Hindawi
Mobile Information Systems
Volume 2023, Article ID 9786865, 1 page
https://doi.org/10.1155/2023/9786865

Retraction

Retracted: Adaptive Feature Analysis in Target Detection and Image Forensics Based on the Dual-Flow Layer CNN Model

Mobile Information Systems

Received 1 August 2023; Accepted 1 August 2023; Published 2 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] N. Liang, H. Xu, W. Zhang, and L. Cui, "Adaptive Feature Analysis in Target Detection and Image Forensics Based on the Dual-Flow Layer CNN Model," *Mobile Information Systems*, vol. 2022, Article ID 7140594, 14 pages, 2022.

Hindawi
Mobile Information Systems
Volume 2022, Article ID 7140594, 14 pages
https://doi.org/10.1155/2022/7140594

Research Article

Adaptive Feature Analysis in Target Detection and Image Forensics Based on the Dual-Flow Layer CNN Model

Nannan Liang,[1,2] **Haifeng Xu**,[1] **WanLi Zhang**,[1] **and Lin Cui**[1]

[1]*School of Informatics and Engineering, Suzhou University, Suzhou 234000, China*
[2]*Key Laboratory of Mine Water Resource Utilization of Anhui Higher Education Institutes, Suzhou 234000, China*

Correspondence should be addressed to Nannan Liang; lnannan@ahszu.edu.cn

Received 31 May 2022; Revised 26 July 2022; Accepted 9 August 2022; Published 28 August 2022

Academic Editor: Shadi Aljawarneh

With the rapid development of artificial intelligence technology, image editing technology has evolved from relying on software such as Photoshop and GIMP for manual modification to using artificial intelligence technology to achieve intelligent and automated tampering of images. Editing, falsifying, and disseminating digital images have become simple and easy, leading to a crisis of confidence in digital images and reducing their reliability as judicial evidence. Therefore, how to identify falsified images, improve their trustworthiness, and avoid judicial injustice has become a problem that must be overcome in the information age. In this paper, we propose a target detection and adaptive feature analysis in image forensics based on a dual-flow layer CNN model, which can effectively perform image forensics. The results show that our algorithm has a clear theoretical basis, a small operational complexity, and a high detection accuracy.

1. Introduction

People's daily life is full of image information data. In the era of image information flooding, it becomes more and more difficult to get effective and valuable image information from it quickly when we are faced with so much image information. Computer vision plays an important role in medical, security, transportation, aerospace, automation, control, and other fields. Humans get 91% of their information in cognition of the world from vision, so computer vision is the basis of cognition of the world by machines and devices. As a major vehicle for information dissemination, humans generally focus on only part of the information in an image. In daily life, only certain people or objects are the focus of attention. For example, the license plate number of a moving vehicle, the seat belt wearing of a driver, a suspicious pedestrian in a crowded place, the helmet wearing of a worker in a factory, and so on. It has become a popular research direction to find out the main targets of people's attention in images quickly and effectively and to label them with the correct classification. Therefore, the research of target detection is important for robotics, video surveillance, automated inspection, road traffic, aerospace, and other fields.

Natural scenes have great complexity, and target detection is affected by scene changes, natural lighting, weather changes, object shapes, sizes, scales, and stacking methods, making detection more difficult. Therefore, how to quickly and effectively separate targets and attenuate the effects of target scale, shape, size variation, and background complexity has become a challenge in the direction of target detection. The equipment for obtaining photos has gradually diversified. Nowadays, professional cameras, ordinary cameras, and smartphones with high-definition photo functions are common in life, among which smartphones with high-definition photo functions are far more popular than other devices in the population with their low prices, portable size, and uncompromising photo functions.

However, everything has both positive and negative aspects, and while photo images are easy to access and edit, they also bring serious negative effects to life in the information age. Some people, for various purposes, maliciously tamper with and disseminate some carefully faked photo images, reversing black and white and confusing people's

minds, causing an uproar in society and subverting the basic common sense that seeing is believing.

Digital photo image tampering forensic technology is a kind of digital image forensic technology, which relies on computer technology to judge whether the original content is still maintained in the process of transmission and dissemination of photo images taken by digital cameras. If we want to forensically examine the authenticity of photo images, we should first understand the means of photo image tampering, and the more common tampering methods include copy-paste of the same image, splicing of different images, image retouching, and image enhancement.

1.1. Copy-and-Paste Tampering with the Same Image. Paste a part of the image to other parts of the image as in Figure 1. The original photo image is on the left, and the tampered image on the right is created by copying the lawn in the figure and covering the person.

Heterogeneous image stitching tampering: a part of one image is stitched into another image by two or more images. The stitching tampering between different images has the following characteristics: (1) the tampering trace is not visually noticeable; (2) some statistical characteristics of the image are changed by the tampering behavior. As shown in Figure 2, the yellow flowers in the original image 2(a) are stitched into the original image 2(b) to obtain the tampered image 2(c).

1.2. Image Retouching. It is an image restoration operation commonly used in artistic photos to make the people in the photos more beautiful; it is also commonly used after image copy-paste tampering or splicing tampering to eliminate the edge traces of tampering. In the original image of Figure 3(a), the human face has more obvious spots, while the face of the tampered Figure 3(b) becomes smooth and more beautiful after retouching.

1.3. Image Enhancement. An operation that blurs or highlights information somewhere in an image. This type of tampering technique is usually achieved by changing the hue or contrast of a certain part of the image. Figure 4(a) blurs a large amount of detailed information by adjusting the contrast and hue, so that the original image is tampered with, creating a tampered Figure 4(b).

Incidents of photo tampering such as the one mentioned above have emerged, seriously affecting the public's correct judgment of things. In the present case, the negative impact of photo tampering and the crisis of confidence caused by it are worrying. If doctored photos are used in news reports, they may distort the facts and mislead the public, which may intensify social conflicts; if doctored photos are used as evidence in court, they may lead to false cases, obstruct justice, and allow criminals who should be punished to escape from the net of justice; if doctored photos are used in insurance claims, they may cause unnecessary economic disputes; in international relations, the use of doctored photo images may lead to political turmoil, diplomatic discord and even military conflict, and other extremely serious consequences.

2. Related Work

The traditional target detection methods are Viola-Jones, HOG + SVM, and DPM. Among them, Viola-Jones uses integral graph features and AdaBoost method. HOG + SVM method detects pedestrians as targets. It first extracts HOG features from the candidate regions of the target and then uses SVM classifier for classification decision. DPM is a variant of HOG feature detection, and DPM adds additional strategies. DPM method is the most effective and best performing method among all traditional target detection methods. Its advantages are the intuitive and simple method, block computing speed, and adaptation to animal deformation. It has been verified by a large number of scholars that its detection accuracy, generalization ability, and detection speed are better than traditional methods.

The two-stage-based target detection model refers to the extraction of features using convolutional neural network (CNN) first, then the recommendation of candidate regions using region candidates, and finally marking the target box location and classifying the marked target boxes. The most typical representative is the RCNN series of networks. The one-stage target detection model is a regression model that directly regresses the position of the target frame without generating a candidate frame in the middle of the network and directly converts the target frame location problem into a regression problem. The most representative one is the yolo series network. The large number of prior frames increases the computation and memory usage. For targets with extremely large aspect ratios in the scene, the method of preset a priori frames is not only time-consuming but also prone to false detection problems. Different data sets require different target detection models, so different a priori frames need to be set, resulting in reduced model generalization capability.

