*, Apr. 2007, pp. 1–6.

[10] F. Peng, J. Liu, and M. Long, "Identification of natural images and computer generated graphics based on hybrid features," in *Emerging Digital Forensics Applications for Crime Detection, Prevention, and Security*. Hershey, PA, USA: IGI Global, 2013, pp. 18–34.

[11] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, "Distinguishing computer graphics from natural images using convolution neural networks," in *Proc. IEEE Workshop Inf. Forensics Security (WIFS)*, Dec. 2017, pp. 1–6.

[12] F. Peng, D.-I. Zhou, M. Long, and X.-M. Sun, "Discrimination of natural images and computer generated graphics based on multi-fractal and regression analysis," *Int. J. Electron. Commun.*, vol. 71, no. 1, pp. 72–81, 2017.

[13] W. Quan, K. Wang, D.-M. Yan, and X. Zhang, "Distinguishing between natural and computer-generated images using convolutional neural networks," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2772–2787, Nov. 2018.

[14] Y. Yao, W. Hu, W. Zhang, T. Wu, and Y.-Q. Shi, "Distinguishing computer-generated graphics from natural images based on sensor pattern noise and deep learning," *Sensors*, vol. 18, no. 4, p. 1296, 2018.

[15] I.-J. Yu, D.-G. Kim, J.-S. Park, J.-U. Hou, S. Choi, and H.-K. Lee, "Identifying photorealistic computer graphics using convolutional neural networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 4093–4097.

[16] G. Gando, T. Yamada, H. Sato, S. Oyama, and M. Kurihara, "Fine-tuning deep convolutional neural networks for distinguishing illustrations from photographs," *Expert Sys. Appl.*, vol. 66, pp. 295–301, Dec. 2016.

[17] Q. Cui, S. Mcintosh, and H. Sun, "Identifying materials of photographic images and photorealistic computer generated graphics based on deep cnns," *Comput. Mater. Continua*, vol. 55, no. 2, pp. 229–241, 2018.

[18] E. R. S. de Rezende, G. C. S. Ruppert, A. Theóphilo, E. K. Tokuda, and T. Carvalho, "Exposing computer generated images by using deep convolutional neural networks," *Signal Process., Image Commun.*, vol. 66, pp. 113–126, Aug. 2018.

[19] H. H. Nguyen, T. N.-D. Tieu, H.-Q. Nguyen-Son, V. Nozick, J. Yamagishi, and I. Echizen, "Modular convolutional neural network for discriminating between computer-generated images and photographic images," in *Proc. 13th Int. Conf. Availability, Rel. Secur.*, 2018, Art. no. 1.

[20] M. He, "Distinguish computer generated and digital images: A CNN solution," *Concurrency Comput., Pract. Exper.*, vol. 31, no. 12, p. e4788, 2019.

[21] P. He, X. Jiang, T. Sun, and H. Li, "Computer graphics identification combining convolutional and recurrent neural networks," *IEEE Signal Process. Lett.*, vol. 25, no. 9, pp. 1369–1373, Sep. 2018.

[22] W. Quan, D.-M. Yan, J. Guo, W. Meng, and X. Zhang, "Maximal Poisson-disk sampling via sampling radius optimization," in *Proc. SIGGRAPH ASIA Posters*, 2016, Art. no. 22.

[23] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[24] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," 2014, *arXiv:1405.3531*. [Online]. Available: https://arxiv.org/abs/1405.3531

[25] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

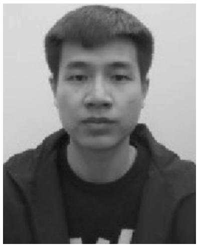
[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[29] B. Shuai, Z. Zuo, B. Wang, and G. Wang, "Dag-recurrent neural networks for scene labeling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3620–3629.

[30] T.-T. Ng, S.-F. Chang, J. Hsu, and M. Pepeljugoski, "Columbia photographic images and photorealistic computer graphics dataset," ADVENT, Columbia Univ., New York, NY, USA, Tech. Rep. 205-2004-5, 2004. [Online]. Available: http://www.ee.columbia.edu/ln/dvmm/downloads/PIM_PRCG_dataset/

[31] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "RAISE: A raw images dataset for digital image forensics," in *Proc. ACM 6th Multimedia Syst. Conf.*, 2015, pp. 219–224.



XUAN NI received the B.S. degree from the Qingdao University of Technology, Shandong, China, in 2018. He is currently pursuing the master's degree with the School of CyberSpace, Hangzhou Dianzi University, Hangzhou, Zhejiang, China. His current research interests include information security, image forensics, deep learning, and computer vision.

LINQIANG CHEN received the B.S. degree in computational mathematics and software applications from Hangzhou University, Hangzhou, China, in 1983, and the M.S. degree in computer graphics from Zhejiang University, Hangzhou, in 1989. He is currently a Professor with the School of CyberSpace, Hangzhou Dianzi University, Hangzhou. His research interests include computer graphics and multimedia information processing.

LIFENG YUAN received the B.S. degree in computer science and technology from Ningbo University, in 2006, and the M.S. and Ph.D. degrees from the Dalian University of Technology, in 2009 and 2017, respectively. He is currently a Lecturer with Hangzhou Dianzi University, and also a Researcher with the Anhui Provincial Key Laboratory of Network and Information Security. His current research interests include secret sharing and information hiding.

GUOHUA WU received the B.S. degree from the Shandong University of Technology, Jinan, China, in 1992, the M.S. degree from the National Institute of Metrology, Beijing, China, in 1995, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 1998. From 1998 to 2000, he was a Lecturer with the Department of Biomedical Engineering, Zhejiang University, and became an Associate Professor, in 2001. Since 2002, he has been with the Department of Computer Science and Engineering, Hangzhou Dianzi University, where he was appointed as a Professor in 2009. His currently research interests include information systems, model driven architecture, and data mining.

YE YAO received the M.S. degree in computer science and the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2005 and 2008, respectively. He was a Visiting Scholar with the New Jersey Institute of Technology, Newark, NJ, USA, from December 2016 to December 2017. He is currently a Lecturer with Hangzhou Dianzi University, and also a Researcher with the Shanghai Key Laboratory of Integrated Administration Technologies for Information Security. His research interests include multimedia forensics and information security.