International Conference on Advances in Materials, Computing and Communication Technologies
In Association with International Journal of Scientific Research in Science and Technology
Volume 9 | Issue 1 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)

Classification of Computer Graphic and Photographic Images Using Convolution Neural Network

Ms. S. Sujitha[1], Dr. I. Muthulakshmi[2]
[1]PG Student, Department of CSE, VV College of Engineering, Tirunelveli, Tamil Nadu, India
[2]Assistant Professor, Department of CSE, VV College of Engineering, Tirunelveli, Tamil Nadu, India

ABSTRACT

With the tremendous development of computer graphic rendering technology, photorealistic computer graphic images are difficult to differentiate from photo graphic images. In this project, a method is proposed based on Maximum Likelihood Principle Component Analysis (MLPCA) image features to distinguish computer graphic from photo graphic images using the CNN classifier. Initially the color image is transform dimension into 128X128 and then converted into gray scale image .The grayscale image can given into a convolution layer has filter or mask operation can performed .The filtered image can be given into ReLU layer. ReLU layer changes the all negative actions to Zero. Maximum Likelihood Principle Component Analysis(MLPCA) can perform feature extraction and reduce the dimensionality of the image .Fully connected layer which are used to generate new features from the existing features.Softmax layer is a classification layer it can be used to classify the computergraphic images from photographic images. Experimental results using Columbia database show that the method achieves reasonable detection accuracy.

Keywords: computergraphics (CG), photographic images (PG),convolutional neural network(CNN),image forensics.

I. INTRODUCTION

Digital image forensics is a research field mostly dedicated to the detection of image falsification. A large part of the current research focuses on the detection of image splicing, copy-past or camera identification [1]. With the development of 3D rendering software and hardware, it becomes more difficult to distinguish Photo graphics from Computer graphics with human eyes. Several industries,such as computer game, can benefit greatly from the advanced rendering techniques to generate more photo-realistic Computer graphics(CGs). This paper addresses the problem of identifying whether an image is a computer graphics or a natural photograph.

A. Computer Graphics vs Natural Photographs

Some recent advances in image processing, like the realtime facial reenactment face2face [2], show the importance of having some tools to distinguish computer graphics (CG) from natural photographic images (PG). Although the distinction between CG and PG depends not only on image properties, but also on cognitive characteristics of viewers [3], people show inaccurate skills at differentiating between altered and non-altered images [4].

B. State of the art

To detect steganography and suggested to apply it to distinguish CG from PG[5]. Like the quasi-totality of the methods that followed, the authors perform a "wavelet-like" decomposition of the images to generate some features vectors, combined with machine learning for classification. Distinguishing CG from PG is by nature strongly related to computer graphics performances in generating photo-realistic images. The first paper to mention the expected difference between CG and PG images, mainly consisting in the image smoothness due to triangles[6].Consider that this difference better resides in the statistical noise properties of the image[7].Consider that the image edges are more relevant for this problem[8].

C. Motivations

Statistical properties of filtered images are good discriminators for distinguishing CG from PG, whether computations involve image gradient [12] or more sophisticated wavelet transformations [9]–[11]. For all these methods, the question of using the best filters is central, i.e. finding the ones that will extract the most meaningful information. Most of the previous works used hand-crafted filtering step which is unlikely to be optimal.

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imager. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics. They have applications in image and video recognition, recommender systems, image classification, medical image analysis, natural language processing and financial time series.

II. DIGITAL IMAGE PROCESSING

A. Image Acquisition

Image acquisition is the process of acquiring the digital image using physical devices and digitizer. The most commonly used image acquisition devices are scanner and video cameras.

B. Image Enhancement

Image enhancement is used to improve the interpretability of information in the image for human viewers and to provide better input for other automated image processing techniques.

1. Spatial Domain Method

The spatial domain is referred as the image plane itself and the approaches are based on direct manipulations of pixels in an image. Spatial domain methods are procedure that operates directly on the pixels. Spatial domain processes is denoted by the expressions:

$$g(x, y) = T[f(x, y)]$$

Where $f(x, y)$ is the input image, $g(x, y)$ is the processed image and T is the operator defined over some neighborhood of(x, y).

