Neural Computing and Applications
https://doi.org/10.1007/s00521-021-06329-4

S.I. : MACHINE LEARNING APPLICATIONS FOR SECURITY

DCU-Net: a dual-channel U-shaped network for image splicing forgery detection

Hongwei Ding[1,2] · Leiyang Chen[1] · Qi Tao[1] · Zhongwang Fu[1] · Liang Dong[1] · Xiaohui Cui[1,2]

Received: 20 January 2021 / Accepted: 10 July 2021

Abstract
The detection and location of image splicing forgery are a challenging task in the field of image forensics. It is to study whether an image contains a suspicious tampered area pasted from another image. In this paper, we propose a new image tamper location method based on dual-channel U-Net, that is, DCU-Net. The detection framework based on DCU-Net is mainly divided into three parts: encoder, feature fusion, and decoder. Firstly, high-pass filters are used to extract the residual of the tampered image and generate the residual image, which contains the edge information of the tampered area. Secondly, a dual-channel encoding network model is constructed. The input of the model is the original tampered image and the tampered residual image. Then, the deep features extracted from the dual-channel encoding network are fused for the first time, and then the tampered features with different granularity are extracted by dilation convolution, and then, the secondary fusion is carried out. Finally, the fused feature map is input into the decoder, and the predicted image is decoded layer by layer. The experimental results on Casia2.0 and Columbia datasets show that DCU-Net performs better than the latest algorithm and can accurately locate tampered areas. In addition, the attack experiments show that DCU-Net model has good robustness and can resist noise and JPEG recompression attacks.

Keywords Image splicing · Image tampering detection · Image forensics · DCU-Net

1 Introduction

Image has always been the most effective information carrier in people's life, and it has penetrated into all aspects of our life. With the development of image editing technology, people can easily modify the content of the image, resulting in the tampered image transmission completely different information. Due to the lack of digital media supervision, malicious image tampering has caused serious adverse consequences in military, political, life, and academic fields [1]. Therefore, there is an urgent need for a detection method of image tampering and forgery.

The common image forgery can be divided into three categories: copy–move [2–4], cut–paste [5, 6], and remove [7–9]. Copy–move refers to copy part of the content of the image and paste it into the same image; cut–paste usually cuts part of the content from one image and pastes it into another image; and remove means to delete part of the image and fill it with background pixels. In order to reduce the public trust crisis caused by image tampering and forgery, many scholars have carried out relevant research [8, 10, 11]. These methods greatly promote the development of digital image tamper detection technology. In this paper, we focus on image splicing forgery detection. On the one hand, image splicing is the most common type of tampering in daily life, which is closely related to the public life. On the other hand, because the tampered area is cut from other images, only by recognizing the difference between the tampered area and the unmodified area can we detect effectively. However, the tampered image is usually post-processed to eliminate the difference after tampering. This makes image splicing forgery detection more challenging than other tamper detection methods. Based on this, the research of image splicing forgery detection has more important significance. As shown in Fig. 1, an

✉ Xiaohui Cui
xcui@whu.edu.cn

[1] School of Cyber Science and Engineering, Wuhan University, Wuhan, China

[2] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, Wuhan, China

Published online: 12 August 2021

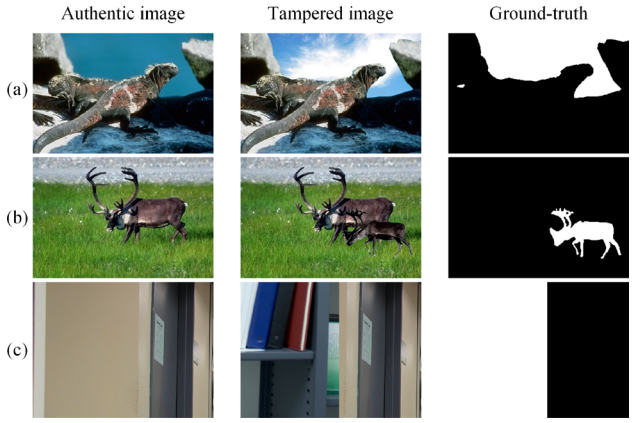

Fig. 1 Splicing tamper forensics sample image. Among them, **a** and **b** are from Casia2.0 dataset and **c** are from Columbia dataset. **a** Modified the sky background, **b** added animals to the image, **c** modified the object on the left

example of image splicing forgery is given, including authentic image, splicing image and ground truth, where white area in ground-truth image is tampered area. As can be seen from the example, it is difficult for the human eye to distinguish the tampered area even with detailed inspection.

