ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

An Image Tampering Detection Algorithm of Qualification Certificate Based on CNN and SVM

Zhongwen Qian[a], Ye Gu[b,*], Wenming Hong[c]

State Grid Zhejiang Procurement Company, Hangzhou, Zhejiang, China
[a]qian_zw@sina.cn, [b]173311327@qq.com, [c]hongwenming@zj.sgcc.com.cn
**Ye Gu*

Abstract: *With the high speed development of digital image processing technology, the phenomenon of malicious tampering with certificate images is rampant. Facing the problems of difficult human eye recognition of tampering traces and poor detection and classification effect of CNN model, a tampering detection algorithm based on CNN combined with SVM is proposed. The algorithm first performs data preprocessing of image compression, difference, resise, and normalization on the image. Then it uses the CNN model to extract the tamper feature, finally completes the calculation and classification of the fully connected layer feature data through SVM, and builds a certified image CNN+SVM model with tamper detection capability. Experimental results show that comparing with the traditional human eye recognition method and the existing deep learning CNN model, the algorithm can reach a detection accuracy of 48.09%, 91.79 %, and 98.67% under the self-built data set of tampering with certificate images. The detection accuracy rate is higher than the previous two detection accuracy rates.*

Keywords: *Convolutional neural networks, support vector machine, qualification certificate image tampering, deep learning*

1. Introduction

Image tampering refers to the purposeful modification of the image content by copying and pasting, splicing and combining, and various transformations, so as to confuse and distort the meaning expressed by the original image, thereby achieving a certain benefit, demand or purpose[1]. With the continuous progress of computer technology, the rapid development of network technology and the increasingly powerful digital image processing software make image tampering easier, more convenient and frequent[2]. However, it is difficult to observe digital image modifications with naked eyes, so it brings convenience to people and meanwhile lays hidden dangers, especially the proliferation of malicious tampering with certificate images, which has a significant impact on the engineering field, public opinion and national security. Therefore, tamper detection has become an important research field as a scheme to identify the integrity and originality of digital multimedia data[3].

Image tampering detection is to complete the detection and determination of the tampering range through various algorithms based on the characteristics of the image. Image tampering detection technology is currently mainly divided into active detection technology and passive detection technology.[4-6] Active detection technology requires watermarking or digital signature processing in advance, while passive detection technology does not perform any pre-processing, and directly analyzes and collects evidence based on the characteristics of the image to be detected. Passive detection technology is recognized as a method of blind forensics, with better adaptability, wider application range, and higher research value.

In recent years, many scholars have proposed a variety of algorithms for image tampering detection, such as N.Ahmed et al. proposed the Discrete Cosine Transform (DCT), which uses the double quantization feature of the algorithm compressed for specific JPEG images to distinguish the traces of tampering[7-8]; David Lowe et al. proposed the SIFT algorithm for image copy-paste passive forensics. The algorithm is more stable in the recognition of local feature tampering; while the HOG algorithm proposed by Navneet Dalal et al. is for object detection feature description, using calculation and statistics of the gradient direction histogram of the local area of the image to determine whether the image has been tampered. In summary, a large part of the existing algorithms are designed to detect

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

certain specific tampering scenarios. These artificially constructed features are highly pertinent, but have poor adaptability. They can only be used in certain specific problems and relatively short-term scenarios. Effective within time, the accuracy of detection is also easily affected.

However, the rapid development of deep learning theory provides a new idea for image tampering detection technology: the use of deep learning methods to extract image features to characterize the tampered area in the image[9-10] Among them, the representative CNN model directly relies on the data and is applicable to almost all images[11]. It has strong applicability and greatly reduces the time and effort to analyze and select specific features; but at the same time, the CNN softmax classification of the model has high requirements on the data sample set, and the accuracy is relatively limited[12]. Therefore, this paper constructs a tampering detection algorithm based on CNN combined with SVM. This algorithm performs image compression,difference, resise, and normalized data preprocessing on the image. The CNN model is used to extract the tampering feature, and the fully connected layer feature data is calculated and classified through SVM[13], and the CNN+SVM model is constructed to realize the detection of the tampering of the certificate image. While using CNN deep learning technology, ELA is used to enhance the features of tampered images, and combined with SVM classification, it reduces the requirement of CNN original softmax classification for sample sets, enhances robustness and improves accuracy.

2. Detection Algorithm based on CNN and SVM

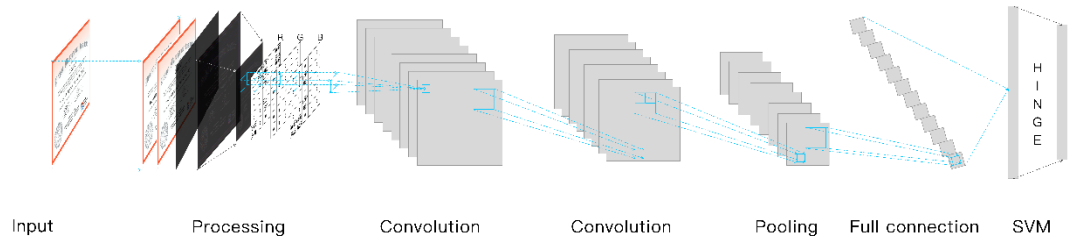

Figure 1: The process of the tampering detection algorithm for certificate image based on CNN and SVM.

Figure 1 shows the process of the tampering detection algorithm based on CNN and SVM. First, perform data preprocessing on the input image, and get the compressed image, the difference image, the brightening image, the resize image, and the normalized image successively; then, configure the convolution parameters and pooling parameters to normalize the three images R, G, and B. The channel value is convolved twice and pooled once; finally, the pooling result is expanded and fully connected, and the classification of tampering is completed through the Hinge function of the SVM[14-15].

In this paper, the experimental process of the algorithm is proposed. A tampered certificate image is substituted into the calculation. The default parameters of each stage are used as follows:

- In the image data preprocessing stage, the compression ratio of the compressed image is 90%, the size of the Resize image is 128×128, the bicubic interpolation (Bicubi) method is used, and the image normalization uses the maximum and minimum normalization (linear function conversion) method.
- In the two convolution stages, the number of convolution kernels for the first convolution and the second convolution is 32, the size is 5×5, the step size is 1×1, and the ReLU activation function is adopted.
- In the primary pooling stage, the pooling method adopts the maximum pooling, the pooling window adopts 2×2, and the pooling step size is 2×2.