2. Frequency Domain Method

Frequency domain processing technique is based on modifying the Fourier transform of the image. Image defect caused by the digitization process or by faults in the imaging set-up can be corrected using image enhancement techniques. If the image is in the good condition, measurement extraction operation is used to obtain useful information from the image.

Image enhancement and measurement extraction on 256 grey-scale images means that each pixel in the image is stored as number between 0 to 255, where 0 represents a black pixel, 255 represents a white pixel and the values in between represent shades of grey. These operations can be extended to operate on color image.

C. Image Restoration

Image restoration is a process that attempts to reconstruct or recover an image that has been degraded by using a prior knowledge of the degradation phenomenon. Similar to the enhancement technique, the function of the restoration technique is to improve the image. Restoration techniques are based on modeling the degradation using a prior knowledge and applying the inverse process in order to restore the original image.

D. Compression

Compression is a technique for reducing the memory required for saving an image or the bandwidth required for transmitting it. Image compression is familiar to most users of computers in the form of image file extensions such as the jpg files.

E. Morphological Processing

Morphological processing is a tool for extracting image component that is useful in the representation and description of shape. Morphological operators often take a binary image and a structuring element as input and combine them using a set operator. They process objects in the input image based on characteristics of its shape which are encoded in the structuring element.

F. Segmentation

Segmentation procedures partition an image into its constituent parts or objects. In general, autonomous segmentations are one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way towards successful solution of imaging problems that require objects to be identified individually.

G. Recognition

Recognition is the process that assigns a label to an object based on its descriptors. Image recognition is used to perform a large number of machine- based visual tasks such as labeling the content of images with meta- tags, performing image content search, guiding autonomous robots, self- driving cars and accident avoidance systems. While human and animal brains recognize objects with ease computers have difficulty with the task.

H. Representation

Representation means quantity that each pixel represents. An important consideration of the image representation is fidelity for measuring quality of an image.

III. IMAGE FEATURES

A. Edges

Edges are points in a boundary between two image regions. In general, edges can be of almost arbitrary shape and may include junctions. The edges are usually defined as sets of points in the image which has a strong gradient magnitude. Edge detection includes a variety of mathematical methods that aim at identifying points in a digital image at which the image brightness changes sharply or more formally and has discontinuities. The points at which image brightness changes sharply are typically organized into a set of curved line segments termed as edges.

B. Corners

Corner detection is an approach used within computer vision systems to extract certain kinds of features that infer the contents of an image. Corner detection is frequently used in motion detection, image registration, video tracking, 3D modelling and object recognition.

C. Blobs

Blob detection methods are aimed at detecting regions in a digital image that differ in properties such as brightness or color compared to surrounding regions. A blob is a region of an image in which some properties are constant or approximately constant. All the points in a blob can be considered in some sense to be similar to each other. The most common method for blob detection is convolution.

Blobs are in the form of region for any complementary description of image structures. It opposes for the corners that are more point-like. Blob descriptor may often contain a preferred point (a local maximum of an operator response or a center of gravity) which means that many blob detectors may also be regarded as interest point operators. Blob detector can detect areas in an image which are too smooth to be detected by a corner detector.

D. Ridges

The ridge is a smooth function of two variables of curve whose points are in one or more ways to be made precise. The local maximum of the function is at least one dimension.

This notion captures the intuition of geographical ridges. For elongated objects the notions of ridge are natural tools. A ridge descriptor computed from grey-level images can be seen as a generalization of a medial axis. From practical viewpoints a ridge can be thought of as a one-dimensional curve that represents an axis of symmetry and in addition has an attribute of local ridge width associated with each ridge points.

IV. METHODOLOGY

Computer graphic images are difficult to discriminate from photographic images by human. But their statistical properties are different. The proposed approach for Photographic (PG) and Computer graphic (CG) image classification is presented using Maximum Likelihood Principal Component Analysis (MLPCA) and Convolutional Neural Network. Input image is pre-processed and features are extracted from the pre-processed image using Maximum Likelihood Principal Component Analysis (MLPCA) and remove the non-important features. Finally, CNN classifier is used for differentiating PG and CG images.