According to the difference of the attributes between the tampered region and the authentic region, the traditional image splicing forgery detection can be roughly divided into four types: based on the difference of imaging equipment [10, 12–14], based on the difference of image essential attributes [15–20], based on the difference of image compression attributes [21–24], and based on the perceptual image hash [25, 26]. Image tamper detection based on the difference of imaging equipment means that different imaging equipment will leave specific fingerprint information in the formed image, through which the tampering and authentic image can be determined. Yao et al. [12] proposed an image splicing detection method for image blocks from different sources by exploring the relationship between noise level function and camera response function according to different noise features in different source regions. Nan et al. [14] designed a noise level function based on the influence of texture and edge of image on the local estimation of noise variance, and regarded the image block not constrained by the noise level function as the splicing region. Tamper detection based on the difference of the essential attributes of the image is to distinguish the tampered image from the authentic image by discovering the change of the inherent attribute in the image. Han et al. [18] employed an effective Markov feature selection algorithm and applied it to image forgery detection. In addition, Wang et al. [27] proposed a color image splicing detection method based on gray-level co-occurrence matrix (GLCM) of chromaticity threshold edge image. Tamper detection based on image compression attribute difference is mainly aimed at JPEG format images, which is divided into double JPEG compression detection and block effect inconsistency detection. In [28] firstly, a region in the image is manually selected, and the DCT coefficients in the region are Fourier transformed, and the original quantization matrix is estimated by the frequency domain characteristics. Then, the block effect of the whole image is calculated by using the quantization matrix of the region. The region with large difference is the tampered region. Iakovidou et al. [24] proposed a new method to detect image forgery by locating grid alignment anomalies in the bitmap of JPEG compressed images. Tamper detection based on perceptual image hashing is to extract the global or local features of the image to form a hash sequence, so as to achieve image splicing tamper detection. Wang et al. [29] employed to use Watson's visual model to extract visual sensitive features, and then generate robust perceptual hash code by combining image block based features and key point based features.

Although the above methods have certain effect on the forgery detection of specific images, when the specific attributes in the image do not exist or are not obvious, the detection failure will occur. For example, (1) when the original part and the tampered part of the tampered image come from the imaging device with the same property, the image will contain the same noise attribute; then, the image tamper detection method based on the difference of the equipment imaging will be invalid. (2) Usually after the image splicing forgery, the corresponding post-processing will be carried out, which will seriously weaken the essential attribute difference between the original part and the tampered part of the image, which will make the corresponding detection method invalid. (3) The difference of image compression attributes is usually suitable for JPEG format image tamper detection, but not for other types of images. (4) The tamper detection method based on perceptual image hash needs the hash value of the original image for tampering detection, so this method cannot be applied to blind detection of unknown images.

The core of deep learning method is to extract the main representation features from the data by using a series of nonlinear transformations. The extracted representation features are multi-dimensional, multi-level, and multi-angle. Therefore, the features extracted by deep learning have stronger generalization and expression ability. As an effective deep learning model of image processing, convolutional neural network has been widely concerned by researchers. For example, Vasan et al. [30] proposed a visualization method of malware, and used the fine-tuned convolution neural network structure to detect the visualized image. Gadekallu et al. [31] employed a novel PCA—which optimization based deep neural network model for