The experiment process attempts to obtain the best parameter configuration of conventional CNN under the scene of tampering with certificate images (self-built data set) from the parameter adjustment experiments at different stages. This paper proposes the best parameter configuration of CNN+SVM and the comparison and analysis of the two before and after.

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

2.1. Image Data Preprocessing

Image data preprocessing refers to the process performed before feature extraction, segmentation and matching of the input image in image analysis. In the CNN model, methods such as purification, data normalization, image size compression, and data enhancement are usually used to eliminate irrelevant information in the image, restore useful real information, enhance the detectability of related information, and maximize the Simplify the purpose of data and increase the reliability of feature extraction, image segmentation, matching and recognition[16].

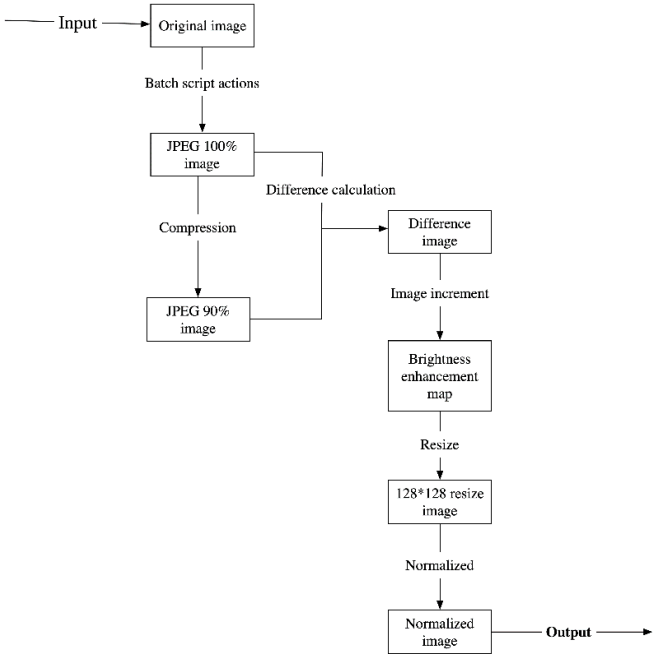

Figure 2: Image data preprocessing process.

The image data preprocessing process of the algorithm proposed in this paper is shown in Figure 2. The training samples are processed into a format that can be recognized by the convolutional network through difference calculation, image brightening, image resize, and image normalization; at the same time, the original is removed. There may be data in the data that is not suitable for this algorithm. Strengthen the feature attributes in the image data and remove the unit limit of the value, so as to improve the convergence speed of the model and improve the speed and accuracy of model detection.

The first step is to use image processing software to execute batch action scripts for the input pictures, convert the format to JPEG format uniformly, and compress JPG images to obtain new pictures in JPEG format with a quality parameter of 90%, without naked eye recognition before and after compression Degree, as shown in Figure 3 and Figure 4.

The second step is to calculate the original image image1 of the certificate and the new image image2 after 90% quality compression pixel by pixel, where out is the output image, as shown in Figure 5.

Figure 3: The original image of the certificate.

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

Figure 4: The 90% quality image of the certificate compressed.

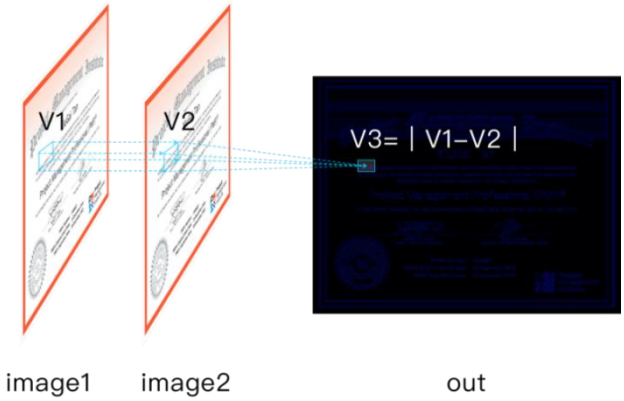

Figure 5: Difference calculation diagram.

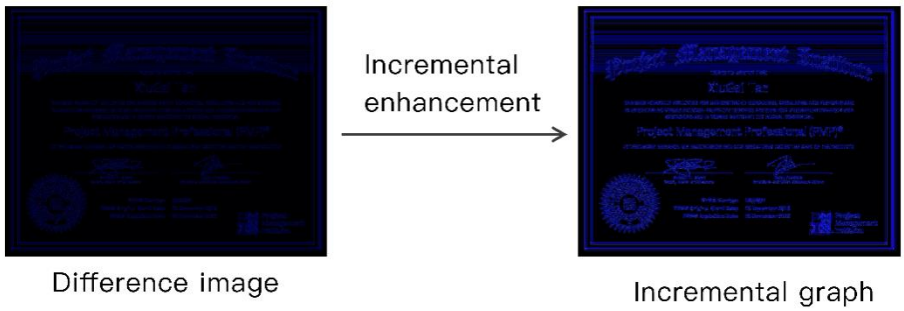

Figure 6: Image brightening map.

The third step is to use the brightness adjustment method to achieve the purpose of image enhancement. According to the difference map RGB value transformation $out = image1 \times alpha$, where out is the output image, image is the input image, and alpha is the constant that controls the enhancement, as shown in Figure 6.

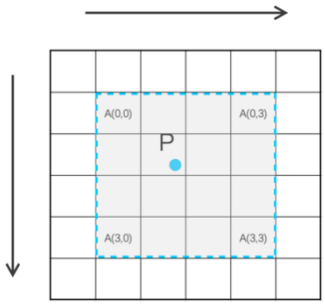

Figure 7: Schematic diagram of the surrounding 16 pixels distribution.

In the fourth step, the Bicubi method is used to perform the Resize calculation to form a Resize graph with a size of 128×128. First, find the 16 pixels around the nearest coordinate in the original image, as shown in Figure 7.