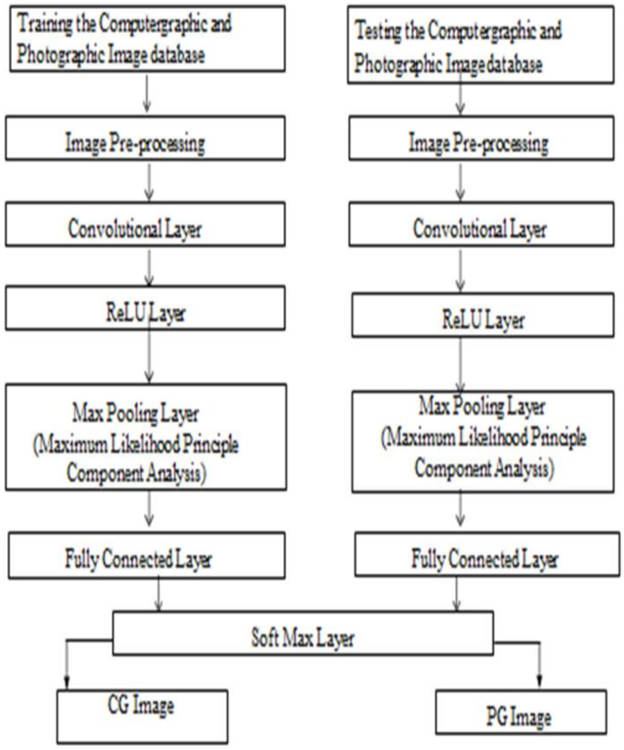

Fig: Flow Diagram of Proposed Method

1) Training the Computer graphic and Photographic Image database.
2) Testing the Computer graphic and Photographic Image database.

Initially, the Computer graphic and Photographic image database are trained.The color image is given to the convolution layer.In this stage preprocessing operation such as filter or mask can performed.The filtered image can be given into ReLU layer.ReLU layer changes the all negative actions to Zero.After Max pooling layer can perform the feature extraction and reduce the dimensionality of the image.Fully connected layer which are used to generate new features from the existing features.Softmax layer is a classification layer it can be used to classify the computergraphic images from photographic images.

A. Input Image

An image is an array, or a matrix, of square pixels (picture elements) arranged in columns and rows .The input element, having the "image" value in its type attribute, represents a graphical submit button, which is a regular image that when pressed, submits the form it belongs to. A digital image is an image composed of picture elements, also known as pixels, each with finite, discrete quantities of numeric representation for its intensity or gray level that is an output from its two-dimensional functions fed as input by its spatial coordinates denoted with x, y on the x-axis and y-axis.

Fig1: Sample of Input images

B. Image Pre-Processing

The aim of pre-processing is the improvement of image data and suppression of unwanted distortions. It enhances some image features that are important for further processing. In this project first the input image is converted into grey scale image for pre-processing.In a greyscale image each picture element has an assigned intensity that ranges from 0 to 255.

Fig2: Sample of Grey scale images

C. Convolution Layer

Convolution layer involves the shift,multiply and sum operations.The main processing component of this layer is a filter or mask which is a matrix of weights.The convolution layer can remove the unwanted noise from the grayscale images.

Fig3:Sample of filtered images

D. ReLU layer

ReLU layer just changes all the negative activations to 0. The purpose of applying the rectifier function is to increase the non-linearity in our images. The reason we want to do that is that images are naturally non-linear.When you look at any image, you'll find it contains a lot of non-linear features.

The rectifier serves to break up the linearity even further in order to make up for the linearity that we might impose an image when we put it through the convolution operation. Rectifier function does to an image like this is remove all the black elements from it, keeping only those carrying a positive value (the grey and white colors).

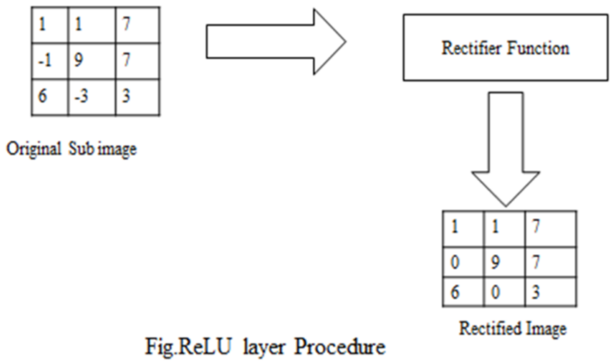

Fig.ReLU layer Procedure

E. Max Pooling Layer

Max pooling is a sample-based discretization process. The objective is to downsample an input representation (image, hidden-layer output matrix, etc.), reducing its dimensionality and allowing for assumptions to be made about features contained in the sub-regions binned.