Photo image tampering forensics is emerged in the last decade. Despite the short development time, photo image tampering forensic technology has gained great progress with the continuous development and improvement of image processing, pattern recognition and artificial intelligence and other related theories, and the continuous discovery of relevant experts and scholars' research. According to the current research theoretical results, the photo image tampering forensic process is briefly summarized, as shown in Figure 5.

Since the tampered part of the tampered photo image differs from the untampered real part in certain types of features, such features can be extracted from each part of the photo image to be tested during forensics, and then the extracted features can be classified to arrive at the verdict of authenticity of the photo image to be tested. According to the different features extracted, this paper divides the photo image tampering forensics into two categories: tampering forensics based on image content features and tampering

(a) (b)

Figure 1: Copy-paste forgery within same image. (a) Original image. (b) Tampered image.

(a) (b) (c)

Figure 2: Splice forgery between different images. (a) Original Figure 1. (b) Tampered figure. (c) Original figure.

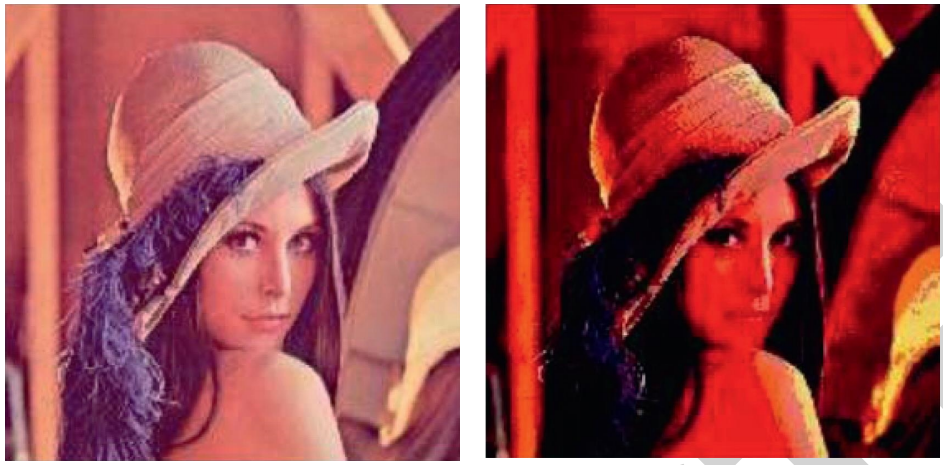
(a) (b)

Figure 3: Image blur forgery. (a) Original image. (b) Tampered image.

forensics based on imaging features, which are briefly introduced below.

2.1. Tampering Forensic Techniques Based on Content Features of Photo Images. The consistency of the content features (such as natural statistical characteristics, key points, lighting direction, and texture) of tampered photo images will be destroyed, and the researcher can make a decision on the authenticity of the image by detecting the changes of these content features.

2.1.1. Forensic Techniques Based on Natural Statistical Features. The natural statistical properties such as mean, variance, histogram, and higher order statistics of images in



(a) (b)

Figure 4: Image enhancement tamper. (a) Original image. (b) Tampered image.

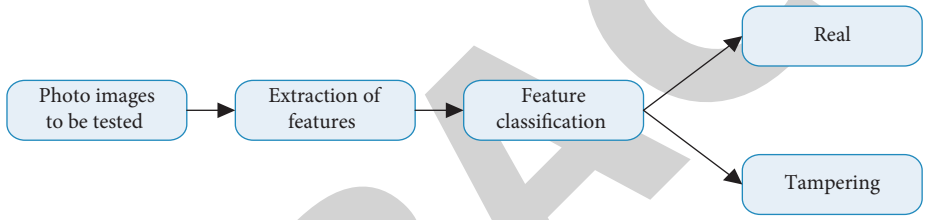


Figure 5: The process of photo image tamper forensics.

the null and transform domains are some basic features of images and an important means to study the essential properties of images.

The literature [1] proposes a copy-paste forensic algorithm based on DCT coefficients, which adopts a sliding window chunking strategy for the image to be tested, then calculates the DCT coefficients of each image block, quantifies the obtained DCT coefficients to construct feature vectors, and then performs dictionary sorting on all feature vectors. If there are similar or identical image blocks in the image to be tested, the positions of their corresponding feature vectors will be closer, and the similar blocks in the image can be identified by calculating the displacement vector to achieve the purpose of tampering detection. On this basis, the DCT quantization coefficients are dimensionalized in the literature [2]. [3] made an improvement on the chunking strategy by using a circular chunking method, followed by the construction of DCT coefficient feature vectors. The literature [4] proposed to construct feature vectors by calculating the difference matrix of the DCT coefficient matrix of an image and then detect them using SVM, and subsequently the improved method achieved an average recognition rate of 97.92% and 91.2% on the stitched image libraries of CASIAv1.0 and CASIAv2.0; however, such methods did not achieve tampering localization. The literature [5] argues that the tampered images are compressed and saved again causing changes in the DCT coefficients, whereby the histogram difference of the DCT coefficients and the double quantized mapping relationship are used to detect the stitched tampered images, respectively, to achieve the localization of the tampered regions. To further improve the forensic effect, [6] suggested extracting LBP features in the DCT domain for detection. Considering that the tampered images may be contaminated by Gaussian blur filtering and Gaussian white noise, the DCT algorithm is improved in the literature [7].

In [8], wavelet transform-based image forensic algorithms are proposed to extract features for matching detection of subband information of the wavelet transform of the image to be detected. For example, the wavelet decomposition has two subbands of low and high frequencies, and the copy-paste block is detected by comparing the correlation of the Zernike moments at the corresponding positions of the two subbands in blocks. The detection of the tampered region by comparing the similarity on the high-frequency subband after wavelet transform as proposed in literature [9]. [10] proposed to extract LBP features in the low-frequency subband to identify tampering.

2.1.2. Key Point-Based Forensic Techniques. For same-image tampering, there are two or more identical or similar regions in an image, and key points are extracted for the whole image. Since the key point characteristics of the identical or similar regions are closer, the tampered region can be located by correlation matching of all key points. Based on this principle, a detection method based on Harris point detection is proposed in the literature [11], which has better

robustness to posttampering compression. The literature [12] uses Harris points combined with the mean value of the circular neighborhood as feature points, which can solve the copy-and-paste operation of the visual structure plane region. The literature [13] extracts image feature points with Harris operator and uses a new forensic feature matching method to improve detection accuracy and efficiency.

The literature [14] proposed the SIFT feature points of the image to be tested are extracted, the feature points are matched using the G2NN matching criterion, and then the key points on the match are clustered to determine the copied and pasted regions, but the detection effect is more dependent on the clustering results. The literature [15] suggests J-linkage clustering, but the algorithm is not accurate in locating pasted blocks after rotating and scaling operations, and the detection efficiency is not ideal. To further improve the robustness of the algorithm, it is proposed to extract SIFT features using e measured after wavelet transform to reduce noise interference.

The SURF key points are proposed, and it is suggested that SURF is extracted, so the localization of the pasted blocks is not ideal; to overcome this drawback, a combination of SURF algorithm and SIFT algorithm is used to extract key points to achieve precise localization while improving the efficiency of the algorithm. The literature [16] combines both SURF and HOG features, and the experimental results are significant.