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

Then, construct the Bicubi function to calculate the weights in the x-axis and y-axis directions, calculated in the following way:

$$W_{(x)} = \begin{cases} (a+2)|x|^3 - (a+3)|x|^2 + 1, & |x| <= 1 \\ a|x|^3 - 5a|x|^2 + 8a|x| - 4a, & 1 < |x| < 2 \\ 0, & |x| >= 2 \end{cases} \quad (1)$$

Among them, |x| refers to the distance between the target coordinate and the nearest coordinate axis around;

Finally, the value and the weight are multiplied and summed to obtain the pixel value of the Resize target coordinate. The calculation formula is as follows:

$$\mathbf{f(x,y)} = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} , w_{(i)} w_{(j)} \quad (2)$$

Among them, represents the pixel value of the coordinate point in the source image, w(i) represents the weight of the coordinate in the source image on the x-axis, w(j) represents the weight of the coordinate in the source image on the y-axis, f(x,y) Represents the pixel value of the desired target. The input certificate image Resize is shown in Figure 8:

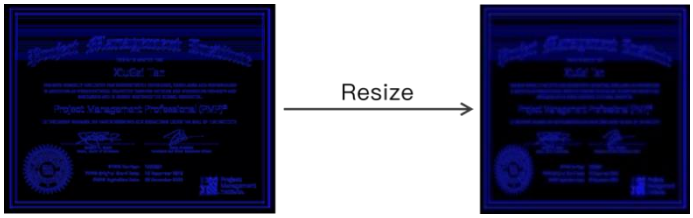

Figure 8: Schematic diagram before and after Resize.

2.2. CNN image Feature Extraction

The product neural network was proposed by Yann Lecun in 1998. The algorithm is good at processing and recognizing various images. It is a neural network specially used to process data with a similar grid structure. It is composed of convolutional layers and down-sampling pooling. It consists of three parts: Pooling and Fully connected[17]. Generally, CNN is mainly used to identify displacement, scaling and other forms of distortion invariance of two-dimensional graphics, especially its special structure of local weight sharing has unique advantages in speech recognition and image processing, and its layout is closer to the actual situation. In the biological neural network, weight sharing reduces the complexity of the network[18-19].

The feature extraction process of the algorithm proposed in this paper is mainly completed by two convolution and one pooling processes. The two convolution operations are mainly to reduce the dimensionality and feature extraction of the input image, using the ReLU activation function; the one-time pooling process is mainly to reduce the model size, increase the calculation speed, reduce the probability of overfitting, and improve the robustness[20].

2.3. Convolution

A convolution in this article is a normalized image (RGB three channels, size 128×128), output from the preprocessing of the data image through the convolution kernel that can be learned (the number is set to 32, and the size is 5×5), Extract the region of the same size as the convolution kernel from the feature matrix, and then multiply the eigenvalues in the region with the weight of the corresponding position of the convolution kernel in order, and add the offset value as the convolution result of the region, and finally follow The moving step of the convolution kernel calculates the convolution results of other regions in turn. When the convolution is completed, the output feature map size is 124*124, the feature map has 32 layers, and the convolution process of each layer of feature

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

map is shown in Figure 9.

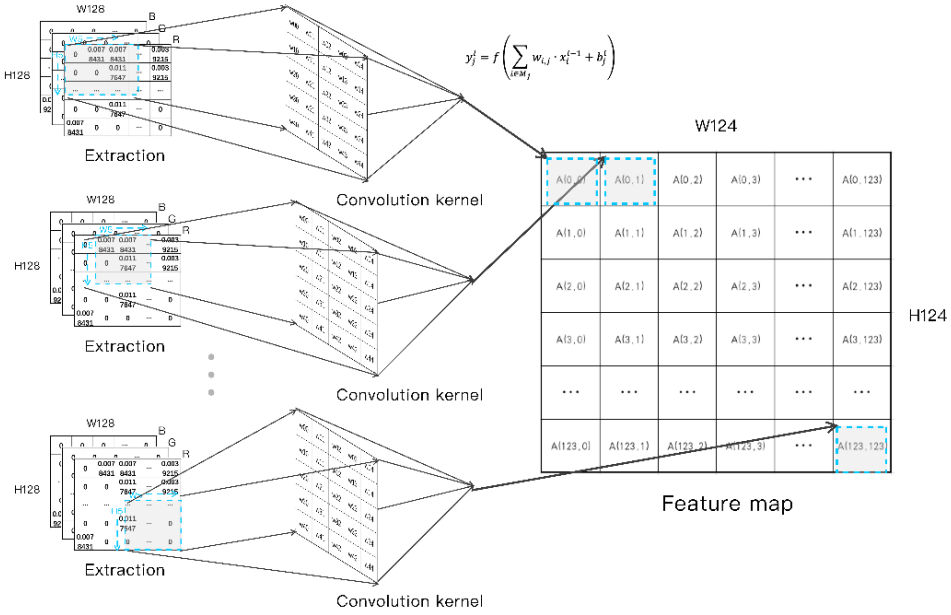

Figure 9: First convolution process.

The convolution formula of each extracted area is:

$$y_j^l = f\left(\sum_{i \in M_j} w_{i,j} \cdot x_i^{l-1} + b_j^l\right) \tag{3}$$

Among them, y_i^l is the feature map of the convolutional layer l whose subscript is j;

$f(\ )$ represents the activation function (in this paper, the ReLU type function $y = max(0, x)$ is nonlinear);

M_i is the set of input subscripts participating in the convolution operation;

x_i^{l-1} represents the input of the upper layer of l whose subscript is i;

$w_{i,i}$ is the convolution kernel between i and j;

b_i^l represents the bias of the convolutional layer l.

The 32-layer feature map output after the inputted certificate normalization image is subjected to a convolution process is shown in Figure 10:

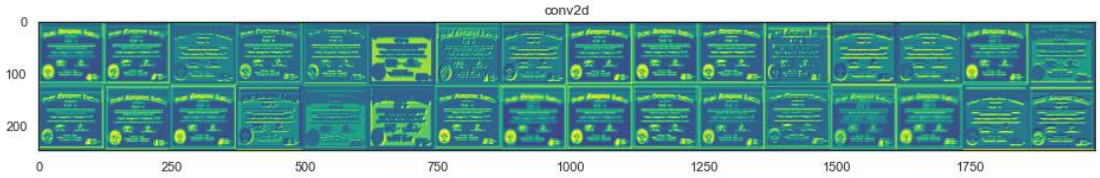

Figure 10: The output result of a convolution.