In this Paper using Maximum Likelihood Principle Component Analysis(MLPCA) algorithm.MLPCA algorithm is a dimensionality reduction algorithm. MLPCA algorithm will be used to extract the features from the input images.

Dimension of an input image	Feature Vector Size
128 X 128 double	1 X 12800 double

F. Fully Connected Layer

Fully connected layers are an essential component of Convolutional Neural Networks (CNNs), which have been proven very successful in recognizing and classifying images for computer vision. The CNN process begins with convolution and pooling, breaking down the image into features, and analyzing them independently.Fully connected layers connect every neuron in one layer to every neuron in another layer. It is in principle the same as the traditional multi-layer perceptron neural network (MLP). The flattened matrix goes through a fully connected layer to classify the images.Fully connected layers are an essential component of Convolutional Neural Networks (CNNs), which have been proven very successful in recognizing and classifying images for computer vision. The CNN process begins with convolution and pooling, breaking down the image into features, and analyzing them independently.

The result of this process feeds into a fully connected neural network structure that drives the final classification decision.

The objective of a fully connected layer is to take the results of the convolution/pooling process and use them to classify the image into a label. The output of convolution/pooling is flattened into a single vector of values, each representing a probability that a certain feature belongs to a label. For example, if the image is of a cat, features representing things like whiskers or fur should have high probabilities for the label cat‖.The fully connected part of the CNN network goes through its own backpropagation process to determine the most accurate weights. Each neuron receives weights that prioritize the most appropriate label. Finally, the neurons vote‖ on each of the labels, and the winner of that vote is the classification decision.

G. Softmax Layer

The softmax function can be used in a classifier only when the classes are mutually exclusive. Many multi-layer neural networks end in a penultimate layer which outputs real-valued scores that are not conveniently scaled and which may be difficult to work with. The softmax function, also known as softargmax or normalized exponential function, is a generalization of the logistic function to multiple dimensions. It is used in multinomial logistic regression and is often used as the last **z** of a neural network to normalize the output of a network to a probability distribution over predicted output classes.

The softmax function takes as input a vector z of K real numbers, and normalizes it into a probability distribution consisting of K probabilities proportional to the exponentials of the input numbers. That is, prior toapplying softmax, some vector components could be negative, or greater than one; and might not sum to 1; but after applying softmax, each component will be in the interval (0,1) and the components will add up to 1, so that they can be interpreted as probabilities. Furthermore, the larger input components will correspond to larger probabilities.The standard (unit) softmax function is defined by the formula

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K \text{ and } \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K$$

V. STATISTICAL FEATURES EXTRACTION

This Section explains how statistical information is extracted from the convoluted images. In a deep-learning context, this operation can be viewed as a pooling layer. Usually, after convolution layers, a local maximum pooling is computed to reduce the dimension of the data representation before classification. As for image forensics, other global statistical quantities are known to be more useful. Thus a special pooling layer is developed for adapting neural nets to this particular task. We explored two different approaches: computing simple statistics or estimating the histogram of the convoluted images

A. Simple statistics

deriving really simple quantities from filtered images, in general low order moments of the distribution, is enough to solve our classification problem.

B. Histograms

The second option is to compute the normalized histogram of the pixel distribution which may capture more information than simple statistical quantities [12]. In order to integrate this layer into our framework, we adapted the way bin values are usually calculated. Their computation includes the use of indicator functions which gradients value zero almost everywhere. This prevents the weights of the convolution layers to be updated during back-propagation. This issue can be avoided using Gaussian kernels instead of indicator function to estimate the value of each histogram.

VI. CLASSIFICATION

Train a classifier to separate the two classes CG and PG. We tried the following popular method.