2.1.3. Forensic Techniques Based on Light Consistency Features. In photographs, there is generally a relatively fixed illumination environment (e.g., sun, interior lighting), which makes the illumination intensity and direction consistent in the photograph. For stitched blocks from other photographs, the illumination will not be consistent with the illumination of the real region in the tampered image. Farid's team proposes an image recognition model under 3D light sources, which uses a spherical harmonic model to estimate the direction of the light sources of objects in the photographs and then detects them based on the consistency of the direction. Using the detection of the consistency of the direction of the shadow region caused by the light with the direction of the light, it is robust to multiple tampered targets. Since the above method is subject to relatively strict assumptions that limit its practical applicability, it has been experimentally shown to improve the light direction estimation error and is more applicable [17].

2.1.4. Forensic Techniques Based on Texture Features. Texture is an important feature to describe and distinguish different objects. Given that it is difficult to keep the texture features of the tampered block consistent with the original image, which inevitably destroys the periodicity, directionality, and randomness of the original image texture, it provides another possible method for stitching tampering forensics. By dictionary sorting, the Tamura texture features of each image block and then calculating the feature similarity based on the Euclidean distance, the forged image regions can be detected and located. The literature [18] proposes an LBP-based texture feature description method with some robustness.

2.2. Tampering Forensic Techniques Based on Imaging Features. The general imaging model of a digital camera is shown in Figure 6. First, an optical filter to filter the color light other than red, blue, and green, after which the color information of each position is recorded by the color filter array, and then the light signal is converted into an electrical signal through the sensor, and then the CFA (Color Filter Array) interpolation algorithm is applied to each. At this time, the signal is then processed by a series of digital image processing techniques such as white balance and gamma correction, and then compressed according to certain rules to obtain the final digital photo image.

The analysis of the camera imaging model shows that in the process of digital photo image generation, after a series of hardware processing and software operations, some imaging features such as CFA interpolation noise, pattern noise, and compression noise are inevitably introduced. Due to the use of different hardware and software processing methods, it makes the photo images taken by different brands and models of cameras, only have the imaging features of that camera, so such features can be used for forensics to detect tampered and forged photo images.

2.2.1. Tampering Forensic Techniques Based on CFA Interpolation Noise. Current photo images usually use a single sensor in the generation process, and each pixel point can only record one color information, and the other two-color information are obtained by interpolating the surrounding pixels, which leads to the correlation between neighboring pixels will exist. Since different cameras may not use the interpolation method, the correlation between pixels will be different and can be used to detect tampering. The literature [19] locates splicing tampering by re-CFA interpolation of images to reconstruct their pixel neighborhood consistency, detects whether there is splicing tampering operation by analyzing the distribution of color difference images at high frequencies, extracts CFA features using Gaussian filtering based on posterior probability estimation of CFA interpolation noise, and achieves forensic purposes by classifying the features. The literature [20] considers the spectral correlation introduced by CFA interpolation and identifies the image authenticity based on this property.

2.2.2. Tampering Forensics Based on Camera Response Function. The process of generating photos of natural scenes through a series of hardware and software operations inside the camera can be called Camera Response Function (CRF). Each camera is an independent individual and its corresponding function is not the same, so the authenticity of the image can be identified by comparing the consistency of CRFs in each region of the image. The literature [21] estimated the CRF of each region by the geometric invariance of the pixels in each region of the image and then used crossover for statistical classification, achieving a

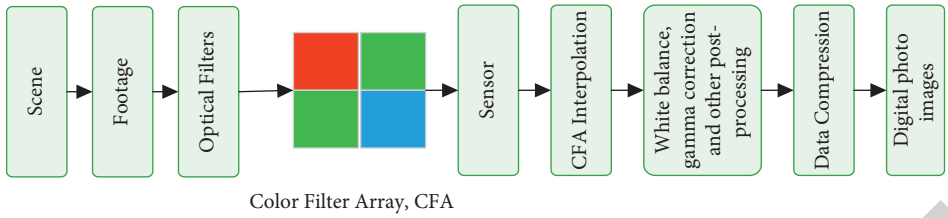


Figure 6: Digital camera imaging model.

detection rate of 87% for stitched images. Based on this, the differential invariants of the images were calculated to estimate the CRF. The literature [22] used a maximum posterior probability model to estimate the normality of the CFR to discriminate the authenticity of the photographs.

2.2.3. Tampering Forensic Techniques Based on Compression Characteristics. Photo images are usually saved in a certain compression format during the generation process, and images are usually compressed one or more times again after tampering, so the differences of individual image blocks after compression are detected to identify tampering. In the literature [23], an iterative method is proposed to estimate the original quantization table of an image to determine the approximate tampered region, and then the estimated original quantization table is used to perform another JPEG compression on the tampered region to precisely locate the tampering according to the difference in pixel values before and after compression. In the literature [24], the similarity of the synthetic images before and after compression is obtained by estimating the quantization factor to identify the location of tampering. Compressing the image again produces a double quantization effect on the DCT coefficients of the real region, whereby tampering is suggested to be identified based on the change in the DCT coefficients of different regions before and after compression. The literature [25] argues that images produce uniform quantization noise after JPEG compression, and tampered blocks corrupt this property and propose a quantization noise estimation model to detect the differences between image blocks. Considering that JPEG compression produces a grid effect, the presence of unaligned grids in the image can be detected to identify the tampered locations.

2.2.4. Pattern Noise-Based Tampering Forensic Technique. Pattern noise is caused by the imperfection of the camera sensor and the inconsistency of the materials used, resulting in the imperfect conversion of light signals into electrical signals, and is stable in every picture taken by the camera. Since the sensor of each camera is unique, its mode noise is also unique; in addition, each pixel point on the sensor is different, resulting in inconsistent performance of mode noise in each pixel point. Based on these two characteristics, the pattern noise can be regarded as a camera fingerprint and applied to photo image tampering forensics, which can be generalized to a variety of tampering operations such as copy-paste of the same image and stitching of different images.

In addition to the above features, Markov features, Fourier-Mellin transform features, image quality features, and color features are often used for photo image tampering detection。

3. Methods

Figure 7 presents a generalized framework for digital image source forensics under the CNN model theory. In the image preprocessing, the image to be detected is first cut into image blocks (P_k in Figure 8(a) indicates the kth image block), and then the image fingerprint characterizing the source of the shot is extracted using CNN, and the detection result Y_k of each image block is output (Y_k in Figure 8(c) indicates the feature extractor predicts the label for the k th image block), and the majority voting algorithm is used to fuse the detection results of the kth image block and output the image-level prediction results, i.e., device model multiclassification identification.

It was found that the FPN feature fusion algorithm improved the detection of small targets but did not improve the detection of large targets, and there was information redundancy after feature fusion. Since then, researchers have proposed some variants of FPN, such as PANet, Libra-RCNN, which are built on the assumption that the weights are the same when the features of two layers are fused, ignoring the feature that the contribution values of features in different layers are different. Therefore, this section proposes a new feature pyramid named dual-stream feature CNN (DS-CNN) using autonomous learning weights and jump connection method.

As shown in Figure 8, FPN is a simple top-down one-way information, while PANet adds bottom-up information flow to FPN to enhance higher-level semantic information for semantic segmentation, which is better than FPN but more computationally intensive. LibraRCNN collects feature information from each layer and then refines the output to the feature layer, with the idea of fusion before segmentation. NASFPN adopts the idea of AutoML and uses search for feature fusion; ASFF learns the weight contribution value of each layer, but it is a fully connected method with high computation and requires higher performance computing equipment, which is not convenient for practical applications.