The secondary convolution in this article also uses the convolution kernel (the number is set to 32, the size is 5×5), and the feature matrix of the 32-layer feature map (size 124×124) is extracted from the feature matrix of the first convolution output. For regions with the same size of the convolution kernel, the convolution result of the region is calculated according to the same convolution method. When the convolution is completed, the output feature map size is 120*120, The feature map has 32 layers, and the convolution process of the feature map of each layer is shown in Figure 11.

The convolution formula of each extraction area is the same as that of the primary convolution. The

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

32-layer feature map output after the secondary convolution process is shown in Figure 12.

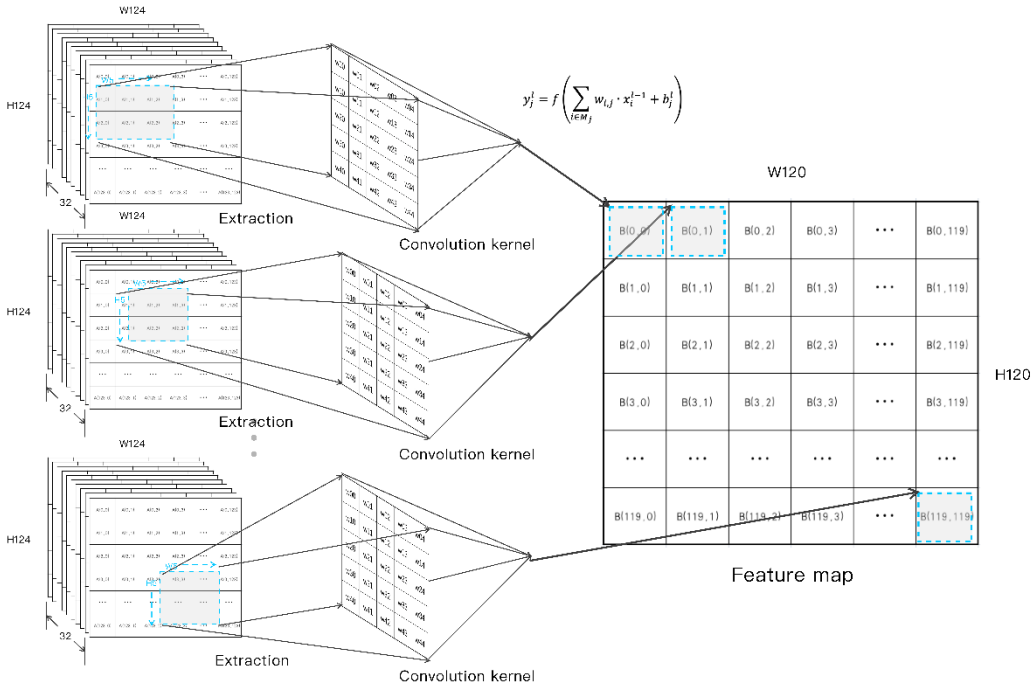

Figure 11: Secondary convolution process.

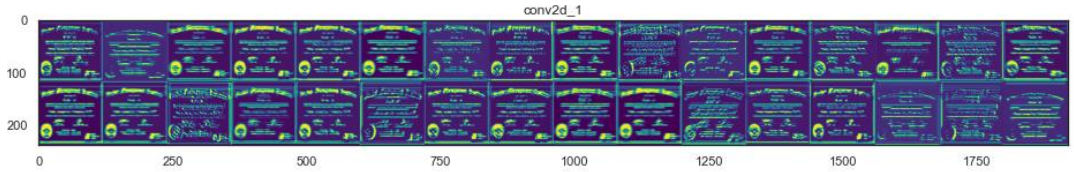

Figure 12: Output result of secondary convolution.

2.4. Pooling

The pooling layer in this article filters and combines visual features into abstract and higher-level visual features by sampling the output results of the secondary convolution. The pooling layer adopts the maximum pooling by default, the pooling window adopts 2×2, and the pooling step size adopts 2×2.The calculation formula is:

$$y_j^l = max\left(y_j^{l-1}\right) \tag{4}$$

Among them, y_j^l is the output of the pooling layer, and y_j^{l-1} is the input of the previous layer. The feature matrix extracted from the previous layer is divided into 2×2 (step size) matrix blocks on average. Take the maximum value for each matrix block, according to the secondary convolution input size and dimension 120×120×32, the final output size dimension is 60×60×32 feature matrix, the maximum pooling process of each layer (32 layers in total) As shown in Figure 13.

The output result of the feature image after processing by the pooling layer is shown in Figure 14.

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

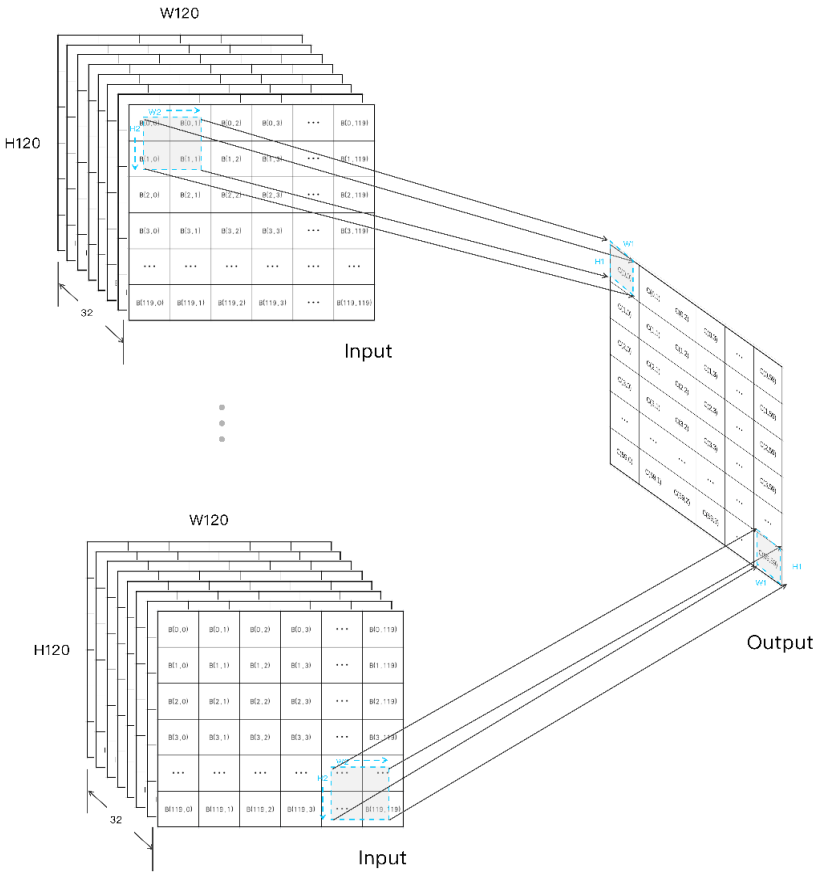

Figure 13: Maximum pooling process.