classification of tomato plant diseases using GPU. Based on its translation invariance and other attributes, convolutional neural network has achieved great success in image classification, semantic segmentation, image generation [32, 33], and object detection. For example, the current outbreak of COVID-19, convolutional neural network for medical image segmentation can play an important role in epidemic prevention and control [34, 35]. In addition, the use of convolutional neural networks for the detection of common diseases in medicine has also achieved good detection results [36–38]. The convolution neural network is applied to image tamper detection [39–41], which can realize the self-learning of tamper features, thus making up for the defects of traditional methods that rely on single image attributes and lack of generalization ability. Bappy et al. [42] proposed a model based on CNN-LSTM to learn the difference characteristics of tampered area and non-tampered region in the shared boundary, so as to locate the tampered area. However, this method uses the block method to detect and cannot effectively combine the feature information of the context for training. In order to make up for the lack of context information in the model, bappy et al. then proposed a hybrid LSTM and encoder–decoder architecture for detecting image forgery [9]. Yang et al. [43] employed a coarse to fine image tamper detection architecture, namely constrained R-CNN. The model can help the network pay more attention to the learning of tampering operation features, but it will cause some feature information loss in the constrained convolution operation. Xiao et al. [44] adopted a detection framework combining cascaded convolutional neural network and adaptive clustering algorithm, which can realize tamper localization from coarse to fine. However, the results of clustering can only roughly locate the tamper location, and only apply to a single tamper object. Cun et al. [45] believed that the splicing area was not only related to local features, but also highly related to global features. Based on this, the image splicing forgery location based on semi global network and full connected condition random field is proposed. Zhou et al. [8] proposed a two-stream Faster R-CNN network detection architecture. The two streams are noise stream extracted by spatial rich model (SRM) and RGB stream of image. However, when the tampered area and the non-tampered region have the same noise, the detection effect of the detection model will be reduced. Reference [46, 47] employed an image tamper detection framework based on improved U-Net structure. Bi et al. [46] constructed a ring-shaped residual U-Net network for image splicing forgery detection. The core idea is to enhance the learning mode of CNN through the residual propagation and feedback process in CNN. Zhang et al. [47] used the spatial rich model (SRM) to capture the residual signal in the image to be detected, and then combined it with the RGB features of the image, and the combined features were used for the training of U-Net model. However, this method has the same defects as Zhou et al. [8], and the detection effect will be reduced when the tampered region and the non-tampered region have the same source. Biach et al. [48] adopted a method based on the encoder/decoder structure to locate the tampered region and used resnet50 in the encoder structure to effectively improve the detection accuracy of tampered region. However, this method has high computational complexity due to its relatively complex model. Rao et al. [49] proposed a multi-semantic attention model and integrated it into a convolutional neural network to realize image forgery detection and localization. The tamper detection performance of this method for small objects needs to be improved. Bi et al. [50] employed an image forgery detection method based on D-Unet. This method is based on the traditional U-Net, uses DWT (Haar discrete wavelet transform) to extract the boundary information of the tampered area, and designs the SPGFE module for global feature extraction. The results show that the method can effectively improve the detection accuracy.

In order to accurately locate the forgery image at the pixel level, we propose a tamper detection method based on dual-channel U-shaped network model. The proposed method is shown in Fig. 2. Firstly, we extract the residual image of the tampered image by high-pass filters. The edge information of residual image extracted by high-pass filters will be enhanced. Because the objects after splicing usually leave tampering information in the boundary, the residual image extracted by high-pass filters will provide additional anomaly feature information for tamper detection model. Secondly, although the residual image extracted by the high-pass filter strengthens the edge information, it will also cause the loss of the content information of the image. Image forgery detection requires the learning of tampering part of the content features, so the learning of the original image features is also very important. The structure of the dual-channel model is designed by the features of the original image and the residual image as the input. Then, two feature fusion processes were carried out between the encoder and decoder in the DCU-Net model. The first feature fusion is the fusion of deep original image features and residual image features extracted by two channel encoders. The second feature fusion is to extract feature information of different scales by using dilation convolution and then to fuse the extracted feature information of different granularity. Finally, the fused features are input into the decoder for decoding. In the process of decoding, the feature information extracted from the encoder is added to the decoder by using the skip connection, so as to reduce the information loss in the convolution process.