As shown in Figure 9, BiFPN is a feature fusion algorithm proposed in the Efficient Det network, in which a jump connection approach and a weighted fusion approach are used for feature fusion, taking into account both efficiency and accuracy. Its calculation equations are shown in (1) and (2).

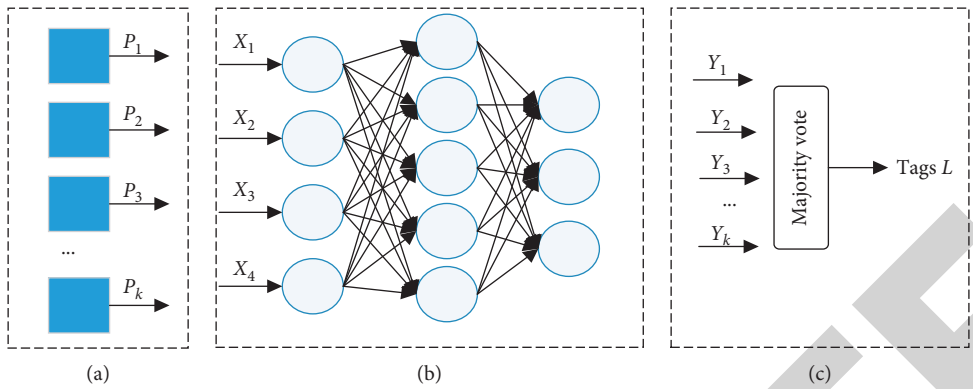


Figure 7: Digital image source framework based on CNN; (a) image preprocessing; (b) image feature extraction; (c) classification result voting.

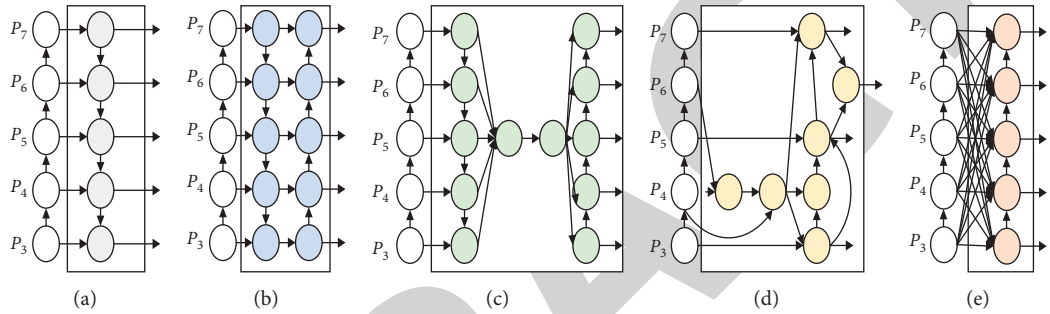


Figure 8: Feature network design diagram; (a) FPN; (b) PANet; (c) LibraRCNN; (d) NAS-FPN; (e) ASFF.

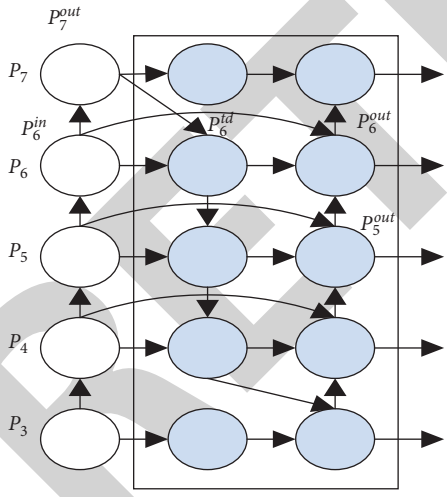


Figure 9: BiFPN basic structure.

$$P_6^{\text{td}} = \text{conv}\left(\frac{w_1 * P_6^{\text{in}} + w_2 * \text{Resize}\left(P_7^{\text{out}}\right)}{w_1 + w_2 + \varepsilon}\right), \quad (1)$$

$$P_6^{\text{out}} = conv\left(\frac{w_1' * P_6^{\text{in}} + w_2' * P_6^{\text{td}} + w_3' \text{Resize}\left(P_5^{\text{out}}\right)}{w_1' + w_2' + w_3' + \varepsilon}\right). \quad (2)$$

Resize denotes up sampling, and nearest neighbor interpolation is used here, and ε is set to 0.0001 to prevent division by zero.

Each feature layer in BiFPN has different weights, which are theoretically normalized to 1 when fused to the same layer, but how to normalize the weights?

(1) Unbounded strategy.

$$O = \sum_i w_i \cdot I_i. \quad (3)$$

(2) Based on softmax.

$$O = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} \cdot I_i. \quad (4)$$

(3) Fast regularity

$$O = \sum_i \frac{w_i}{\varepsilon + \sum_j w_j} \cdot I_i. \quad (5)$$

It is found that the BiFPN algorithm with fast regularization can eliminate the exponential operation in the softmax method, reduce the computational overhead, and speed up the computation although the accuracy is slightly lower than the softmax method, but the performance is the best among the three strategies.

We propose a new dual-stream feature CNN (DS-CNN), for the FCOS structure of fused area-ness. It is mainly improved based on the BiFPN algorithm, following its jump connection and weighted fusion, while improving the FCOS model structure based on the fused area-ness in Chapter 3. The general structure is shown in Figure 10.

As shown in Figure 10, the network is obtained from ResNeXt101 to obtain C_3, C_4, C_5 layers, and then it is convolved by $1 * 1$ for dimensionality reduction to obtain a 256-dimensional P_3, P_4, P_5 feature map, which is convenient for feature fusion. P_6, P_7 is the feature map obtained after down sampling P_5, P_6 separately.

The dual-stream feature CNN (DS-CNN) can enhance the semantic information of each prediction layer though. However, the analysis of information inflow from each node shows that the information inflow and outflow at P_5 is unbalanced. From Figure 10, it can be seen that layer P_5 has only one information input in layer C_5, but three information outflows (the arrows can indicate the information inflow and outflow of P_5). Whether sufficient information can be obtained has an important impact on the subsequent feature fusion of node P_6, P_7. Therefore, in this thesis, the information of node P_5 is enhanced based on the dual-stream pyramid, and the information of layer C_3, C_4 is also fused directly to layer P_5. However, layer C_3, C_4, C_5 has different contribution values to P_5 feature layers, so it is necessary to learn different weights for layer C_3, C_4, C_5 first before feature fusion. In this paper, we name this feature fusion method as Adaptive Dual Streaming Feature CNN (A-DS-CNN), and the specific structure is shown in Figure 11.

As in Figure 11, layer C_3, C_4 will increase the information inflow in layer P_7 by means of adaptive feature fusion. The adaptive weights are calculated as in -) (6):

$$y_{ij} = conv\left(\alpha_{ij}^3 * x_{ij}^3 + \beta_{ij}^4 * x_{ij}^4 + \gamma_{ij}^5 * x_{ij}^5\right), \tag{6}$$

$$\begin{aligned} \alpha_{ij}^3 &= \frac{e^{\lambda_{\alpha_{ij}}^3}}{e^{\lambda_{\alpha_{ij}}^3} + e^{\lambda_{\beta_{ij}}^4} + e^{\lambda_{\gamma_{ij}}^5}}, \\ \beta_{ij}^4 &= \frac{e^{\lambda_{\beta_{ij}}^4}}{e^{\lambda_{\alpha_{ij}}^3} + e^{\lambda_{\beta_{ij}}^4} + e^{\lambda_{\gamma_{ij}}^5}}, \\ \gamma_{ij}^5 &= \frac{e^{\lambda_{\gamma_{ij}}^5}}{e^{\lambda_{\alpha_{ij}}^3} + e^{\lambda_{\beta_{ij}}^4} + e^{\lambda_{\gamma_{ij}}^5}}, \end{aligned} \tag{7}$$

where $\alpha_{ij}^3, \beta_{ij}^4, \gamma_{ij}^5$ denotes the weight of C_3, C_4, C_5, respectively, i, j denotes the location coordinates of the feature map, $x_{ij}^3, x_{ij}^4, x_{ij}^5$ denotes the value at location (i, j) in layer C_3, C_4, C_5, and y_{ij} denotes the final output to the value at location (i, j) in P_5.