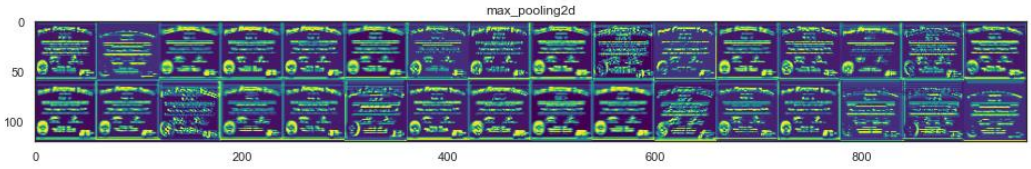

Figure 14: Maximum pooling output result.

2.5. Fully Connected Layer

The fully connected process of the algorithm proposed in this paper firstly expands the 60×60×32 feature matrix of the pooling result as the input basis of the subsequent fully connected layer; then, in order to enhance the nonlinear mapping ability of the network, the previous one is expanded by the features. All neurons of the layer network are connected to all neurons of the current network, as shown in Figure 15.

As shown in the figure above, the development of each layer of 60×60 feature map of 60×60×32 can get fl expanded layer. The connection between fl and all neuron items satisfies the following formula:

$$y_j^l = f\left(\sum_{i=1}^{n} W_{j_i}^l \cdot x_i^{l-1} + b_j^l\right) \tag{5}$$

Among them, l represents the current number of network layers, n represents the number of neurons in the $l-1$th layer of the network, $W_{j_i}^l$ represents the distance between the l th layer of network neuron j and the $l-1$ th layer of network neuron i The connection weight of x_1^{l-1} is the input value of the i th

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

neuron of the $l-1$ network, and b_j^l is the bias of the neuron j of the l th network.

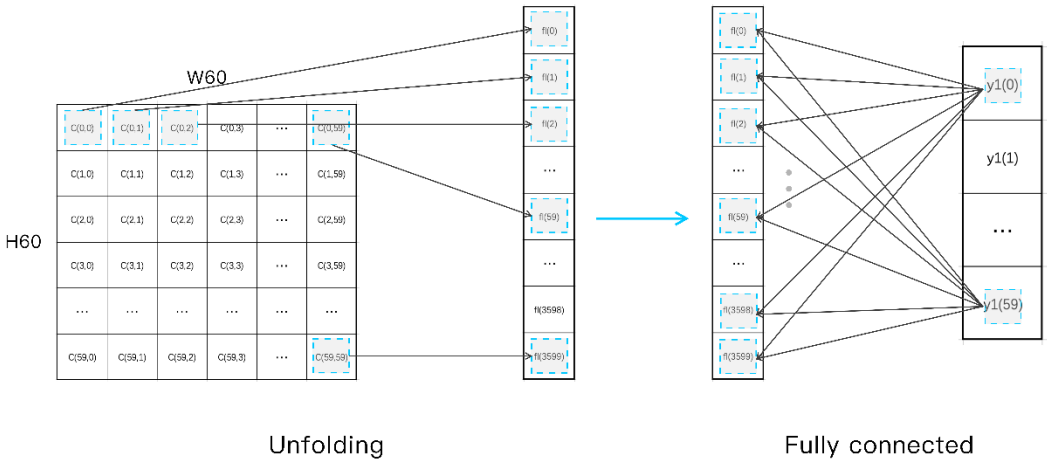

Figure 15: Feature expansion and full connection process.

2.6. SVM

SVM is a new and very promising classification technology proposed by the AT&TBell laboratory research group led by Vanpik in 1963. It is a pattern recognition method based on statistical learning theory. In the field of machine learning, it is a supervised learning model. Usually used for pattern recognition, classification and regression analysis[21].

The SVM classification process proposed in this paper is aimed at practical application scenarios where the tampering detection of certificate images often has limited training samples[22]. According to the limited sample information, a good balance is achieved between the complexity of the model and the learning ability to obtain better generalization ability. The output data of the fully connected layer is used as the input, and the loss function hinge of SVM is used to back-propagate the neural network. The weight and bias of each previous layer are learned. The purpose is to make the loss result as small as possible, so that The predicted value of the classification result is closer to the actual result. The formula of the loss function is as follows:

$$l(y) = \max\left(0,1 - y_t y_p\right) \tag{6}$$

Among them, y_t is the actual value of the classification, and y_p is the predicted value of the classification. In actual implementation, suppose the sample to be classified is $x_i \in M$, and M is the set of samples to be classified. The purpose of the SVM algorithm in this paper is to find a hyperplane:

$$0 = w^T x + b \tag{7}$$

So that it can separate the samples of the two categories, and all the samples are as far away from this plane as possible, and the classification result is that the value of y_i is 1 or -1 to represent different classification results:

$$y_i = \begin{cases} w^T x_i + b \geq 1, y_i = 1 \\ w^T x_i + b \leq -1, y_i = -1 \end{cases} \tag{8}$$

Bring the sample to be classified into the $w^T x + b$ formula. If the output value is >0, it is classified as 1 (tampered), and if it is <0, it is classified as -1 (not tampered), as shown in Figure 16.