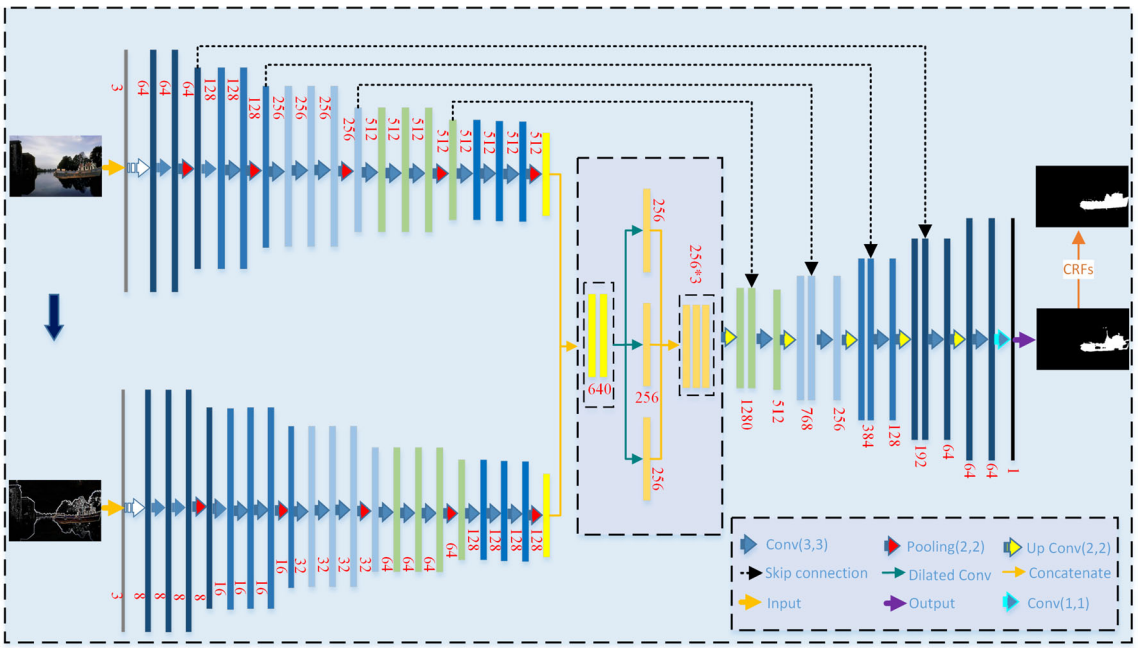

Fig. 2 Proposed splicing tamper detection model framework. The input of the framework is a dual-channel structure. The upper channel takes the tamper image itself as the input, and the input of the lower channel is the residual image extracted by high-pass filters

In summary, the major contributions of this study are described as follows:

- A dual-channel U-shaped network model, that is, DCU-Net, is proposed, which can accurately locate the tampered area at pixel level.
- A training method is combining the residual image extracted by high-pass filters and the original tamper image is proposed, which provides more tamper feature clues for model learning.
- The proposed model adds multi-feature fusion mechanism between encoder and decoder, which can extract tamper features more effectively and extract feature information of different granularity better.

The rest of this paper is arranged as follows. The second section introduces the implementation details of DCU-Net model, including residual image extraction, model structure design and post-processing. The third section describes the experimental results on different datasets and discusses the experimental results based on subjective evaluation and index evaluation. The fourth section is the conclusion, which summarizes the model and experimental results, and describes the future work.

2 Proposed detection method

We regard the features of the tampered area in the image as anomaly features [51] and, based on this, use U-Net for the learning and extraction of anomaly features of the forged area. The literature [46, 47] shows that U-Net has achieved good results for forgery detection. The proposed image splicing forgery detection framework is shown in Fig. 2. The design structure of the framework is specially designed for image splicing tamper detection. The dual-channel structure can make the model learn more tamper features and content features of tampered area. At the end of the model, we add the full connection conditional random field (FCRF) [52] and morphological opening operation to do further post-processing for the predicted image, which is used to remove the over segmented pixels and supplement the missing pixels. This section mainly introduces the following three parts: Sect. 2.1 introduces the process of extracting residual image using high-pass filters; Sect. 2.2 details the model structure; and Sect. 2.3 describes the post-processing based on FCRFs and morphological opening.

2.1 Residual image

Considering that high-pass filter has shown good results in different image forgery detection [53, 54], we use the residual image extracted by high-pass filter as additional information for image splicing detection. We have designed filters in the vertical and horizontal directions. Use this filter at any point in the image, which can be used as a measurement of the amount of change in the image in the vertical and horizontal directions. The filters consists of

two 3×3 matrices(S_x, S_y), which are used for horizontal and vertical operations, respectively.

$$S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (1)$$

When high-pass filters are convoluted with the image in plane, the approximate values of measurement difference in horizontal and vertical directions can be obtained, respectively. Finally, the residual image can be obtained by superimposing the horizontal and vertical filtering images. An example of residual image is shown in Fig. 3.