The network structure of the final target detection algorithm in this thesis is shown in Figure 12.

To improve the basic convolutional network, the Backbone layer adopts the ResNeXt101 structure and uses 64 paths with each path width of 4 to reduce the computational effort. For better feature extraction, the Channel of C_3, C_4, C_5 layers is first converted to 256 dimensions and then input to A-DFN for feature fusion. The adaptive feature fusion is performed to P_5 for layer C_3, C_4, C_5, where the weights are normalized by Soft max, in order to provide more information for the subsequent feature fusion without sacrificing accuracy. And the subsequent weight normalization used for weighted feature fusion from layer P_3, P_4, P_5, P_6, P_7 is fast regular, which aims to make the network run better by slightly sacrificing accuracy while ensuring performance. Head layer is divided into a shared part and a branch part, and since area-ness is more closely related to location, it is divided into the same branch as regression to help the model perform better target box position regression.

4. Experiments

The COCO dataset, known as Microsoft Common Objects in Context, is a dataset acquired by the Microsoft team to perform target recognition, target segmentation, and target detection competitions. The schematic diagram of the COCO dataset is shown in Figure 13. COCO dataset is divided into 2014 version and 2017 version. The current version used in this thesis is the 2017 version, which contains 80 target categories, 118287 training sets, totaling 19.3 G, 5000 validation sets, totaling 1814.7 M, so the 2017 version COCO dataset has 123287 sheets.

From Figure 14, it can be seen that the COCO dataset has more categories than the PASCALVOC dataset, and the number of instances corresponding to each category is also higher. Therefore, the COCO dataset is more difficult to detect the target and can better represent the performance of the target detection model.

As shown in Figure 15, the area of most targets in the COCO dataset is only about 6% of the image size; 41.43% of all targets appearing in the COCO training dataset are small targets, 34.4% are medium targets, and 24.2% are large targets. By analyzing the COCO data, it is found that small and medium targets account for a larger proportion, so this dataset is more concerned with the detection of small and medium targets.

We mainly detect the coco dataset and divide it. During training, the image data are first preprocessed, resize the image to match the target detection model size, and then input to the target detection model, train the appropriate number of iterations, and get the final detection results. Finally, the test-test-dev data results are submitted to the coco Detection Challenge competition to obtain AP values and AR values.

In order to better compare the anchor-base and anchor-free methods, this thesis is based on a unified benchmark, i.e., the COCO dataset and uses a unified evaluation criterion: MAP values (MAP values are equivalent to AP values in coco data), and compares their MAP values for large, medium, small targets and different IOU thresholds. The MAP values are compared for large, medium, small targets and different IOU thresholds.

According to Table 1, CenterNet511 and CornerNet511 take longer to test one image under the same conditions,

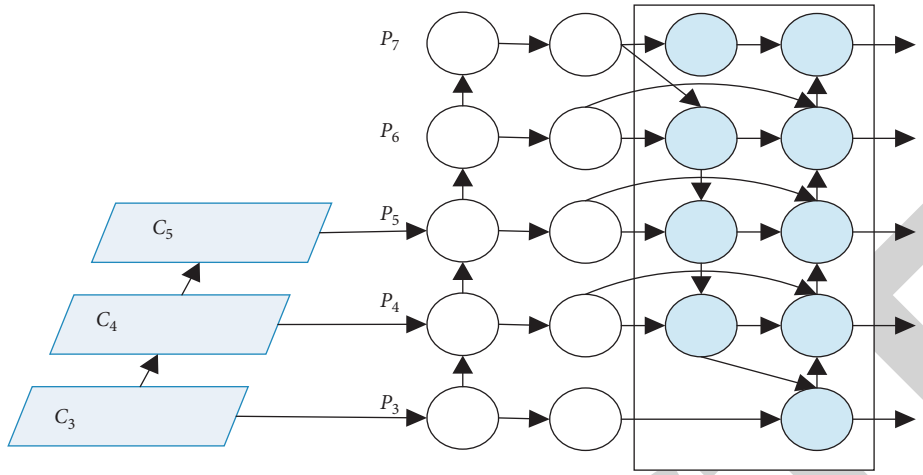


Figure 10: Dual-stream feature CNN.

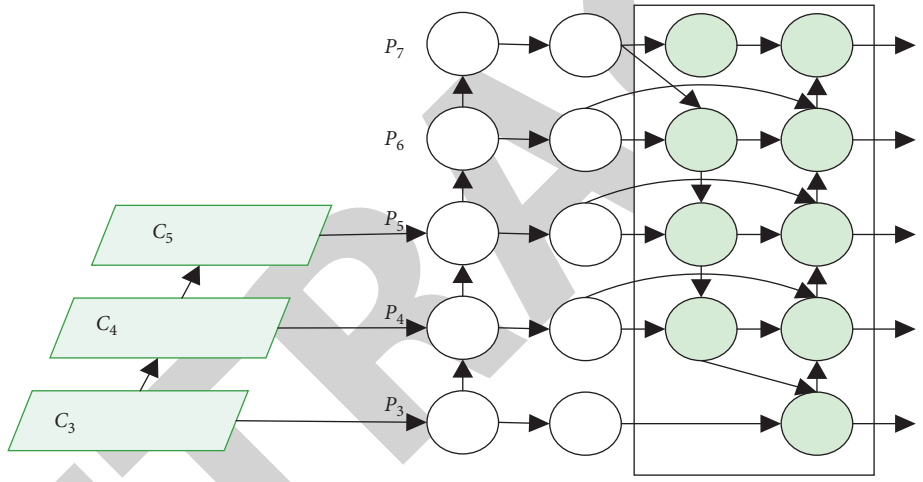


Figure 11: Adaptive dual-stream feature CNN.

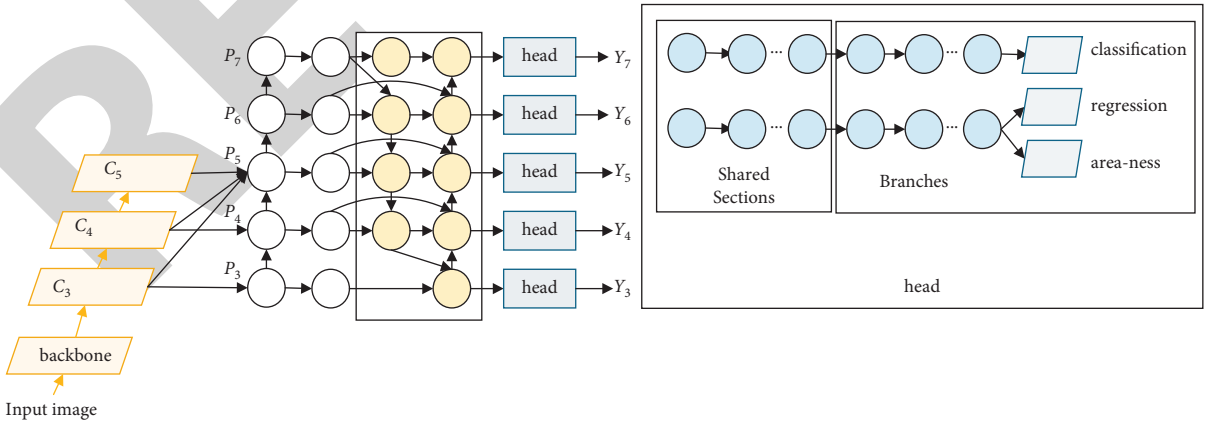


Figure 12: Schematic diagram of the overall network structure.

indicating that CenterNet511 predicts slower because CenterNet511 predicts one more centroid than CornerNet511, which brings more computation. Both CenterNet511 and CornerNet511 use the hourglass network as the backbone layer, which is computationally intensive and has a slow computing speed. In contrast, the ResNeXt-

Figure 13: Schematic diagram of coco target detection dataset.