A. Maximum Likelihood Principle Component Analysis(MLPCA)

The theoretical principles and practical implementation of a new method for multivariate data analysis, maximum likelihood principal component analysis (MLPCA), are described. MLCPA is an analog to principal component analysis (PCA) that incorporates information about measurement errors to develop PCA models that are optimal in a maximum likelihood sense. The theoretical foundations of MLPCA are initially established using a regression model and extended to the framework of PCA and singular value 31 decomposition (SVD).

An efficient and reliable algorithm based on an alternating regression method is described. Generalization of the algorithm allows its adaptation to cases of correlated errors provided that the error covariance matrix is known. Models with intercept terms can also be accommodated. Simulated data and near- infrared spectra, with a variety of error structures, are used to evaluate the performance of the new algorithm. Convergence times depend on the error structure but are typically around a few minutes. In all cases, models determined by MLPCA are found to be superior to those obtained by PCA when non-uniform error distributions are present, although the level of improvement depends on the error structure of the particular data.

VII. OUTPUT AGGREGATION

A regular image is usually big enough to be composed of many tiles. The classification of a full image will be decided according to the classification probability of each of its tiles. In practice, we use a weighted voting scheme where each patch contribution is the log likelihood of the label:

$$y_{pred} = \text{sgn} \sum_{i=1}^{Np} \log \left(P(Y = 1 \mid Xi = xi) \,/\, P(Y = -1 \mid Xi = xi) \right)$$

with Y and Xi the random variables modeling the labels and the patches, xi the real observations and sgn(x) a function that returns ±1 according to the sign of x. We used this rule because it is fairly easy to obtain posterior probabilities (with our MLP for example, it just corresponds to the two output values of the read-out layer) but also because it can be interpreted as a global Maximum Likelihood Estimation criterion for the parameter Y.

VIII. TESTS AND RESULTS

A. Database

CG were downloaded from the Level-Design Reference Database [14] which contains more than 60,000 good resolution (1920×1080 pixels) video-game screenshots in JPEG format. Only 5 different video-games were judged photo-realistic enough which reduces the set to 1800 images.

The Photographic images are high resolution images (about 4900 × 3200 pixels) taken from the RAISE dataset [15], directly converted from RAW format to JPEG.

From those 3600 images, we constructed 3 databases on which our tests were carried out. Firstly, we selected the green channel of each image. Each class was then divided into training (70%), testing (20%) and validation (10%) to form the Full-size database. From this original database, we constructed a lower size one by cropping each image to 650×650. Finally, we randomly extracted 43000 patches sized at 100 × 100 for training the patch classifier.

TABLE I: CG vs PG datasets description

Name	Ntrain	Ntest	Nval	Size
VFull-size	2520	720	360	Various
Low-Size	2520	720	360	650 × 650
Patch	40000	2000	1000	100 × 100

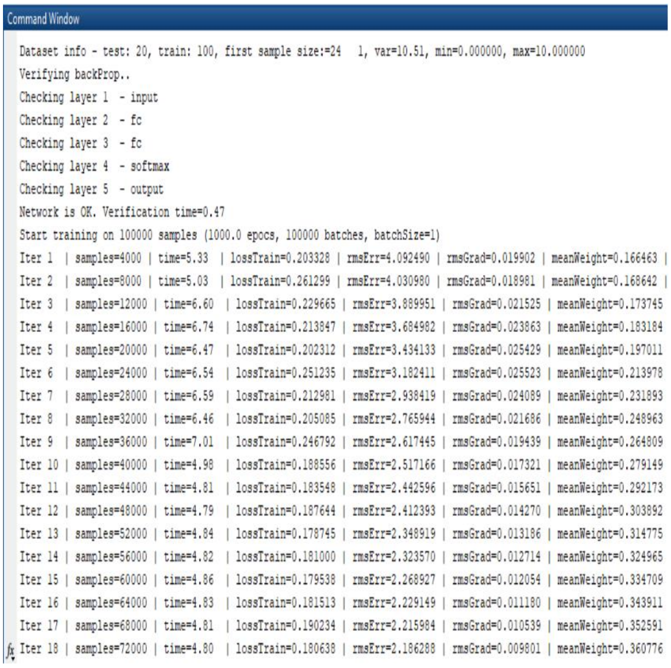

Fig4:CNN classification Training process

A. Experimental setup

Our method was implemented using the Tensorflow framework . Our statistical and histogram layers were directly integrated to the pipeline, using Tensorflow native functions.