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

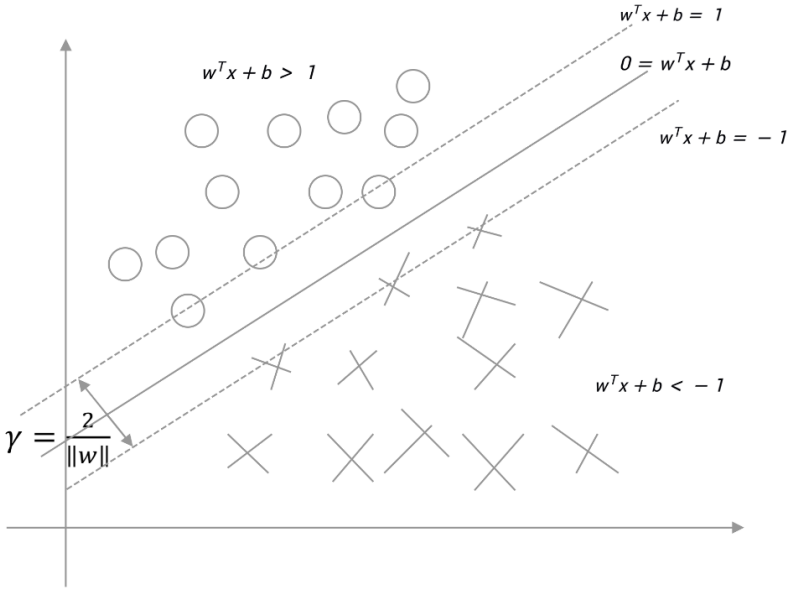

Figure 16: SVM classification hinge function process.

In this paper, input the certificate image to expand the feature matrix, use the model trained in the data set, perform the full connection and substitute it into the hint function calculation of the SVM, and finally get $y = 1.60347416$, the judgment result is 1, that is, the certificate image is a tampered piece. The image is consistent with the tampering "revised" traces on the signature of the certificate image, and the correct test result is obtained.

3. Experimental Results and Discussion

This paper verifies the performance of the proposed algorithm through experiments. The operating environment is an Intel Core i5 quad-core 2.3GHz processor, 8GB of running memory, the programming environment is Python3.8, and the deep learning is based on the Keras framework.

3.1. Second Section

In view of the particularity of the application scenario of the tampering detection of certificate images, the image features are different from the existing public data sets, therefore, this paper uses self-built data sets to conduct experiments. The data set consists of 1504 original certificate images (resolution 1190*498) that are not repeated, and 761 of them are tampered with: 655 are tampered with random content by PS action scripts, and 106 are manually processed through various processes The method completes the tampering, and realizes that the basic PS traces are not recognized by the naked eye and have been eliminated. The data set is divided into training set, validation set and test set according to the ratio of 6:2:2, as shown in Table 1.

Table 1: Data set division table.

Data set	Training set (photos)	Validation set(sheets)	Test set(sheets)
Number of unmodified pictures	447	157	157
Number of tampered pictures	457	143	143

3.2. Parameter Settings

In the algorithm proposed in this paper, the result of substituting a tampered certificate image into the detection will be affected by different parameter settings at each stage. In the image data preprocessing stage, the corresponding parameters involve the compression coefficient setting before and after the compressed image, the setting of the Resize image size and the selection of the

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

interpolation method; different methods used in the image normalization stage will affect the degree of redundancy of image irrelevant information The degree of enhancement and detectability of the real information; the number of convolution and pooling processes, the size and step length of the window, and the number of channels will affect the reliability of image segmentation and feature extraction; tampering with the image classification process for CNN The comparison and analysis before and after the choice of the hinge function of softmax and SVM is the focus of this paper. Therefore, in order to more accurately detect the tampering of the credential image with moderate time complexity, the experimental environment in this paper replaces different parameters under the configuration of the default parameters. The detection results are shown in Table 2 to Table 14(Marked with * is the default parameter), Accuracy refers to the recognition accuracy of tampered images in the environment of different variables; loss value represents the value of the loss function hinge in the environment of different variables; training time is the training time required for the current model in the environment of different variables ; The detection time refers to the time required to identify and tamper a certain picture.

Table 2: Results and timetable of tampering detection under different compression ratios.

Compression ratio	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
* 90%	93.023258	2.096832	688	0.005974
80%	92.691028	2.074883	636	0.005024
70%	94.352162	2.030231	658	0.005488
60%	93.687708	1.965886	639	0.005049

Table 3: Results and timetable of tampering detection under different Resize sizes.

Resize size	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
256×256	93.355483	2.034732	3098	0.020397
$* 128 \times 128$	93.023258	2.096832	688	0.005974
64×64	92.358804	2.038125	149	0.001323

Table 4: Results and timetable of tampering detection under different Resize interpolation.

Resize interpolation	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
* Bicubic	93.023258	2.096832	688	0.005974
Nearest	98.671097	2.061803	646	0.005913
Bilinear	91.362125	1.975484	633	0.004917
Lanczos	91.029900	2.092229	635	0.005043

Table 5: Results and timetable of tampering detection under different normalizations.

Image normalization	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
*Linear function conversion	93.023258	2.096832	688	0.005974
Logarithmic function conversion	90.033221	2.037509	639	0.005102
Arc cotangent function conversion	89.368773	2.005045	661	0.005572

Table 6: Results and timetable of tampering detection under different convolution kernel sizes.

The first convolution kernel size	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
* 5*5	93.023258	2.096832	688	0.005974
7*7	92.358804	1.968510	644	0.005278
3*3	89.368773	2.066447	665	0.004964

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

Table 7: Results and timetable of tampering detection under different number of convolution kernels at a time.

Number of first convolution kernels	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
64	92.691028	2.064287	1173	0.008149
* 32	93.023258	2.096832	688	0.005974
16	91.694355	2.015028	448	0.003335

Table 8: Results and timetable of tampering detection under different conditions of a convolution kernel step size.

The first convolution kernel step size	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
5×5	59.136212	2.327825	31	0.000659
3×3	92.358804	2.135965	83	0.001004
$*1 \times 1$	93.023258	2.096832	688	0.005974

Table 9: Results and timetable of tampering detection under different numbers of secondary convolution kernels.

Number of second convolution kernels	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
64	91.362124	1.979561	958	0.008269
* 32	93.023258	2.096832	688	0.005974
16	92.358804	2.056703	453	0.003407

Table 10: The tampering detection results and timetable under different sizes of the secondary convolution kernel.

The second convolution kernel size	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
7×7	94.352162	2.140940	1029	0.008063
$*5 \times 5$	93.023258	2.096832	688	0.005974
3×3	92.691028	1.970268	388	0.003611

Table 11: The results and timetable of tampering detection under different steps of the secondary convolution kernel.

Second convolution kernel step size	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
5×5	77.408636	2.192767	122	0.001655
3×3	90.033221	1.973250	152	0.001785
$*1 \times 1$	93.023258	2.096832	688	0.005974

Table 12: The results and timetable of tampering detection under different pooling methods.