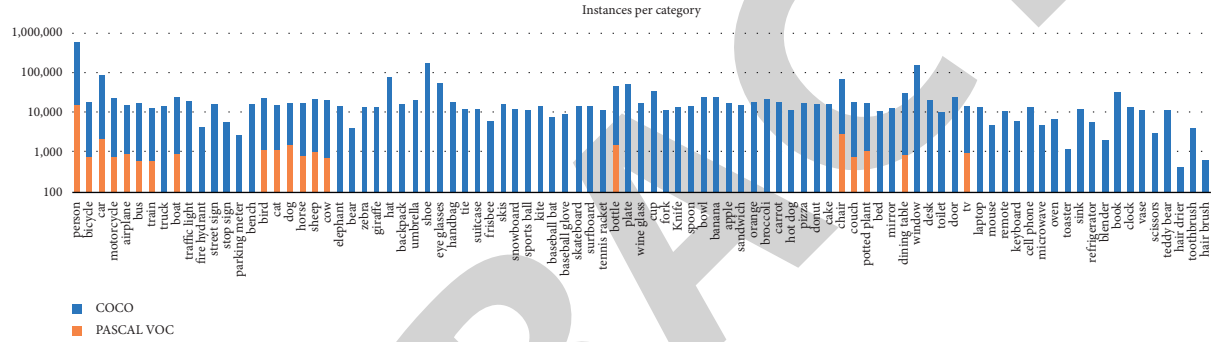


Figure 14: Comparison of the number of instances between the COCO dataset and the PASCAL VOC dataset.

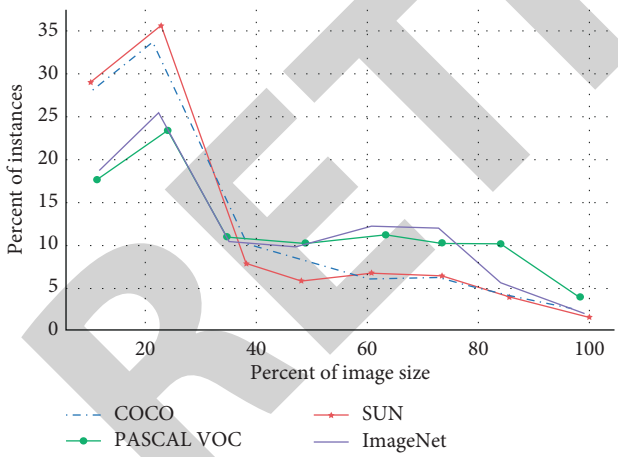


Figure 15: Example image size and percentage of the COCO, PASCAL VOC, ImageNet, and SUN datasets.

Table 1: Comparison of FCOS and CenterNet511, CornerNet511 prediction speed.

Method	Backbone	Testing time/image
CenterNet511	Hourglass52	270 ms
CenterNet511	Hourglass104	300 ms
CenterNet511	Hourglass104	340 ms
FCOS	ResNeXt-101	112 ms

(Note: 511 means the input image size is 511 ∗ 511).

Table 2: FCOS prediction accuracy values using different backbone.

Method	Backbone	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
FCOS w/ FPN	ResNet-101	41.6	60.6	45.1	24.3	44.9	51.5
FCOS w/ FPN	ResNeXt-32x8d-101	42.6	62.3	46.2	26.1	45.5	52.5
FCOS w/ FPN	ResNeXt-64x4d-101	44.8	64.1	48.5	27.5	47.4	55.7

(Note: 32 ∗ 8d means 32 paths, the width of each path is 8).

101 prediction used by FCOS is fast and has a better performance in terms of average accuracy mean, so this thesis uses the FCOS algorithm as the base algorithm for improvement.

According to Table 2, the MAP value can reach 44.8 when FCOS adopts ResNeXt-64x4d-101-FPN (i.e., 64 paths, each with a width of 4) as the backbone, which is 3 and 2 points higher than that of ResNet-101 and ResNeXt-32x8d-101, respectively. Therefore, ResNeXt-64x4d-101 is used as the backbone of the target detection model in this thesis.

According to Table 3, the mean accuracy of target detection using center-ness is 0.2 points lower than the area-ness designed in this paper when the feature fusion method is FPN. In the DS-CNN fusion method, the area-ness is 0.4 points higher than the MAP value of center ness. It means

Table 3: MAP values of area-ness and center-ness on FCOS algorithm.

Method	Centrality method	Feature fusion method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
FCOS	Center-ness	FPN	44.8	64.0	48.5	27.5	47.6	55.7
FCOS	Area-ness	FPN	44.8	64.4	48.6	27.4	47.5	55.8
FCOS	Center-ness	DS-CNN	46.2	65.7	49.9	26.8	49.7	58.4
FCOS	Area-ness	DS-CNN	46.6	66.1	50.3	27.3	49.5	59.2

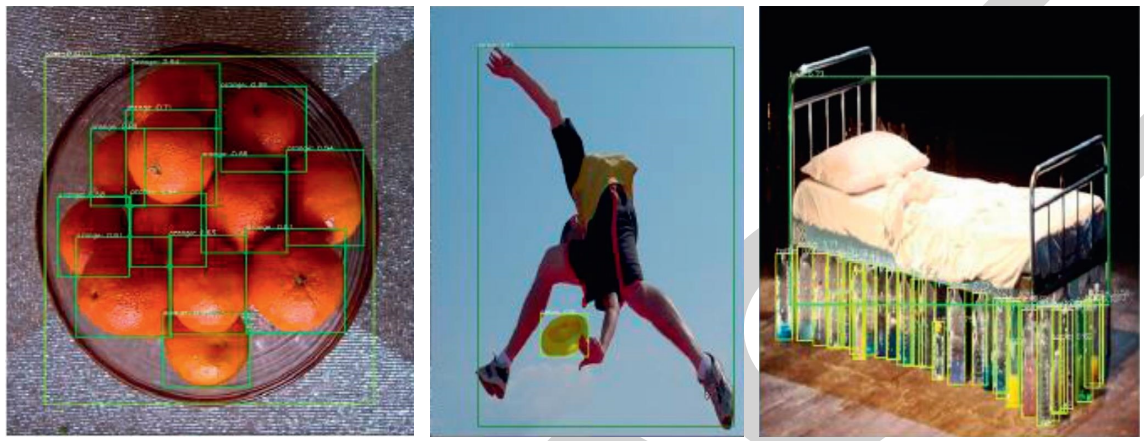


Figure 16: Visualization results of partial object detection by the algorithm proposed in this thesis.