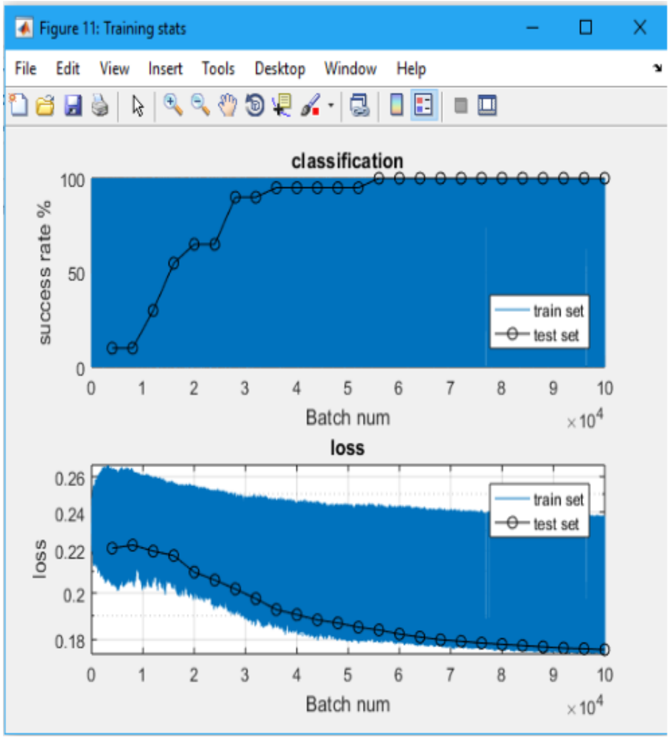

Fig5:CNN Classification Success rate and Loss

Number of Images in each class	Accuracy (in %)
50	71.87
100	72
150	72.85
200	73.05
250	73.24
300	74.67
350	75
400	75.82
450	76.17
500	77.50
550	78
600	80.05
650	80.24
700	80.37
750	80.67
800	81

Comparision of Classification Accuracy Table

B. Accuracy Calculation

ACCURACY = (TP+FP) / TP+TN+FP+FN

Parameter	Proposed	Existing
Accuracy	0.81	0.68

IX. CONCLUSION

Classification of computergraphics from photographics has become an important research topic in the field of passive image authentication. Photographic image and computergraphic image feature extracted by using Maximum Likelihood Principle Component Analysis(MLPCA).CNN classifier is used for classification. Accuracy of 81% is obtained for classifying 800 photographic images from 800 computergraphic images.

X. REFERENCES

[1]. Archana VMire, S.B. Dhok, N JMistry, P. D. Porey, “Resampling Detection in Digital Images: A Survey”, Volume 84 – No 8, IJCA-2013.

[2]. Duc-tien Dang-Nguyen,Giulia boato and Francesco G.B.Denatale, “Discrimination between computer generated and natural human faces based on asymmetry information”, EUSIPCO-2012.

[3]. Hanghang Tong, Mingjing Li, Hong Jiang Zhang, Jingrui He, Changshui Zhang, “Classification of Digital Photos Taken by Photographers or Home Users”, pp.198- 205,Volume 3331,PCM-2004.

[4]. Jassem Mtimet, Hamid Amiri, “Image Classification using Statistical Learning Methods”, pp.200-203,JSEA-2012

[5]. Michael Haller, “Photorealism or/and Non-Photorealism in Augmented Reality”,pp.189-196,VRCAI-2004.

[6]. Micah K. Johnson, Kevin Dale, Shai Avidan, Hanspeter Pfister, William T. Freeman, “CG2Real: Improving the Realism of Computer Generated Images using a Large Collection of

Photographs", pp.1273-1285,Volume:17, Issues:9, ITVCG-2010.