Pooling method	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
* Max pooling	93.023258	2.096832	688	0.005974
Average pooling	91.029900	2.043743	683	0.005586

Table 13: Results and timetable of tampering detection under different pooling windows.

Pooling window	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
6×6	89.368773	2.249599	296	0.004588
4×4	90.697676	2.209584	548	0.005094
$*2 \times 2$	93.023258	2.096832	688	0.005974

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

Table 14: Results and timetable of tampering detection under different classification methods.

Classification method	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
softmax	91.796011	0.220358	180	0.001904
* hinge	93.023258	2.096832	688	0.005974

From Table 2, Table 3, Table 4, and Table 5, it can be seen that in the image data preprocessing process, when the compression ratio is 70%, Resize is 256*256, Resize interpolation is Nearest, and normalized to linear function conversion, it has a higher rate of tampering identification accuracy.

It can be seen from Table 6, Table 7, and Table 8 that in a convolution process, the size of the convolution kernel is 5*5, the number of convolution kernels is 32, and the step size of the convolution kernel is 3*3, the accuracy rate is higher; as can be seen from Table 9, Table 10 and Table 11, in the secondary convolution process, the size of the convolution kernel is 7*7, the number of convolution kernels is 32, and the step size of the convolution kernel is 1. *1, the recognition accuracy is high.

It can be seen from Table 12, Table 13, and Table 14 that in the pooling process, the maximum pooling is used, and the pooling window is 2*2, and the pooling step size is 2*2, when the detection time is similar, the recognition accuracy is higher.

3.3. Experimental Results and Comparison

The algorithm proposed in this paper is aimed at the tampering detection scene of the certificate image, using a self-built data set as the experimental data, and comparing the three tampering detection methods of human eye, CNN, SVM+CNN, to further verify the effectiveness of the algorithm in this paper. Obtained from Table 15:

Table 15: Detection results and timetables corresponding to different tampering methods.

Tamper detection	Accuracy(%)	Loss value	Training time (s)	Detection time (s)
human eye	48.090000			
CNN	91.796011	0.220358	180	0.001904
CNN+SVM	98.671097	2.061803	646	0.005913

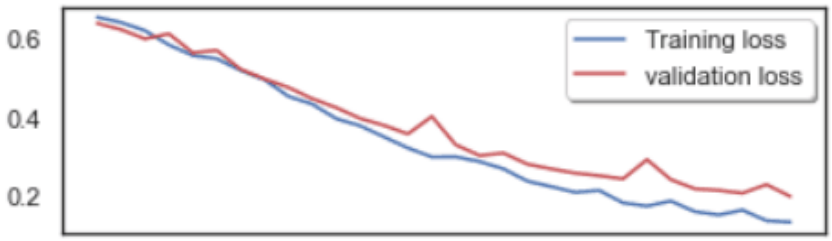

Figure 17: Loss curve of CNN.

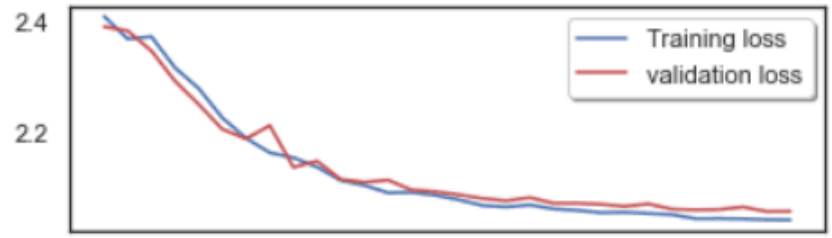

Figure 18: Loss curve of CNN+SVM.

The CNN+SVM algorithm proposed in this document tampering scenario can perform image tampering detection like the CNN algorithm, and compared with the CNN algorithm and the human eye to identify tampered images, it has a higher recognition accuracy, and its accuracy is up to 98.67%, it can effectively solve the problem that the current digital image is difficult to be recognized by the naked eye after modification, especially for the difficult recognition problem in the data tampering scenario in this article.

In addition, although compared with the CNN algorithm, the CNN+SVM model consumes more

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

time during the training process, but it can effectively improve the recognition accuracy of 8%. From Table 15, it can be seen that the CNN+SVM model is training on a small sample set. Time is more advantageous. For the current scenario of small sample sets of certification images in the corporate procurement process, the actual application is stronger, and the advantages are obvious under the same conditions.

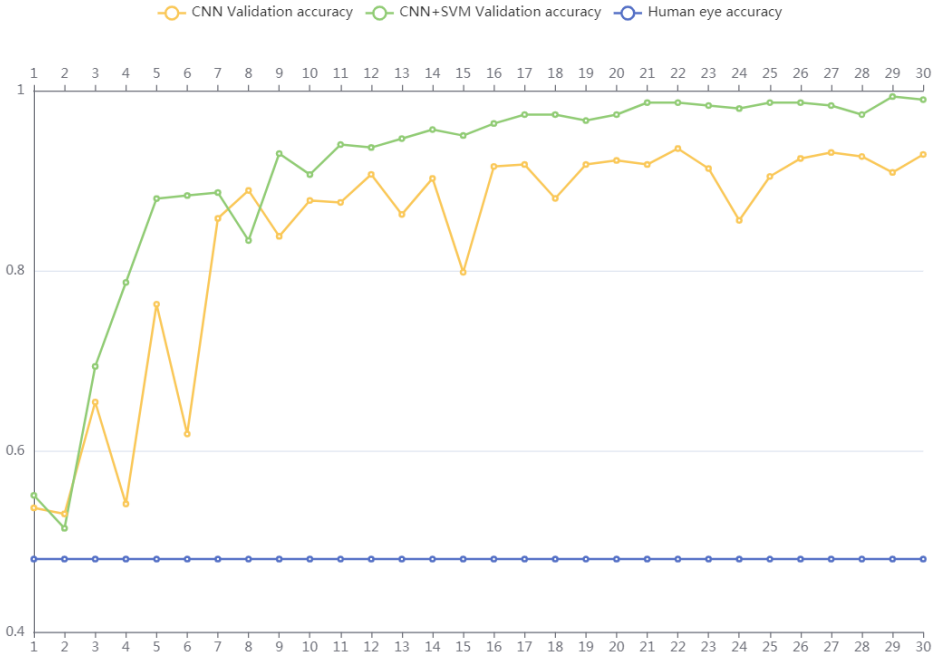

Figure 19: Accuracy chart of the three methods.