Table 4: Comparison of detection results.

Algorithm		TPR/%	FAR/%	Average time to detect an image
Traditional pixel-by-pixel sliding window fixed threshold algorithm	$\tau = 0.006$	88.5	1.903	727.5 s
	$\tau = 0.001$	95.7	4.942	
	$\tau = 0.014$	97.7	9.029	
	$\tau = 0.03$	98.9	15.37	
The proposed algorithm		98.9	1.896	26.75 s

that the area-ness designed in this paper is better than the center-ness of the original FCOS.

As shown in Figure 16, the algorithm of this thesis is able to accurately detect even a compact orange placed in the fruit tray or an empty water bottle placed by the bed, indicating that the prediction layer is able to acquire a sufficient number of features. When the target frames of people and Frisbees overlap, the algorithm of this thesis is still able to predict them each.

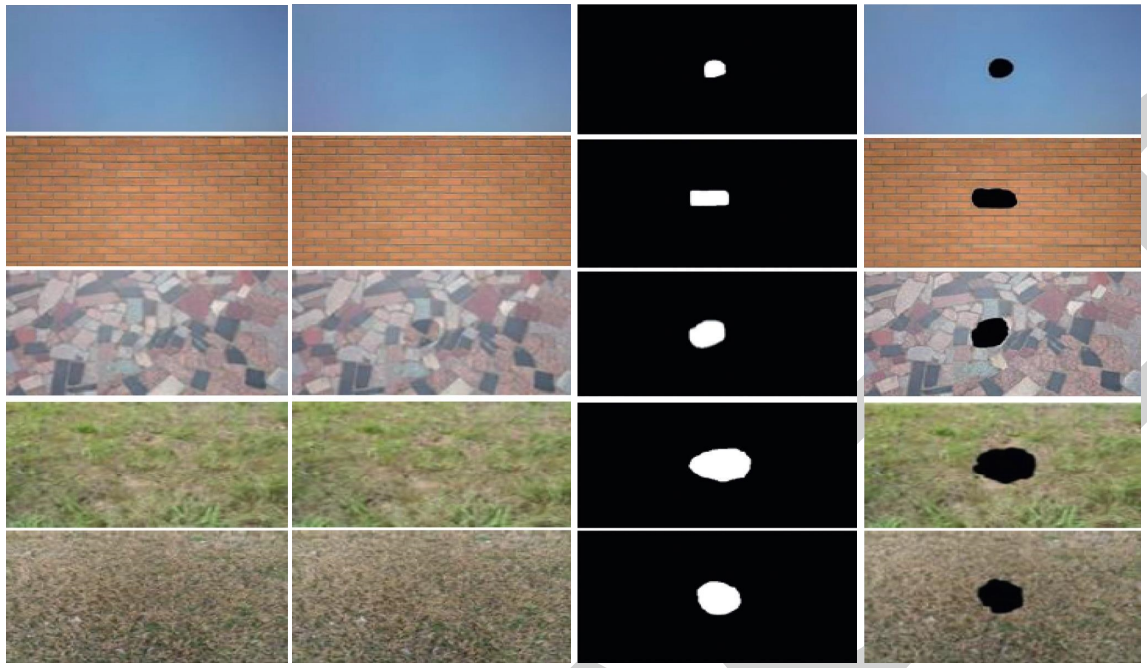
The authenticity detection results for each test image can be divided into two categories: tampered and true. To evaluate the performance,

$$\begin{aligned} \mathrm{TPR} &= \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}, \\ \mathrm{FAR} &= \frac{\mathrm{FN}}{\mathrm{FN} + \mathrm{TN}}. \end{aligned} \tag{8}$$

Authenticity detection results: tampering detection experiments are performed on 500 real images and 500 tampered images from the image library given in Table 4 using the traditional fixed-threshold sliding window method based on correlation coefficients and the proposed SPCE-based adaptive threshold nonoverlapping chunk matching + ZNCC algorithm, respectively.

The methods based on fixed thresholds will have different detection results at different thresholds, and four more desirable thresholds of 0.006, 0.01, 0.014, and 0.03 were selected for comparison through experiments. In order to be able to evaluate the detection results objectively, the pattern noise in both types of algorithms is obtained by wavelet noise reduction and then processed with ZM + WF. In the calculation of TPR and FAR, if the number of pixels of a certain image tampering localization result is less than 20, it is judged to be a real image and vice versa; it is considered as tampering. The detection results of the two algorithms are shown in Table 4.

The proposed adaptive thresholding algorithm has a TPR of 98.9% and FAR of 1.896% for 1000 images to be tested, while the fixed thresholding algorithm has different detection results at different threshold values. It is 0.01, 0.014, and 0.03, and although the TPR is similar or equal to it, the FAR is much higher than this paper. At 0.006, the FAR is similar to the algorithm in this paper, but the TPR is much lower than that in this paper. Meanwhile, the average detection time of the two algorithms on 1000 images is given in Table 4, and the comparison results show that the proposed algorithm effectively reduces false alarms while maintaining a high detection rate and detection efficiency.



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

Figure 17: The location results of tampered images with different texture. (a) Original image. (b) Tampered image. (c) Tampered position. (d) Positioning effect.

Table 5: The detection accuracy of Seam-carving tampering images.

P-value ratio selection	Tampering				
	Untampered image (%)	0.9 (%)	0.8 (%)	0.7 (%)	0.6 (%)
G.R. sheng	67.86	64.27	85.72	83.92	87.50
0.9	67.84	78.58	89.28	85.72	87.50
0.8	69.65	76.77	87.51	89.28	91.08
0.7	71.43	78.56	94.63	85.72	89.28

(Note: G.R. Sheng is the detection result based on the extended Markov feature, the rest are the detection results when different thresholds are selected by the adaptive detection feature extraction method, the results in the table are all correct results).

Table 6: The comparison between the proposed method and G.R. Sheng's method.

P-value ratio selection	Tampering				
	Untampered image (%)	0.9 (%)	0.8 (%)	0.7 (%)	0.6 (%)
G.R. sheng	67.86	64.27	85.72	83.92	87.49
Algorithm of this paper	69.25	77.98	90.488	86.89	89.31

(Note: The detection result of this algorithm is the average correct rate using three thresholds of 0.7, 0.8 and 0.9).

Compared with the traditional fixed threshold judgment method, the adaptive threshold judgment method, which is based on the texture complexity of the image block to be tested, selects a suitable threshold value, thus realizing "specific problem specific analysis."

In Figure 17, the first to fourth columns show the original image, the tampered image, and the tampered location, respectively. The second column gives five tampered images with simple to complex texture complexity, where the texture complexity of the first local block of blue sky image is $k \in [0.1857, 0.2886]$; the texture complexity of the second local block of wall image is $k \in [0.3288, 0.4372]$; the texture complexity of the third local block of floor image is $k \in [0.3511, 0.5296]$; the texture complexity of the fourth local block of green grass image is $k \in [0.6601, 0.8442]$; and the texture complexity of the fifth local block of dead grass image is $k \in [0.6927, 0.9463]$.

Observing the localization results of the proposed adaptive thresholding algorithm for five tampered inspection images shows that whether the texture of the tampered image is simple or complex, which effectively eliminates the influence of texture on forensics (see Tables 5, 6).