2.2 Model structure

Most of the deep learning models for image tampering localization use continuous convolution and deconvolution operations, which will inevitably cause information loss in the process of feature extraction. Compared with the traditional model, the significant improvement of U-Net model [55] is to add the intermediate skip connection structure, which can add low-level semantic information lost by the front-end network to the later high-level semantic information, thus avoiding the loss of low-level semantics. Therefore, we use a model similar to U-Net structure for image splicing tampering detection. Figure 4 shows the structure of the model.

2.2.1 Feature encoder module

In the traditional U-Net model, each convolution module of encoder contains two convolution layers and one maximum pooling layer. Our proposed method includes two encoders, one is used to extract the RGB feature information of tampered image, and the other is used to extract the tampered feature information from residual image. The model structure diagram is shown in Fig. 4. The input of the left encoder is the splicing tampered image, and the input of the right encoder is the residual image. For the encoder used for RGB feature extraction of tampered images, we replace the feature encoder with vgg16 structure [56], which retains five feature extraction blocks without full connection layer. The first and second convolution blocks contain two convolution layers and a max-pooling layer. The third, fourth and fifth convolution blocks contain three convolution layers and a max-pooling layer, respectively. Compared with the original encoder structure, using the pre-trained vgg16 weight can make the model achieve the optimal training effect faster and avoid falling into the local optimal solution, and can extract deeper feature information. For the encoder used in residual image feature extraction, we use ResNet module [57] instead of the traditional convolution module and set the stride to 2 instead of pooling operation. Compared with the original convolution module, ResNet adds a shortcut connection, which can effectively avoid the vanishing gradient and increase the convergence speed of the model.

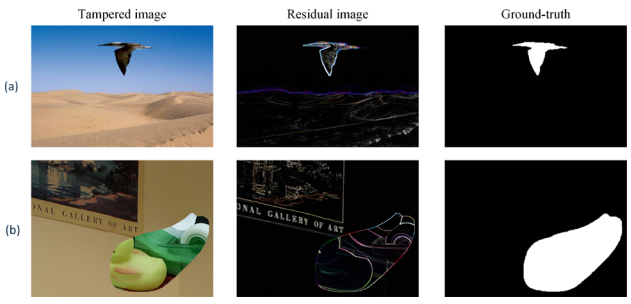

Fig. 3 Example of residual image. **a** and **b** are from Casia2.0 and Columbia datasets, respectively. The first column is tampered image, the second column is residual image, and the third column is ground-truth image

Residual learning: since the tamper feature information in the residual image is weak, in order to reduce the information loss in the convolution process, we use ResNet module to replace the original convolution block. The original convolution block and ResNet module structure are shown in Fig. 5a and b. Traditional convolution blocks are usually composed of several convolution layers. The underlying mapping fitted by several convolution layers can be expressed as $H(x)$. The results calculated from these stacked convolution layers are close to $H(x)$. Therefore, the definition of ordinary convolution block can be expressed as follows:

$$y = F(x, w) \quad (2)$$

where x and y represent the input and output of the convolution block, $F(\cdot)$ represents the mapping function, and w represents the weight to be learned. In the residual learning, it is assumed that the input of the residual module is x and the expected output is $H(x)$. If we pass input x directly to output, let $H(x) = F(x) + x$. Then our module only needs to learn a residual function $F(x) = H(x) - x$, instead of approaching $H(x)$ directly. Therefore, the definition of residual can be expressed as follows:

$$y = F(x, w) + x \quad (3)$$

where the function $F(\cdot)$ represents the residual mapping learned from the residual block. The residual structure is shown in Fig. 5b, and the residual learning can be realized by the shortcut connection. When the network deepens, using this simple residual structure can better solve the problem of information loss in the convolution process.