[7]. Olivia Holmes, Martin s. Banks, Hany Farid, "Assessing and Improving the Identification ofComputer-Generated Portraits",ACM Transactions on Applied Perception-2016.42

[8]. Peisong He, Xinghao Jiang, Tanfeng Sun, Haoliang Li, "Computer Graphics Identification Combining Convolutional and Recurrent Neural Network",pp.1369-1373,Volumes:25,Issue:9,ISPL-2018.

[9]. Rong Zhang, Rang-Ding Wang, and Tian-Tsong NgY.Q. Shi, H.J. Kim, and F.Perez-Gonzalez, "Distinguishing Photographic Images and Photorealistic Computer Graphics Using VisualVocabulary on Local Image Edges", pp.292– 305, IWDW - 2011.

[10]. Tian-Tsong Ng and Shih-Fu Chang, "An Online System for Classifying Computer Graphics Images from Natural Photographs", SPIE-2006.

[11]. Tian-Tsong Ng, Shih-Fu ChangJessie Hsu, Lexing Xie, "Physics-Motivated Features for Distinguishing Photographic Images and Computer Graphics", pp.239-248,MULTIMEDIA-2005.

[12]. Wen Chen, Yun Q. Shi, Guorong Xuan, "Identifying computer graphics using HSV color model and statistical moments of characteristic functions",pp.189- 196,VRCAI-2004.

[13]. Ye Yao, Weitong Hu, Wei Zhang, Ting Wu and Yun-Qing Shi, "Distinguishing Computer-generated Graphics from Natural Images Based on Sensor Pattern Noise and Deep Learning",MDPI-2018.

[14]. Yingda Lv, XuanJing Shen, Guofu Wan, HaiPeng Chen, "Blind Identification of Photorealistic Computer Graphics Based on Fractal Dimensions",CCIT-2014.43

[15]. Yuanhao Chen,Zhiwei Li,Mingjing Li,Wei-YingMa,— "Automatic Classification of Photographs and Graphics",ICME-2006.

[16]. R. Wang, S. Fan, and Y. Zhang, "Classifying computer generated graphics and natural imaged based on image contour information," J. Inf. Comput. Sci, vol. 9, no. 10, pp. 2877–2895, 2012.

[17]. Y. Wang and P. Moulin, "On discrimination between photorealistic and photographic images," in Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, vol. 2. IEEE, 2006, pp.

[18]. X. Cui, X. Tong, G. Xuan, and C. Huang, "Discrimination between photo images and computer graphics based on statistical moments in the frquency domain of histogram," Procedings of the CIHW2007, Nanjing, pp. 276–283, 2007.

[19]. W. Chen, Y. Q. Shi, G. Xuan, and W. Su, "Computer graphics identification using genetic algorithm," in 19th International Conference on Pattern Recognition, ICPR 2008. IEEE, 2008, pp. 1–4. [20] R. Wu, X. Li, and B. Yang, "Identifying computer generated graphics via histogram features," in 18th IEEE International Conference on Image Processing (ICIP). IEEE, 2011, pp. 1933–1936.

[20]. W. Li, T. Zhang, E. Zheng, and X. Ping, "Identifying photorealistic computer graphics using second-order difference statistics," in 2010 Seventh International Conference on Fuzzy Systems an Knowledge Discovery (FSKD), vol. 5. IEEE, 2010, pp. 2316–2319.

[21]. N. Khanna, G. T.-C. Chiu, J. P. Allebach, and E. J. Delp, "Forensic techniques for classifying scanner, computer generated and digital camera images," in IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2008. IEEE, 2008, pp. 1653–1656.

[22]. A. C. Gallagher and T. Chen, "Image authentication by detecting traces of demosaicing," in IEEE Computer Society Conference on Computer Vision and Pattern

Recognition Workshops, CVPRW'08. IEEE, 2008, pp. 1–8.

[23]. E. Tokuda, H. Pedrini, and A. Rocha, "Computer generated images vs. digital photographs: A synergetic feature and classifier combination approach," Journal of Visual Communication and Image Representation, vol. 24, no. 8, pp. 1276–1292, 2013.

[24]. T.-T. Ng and S.-F. Chang, "Discrimination of computer synthesized or recaptured images from real images," in Digital Image Forensics. Springer, 2013, pp. 275–309.