5. Conclusion

In the information age, photography is one of the most important means to ensure public access to information, but the continuous tampering of photos forces people to reexamine. As Professor Hani Farid of Dartmouth said, "we live in a place where we no longer believe what we hear. As an important information carrier, digital photos should have broad application space. Justice and safeguarding justice are indisputable. This paper proposes an adaptive characteristic analysis method based on two-layer CNN model, which effectively solves this problem and is of great significance to the research in this field.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Acknowledgments

This work was supported in part by the Key Laboratory of Mine Water Resource Utilization of Anhui Higher Education Institutes, Suzhou University (Grant no. KMWRU202107), in part by the Outstanding Youth Talents in Anhui Provincial Education Department (Grant no.2019gxbjZD43), in part by the Open Research Fund of National Engineering Research Center for Agro-Ecological Big Data Analysis and Application, Anhui University (Grant no.AE202201) , in part by the Open Foundation of the Anhui Key Laboratory of Intelligent Building and Building Energy Conservation (Grant no.IBES2020KF03), in part by the Key Research and Development Project of Anhui Province in China (Grant no. 202004b11020023), in part by the Academic support project for top-notch talents in disciplines (majors) of Colleges and Universities at Anhui Province in China (Grant no. gxbjZD21081), and in part by the school-level key disciplines of computer science and technology at Suzhou University in China (Grant no. 2019xjzdxk1) and in part by the Collaborative Innovation Center—cloud. computing industry (Grant no. 4199106).

References

[1] A. M. Al-Azab, A. A. Zaituon, K. M. Al-Ghamdi, and F. M. A. Al-Galil, "Surveillance of dengue fever vector *Aedes aegypti* in different areas in Jeddah city Saudi Arabia," *Advances in Animal and Veterinary Sciences*, vol. 10, no. 2, pp. 348–353, 2022.

[2] L. Cai, Q. Sun, T. Xu, Y. Ma, and Z. Chen, "Multi-AUV collaborative target recognition based on transfer-reinforcement learning," *IEEE Access*, vol. 8, pp. 39273–39284, 2020.

[3] R. Tong, Y. Zhang, H. Chen, and H. Liu, "Learn the temporal-spatial feature of sEMG via dual-flow network," *International Journal of Humanoid Robotics*, vol. 16, no. 04, Article ID 1941004, 2019.

[4] A. R. Alqahtani, A. Badry, S. A. Amer, F. M. A. Al Galil, M. A. Ahmed, and Z. S Amr, "Intraspecific molecular variation among Androctonus crassicauda (Olivier, 1807) populations collected from different regions in Saudi Arabia," *Journal of King Saud University Science*, vol. 34, no. 4, Article ID 101998, 2022.

[5] Di Wu, Y. Lei, M. He, C. Zhang, and Li Ji, "Deep reinforcement learning-based path control and optimization for unmanned ships," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–8, Article ID 7135043, 2022.

[6] R. Ali, M. H. Siddiqi, and S. Lee, "Rough set-based approaches for discretization: a compact review," *Artificial Intelligence Review*, vol. 44, no. 2, pp. 235–263, 2015.

[7] M. Afrasiabi, H. Khotanlou, and T. Gevers, "Spatial-temporal dual-actor CNN for human interaction prediction in video," *Multimedia Tools and Applications*, vol. 79, no. 27-28, pp. 20019–20038, 2020.

[8] G. Cai, Y. Fang, J. Wen, S. Mumtaz, Y. Song, and V. Frascolla, "Multi-carrier M-ary DCSK system with code index modulation: an efficient solution for chaotic communications," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 6, pp. 1375–1386, Oct, 2019.

[9] I. Castillo Camacho and K. Wang, "A comprehensive review of deep-learning-based methods for image forensics," *Journal of Imaging*, vol. 7, no. 4, p. 69, 2021.

[10] K. Chandra, A. S. Marcano, S. Mumtaz, R. V. Prasad, and H. L. Christiansen, "Unveiling capacity gains in ultradense networks: using mm-wave NOMA," *IEEE Vehicular Technology Magazine*, vol. 13, no. 2, pp. 75–83, June 2018.

[11] X. Liao, K. Li, X. Zhu, and K. J. R. Liu, "Robust detection of image operator chain with two-stream convolutional neural network," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 955–968, 2020.

[12] F. B. Saghezchi, A. Radwan, J. Rodriguez, and T. Dagiuklas, "Coalition formation game toward green mobile terminals in heterogeneous wireless networks," *IEEE Wireless Communications*, vol. 20, no. 5, pp. 85–91, 2013.

[13] B. Chen, W. Tan, G. Coatrieux, Y. Zheng, and Y. Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment," *IEEE Transactions on Multimedia*, vol. 23, pp. 3506–3517, 2021.

[14] S. Palanisamy, B. Thangaraju, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and F. Alassery, "A novel approach of design and analysis of a hexagonal fractal antenna array (HFAA) for next-generation wireless communication," *Energies*, vol. 14, no. 19, p. 6204, 2021.

[15] B. Bayar and M. C. Stamm, "Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2691–2706, 2018.

[16] O. Mayer and M. C. Stamm, "Forensic similarity for digital images," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1331–1346, 2020.

[17] A. Abd, A. Fahd Mohammed, and S. P. Zambare, "New species of flesh fly (Diptera: sarcophagidae) Sarcophaga (Liosarcophaga) geetai in India," *J Entomol Zool Stud*, vol. 4, no. 3, pp. 314–318, 2016.

[18] S. Nagi Alsubari, S. N Deshmukh, A. Abdullah Alqarni et al., "Data analytics for the identification of fake reviews using supervised learning," *Computers, Materials & Continua*, vol. 70, no. 2, pp. 3189–3204, 2022.

[19] D. Bhardwaj and V. Pankajakshan, "A JPEG blocking artifact detector for image forensics," *Signal Processing: Image Communication*, vol. 68, pp. 155–161, 2018.
[20] W. Zhang, Q. Li, Q. M. J. Wu, Y. Yang, and M. Li, "A novel ship target detection algorithm based on error self-adjustment extreme learning machine and cascade classifier," *Cognitive Computation*, vol. 11, no. 1, pp. 110–124, 2019.
[21] Q. Liu, C. Liu, and Y. Wang, "etc. Integrating external dictionary knowledge in conference scenarios the field of personalized machine translation method," *Journal of Chinese Informatics*, vol. 33, no. 10, pp. 31–37, 2019.
[22] S. Walia and K. Kumar, "Digital image forgery detection: a systematic scrutiny," *Australian Journal of Forensic Sciences*, vol. 51, no. 5, pp. 488–526, 2019.
[23] S. A. Bansode, V. R. More, S. P. Zambare, and M. Fahd, "Effect of constant temperature (20 0C, 25 0C, 30 0C, 35 0C, 40 0C) on the development of the Calliphorid fly of forensic importance, *Chrysomya megacephala* (Fabricus, 1794)," *Journal of Entomology and Zoology Studies*, vol. 4, no. 3, pp. 193–197, 2016.
[24] G. Boato, D. T. Dang-Nguyen, and F. G. B. De Natale, "Morphological filter detector for image forensics applications," *IEEE Access*, vol. 8, pp. 13549–13560, 2020.
[25] F. A. Al-Mekhlafi, R. A. Alajmi, Z. Almusawi et al., "A study of insect succession of forensic importance: Dipteran flies (diptera) in two different habitats of small rodents in Riyadh City, Saudi Arabia," *Journal of King Saud University Science*, vol. 32, no. 7, pp. 3111–3118, 2020.