4. Conclusions

Aiming at the problem of detecting and identifying the tampering traces of certificate images, this paper proposes a method based on convolutional neural network (CNN) combined with support vector machine (SVM) based on the high requirements of the existing CNN model detection data sample set and limited detection accuracy. In the course of multiple rounds of experiments, it was found that the accuracy results of the same parameters would fluctuate. The final accuracy depends on the data set and is affected by the size of the data set and the average sample distribution. The experimental results show that under the self-built certificate image tampering data set, the CNN+SVM certificate image tampering detection algorithm can achieve 98.67% accuracy, and it has good results in this scenario. Compared with the CNN algorithm, this article The constructed CNN+SVM algorithm can enhance the tampering characteristics of the image after using the image compression, difference, Resise, and normalization of the image data. In particular, SVM can reduce the requirements of the CNN softmax classification function for the data sample set , To a certain extent, has a higher detection accuracy.

With the development of image digital technology and the further research of deep learning theory, Convolutional Neural Network (CNN) will become more and more mature, commercialized, and civilian. In future research, we will not only be limited to the current CNN+SVM model, but consider integrating multiple classifiers to explore the differences in algorithm robustness, detection speed and recognition accuracy, and further explore the Better solutions for algorithms and models in tampering scenarios.

References

[1] R. Thakur and R. Rohilla, "Recent advances in digital image manipulation detection techniques: A brief review," Forensic Science International, vol. 312, p. 110311, 2020.
[2] L. Zheng, Y. Zhang, and V. L. Thing, "A survey on image tampering and its detection in real-world photos," Journal of Visual Communication and Image Representation, vol. 58, pp. 380 – 399, 2019.
[3] K. A. da Costa, J. P. Papa, L. A. Passos, D. Colombo, J. D. Ser, K. Muhammad, and V. H. C. de

ISSN 2616-5775 Vol. 4, Issue 7: 24-38, DOI: 10.25236/AJCIS.2021.040705

Albuquerque, "A critical literature survey and prospects on tampering and anomaly detection in image data," Applied Soft Computing, vol. 97, p. 106727, 2020.

[4] T. Anbu, M. M. Joe, and G. Murugeswari, "A comprehensive survey of detecting tampered images and localization of the tampered region," MULTIMEDIA TOOLS AND APPLICATIONS, vol. 80, no. 2, pp. 2713 – 2751, 2021.

[5] Bhowmik, Deepayan, and T. Feng, "The multimedia blockchain: A distributed and tamper-proof media transaction framework," in 2017 22nd International Conference on Digital Signal Processing (DSP), 2017, pp. 1 – 5.

[6] M. Ankita, M. Sankar, and P. Shubhangi, "Digital image watermarking technique for tamper detection and restoration," in 2020 First International Conference on Power, Control and Computing Technologies (ICPC2T), 2020, pp. 56 – 61.

[7] A. Kuznetsov and V. Myasnikov, "A new copy-move forgery detection algorithm using image preprocessing procedure," Procedia Engineering, vol. 201, pp. 436 – 444, 2017.

[8] E. Ramadhani, "Photo splicing detection using error level analysis and laplacian-edge detection plugin on gimp," Journal of Physics: Conference Series, vol. 1193, p. 012013, 2019.

[9] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," IEEE Access, vol. 7, pp. 53 040 – 53 065, 2019.

[10] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," Neurocomputing, vol. 187, pp. 27 – 48, 2016.

[11] C.-C. J. Kuo, "Understanding convolutional neural networks with a mathematical model," Journal of Visual Communication and Image Representation, vol. 41, pp. 406 – 413, 2016.

[12] H. Na, J. He, and N. Zhu, "A novel method for detecting image forgery based on convolutional neural network," in IEEE International Conference on Big Data Science and Engineering;IEEE International Conference on Trust, Security and Privacy In Computing and Communications.

[13] T. Han, L. Zhang, Z. Yin, and A. C. Tan, "Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine," Measurement, vol. 177, p. 109022, 2021.

[14] HU Xiaoyi ,JING Yunjian,SONG Zhikun ,HOU Yinqing. Bearing fault identification by using deep convolution neural networks based on CNN-SVM [J]. JOURNAL OF VIBRATION AND SHOCK,2019,38(18):173-178.(in Chinese).

[15] B. Xiao, Y. Wei, X. Bi, W. Li, and J. M, "Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering," Information Sciences, vol. 511, pp. 172–191, 2020.

[16] D. C. Jeronymo, Y. C. C. Borges, and L. dos Santos Coelho, "Image forgery detection by semi-automatic wavelet soft-thresholding with error level analysis," Expert Systems with Applications, vol. 85, pp. 348 – 356, 2017.

[17] ZHOU Fei-Yan,JIN Lin-Peng,DONG Jun. Review of Convolutional Neural Network [J].CHINESE JOURNAL OF COMPUTERS,2017,40(06):1229-1251. (in Chinese)

[18] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection," in Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec 2017, Philadelphia, PA, USA, June 20-22, 2017, M. K. Matthew C. Stamm and S. Voloshynovskiy, Eds. ACM, 2017, pp. 159 – 164.

[19] C. Song, P. Zeng, Z. Wang, T. Li, L. Qiao, and L. Shen, "Image Forgery Detection Based on Motion Blur Estimated Using Convolutional Neural Network," IEEE SENSORS JOURNAL, vol. 19, no. 23, pp. 11 601 – 11 611, 2019.

[20] Z. Shi, X. Shen, H. Kang, and Y. Lv, "Image Manipulation Detection and Localization Based on the Dual-Domain Convolutional Neural Networks," IEEE ACCESS, vol. 6, pp. 76 437 – 76 453, 2018.

[21] D. Liang, C. Lu, and H. Jin, "Soft multimedia anomaly detection based on neural network and optimization driven support vector machine," Multimedia Tools and Applications, 2017.

[22] P. Kumar and A. K. Sharma, A Robust Digital Image Watermarking Technique Against Geometrical Attacks Using Support Vector Machine and Glowworm Optimization. Intelligent Data Communication Technologies and Internet of Things, 2020.