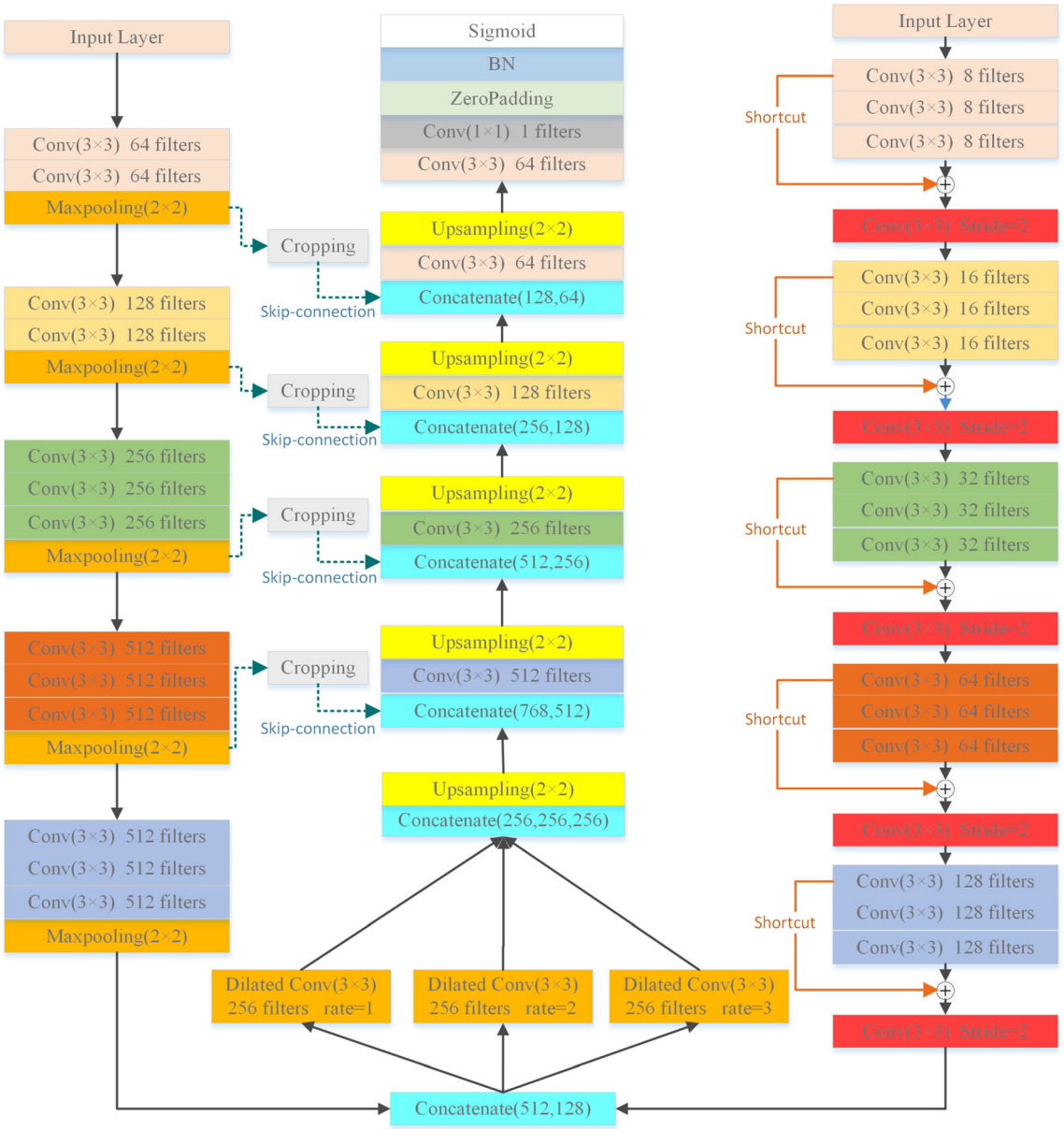

Fig. 4 DCU-Net model structure for image splicing forgery detection. There are two encoders on the left and right sides. The input is tampered image and residual image. At the bottom is the feature fusion module, the first is the fusion of image features and residual features extracted by two convolution channels, and the second is the fusion of different granularity features extracted by dilated convolution. In the middle is the decoder, which is used for the final image splicing forgery positioning

2.2.2 Feature fusion

The model structure proposed in this paper is a dual-channel model structure. As shown in Fig. 4, the two channels are convoluted layer by layer to extract the deep feature information of tampered image and residual image. The target of the first fusion is the deep features extracted from the two channels; the second fusion is the feature fusion after multi-scale dilated convolution of the first fusion results.

Due to the complexity of the tampered parts in the tampered images, the tampering positions of different tampered images are quite different. For example, some tampered objects are larger and some tampered objects are very small. Based on this, we propose a method of multi-scale dilated convolution [58] fusion. The proposed fusion method mainly depends on different dilated rate to expand the field of view of the filter, so as to better detect different sizes of tampered objects. Mathematically, the dilated convolution under 2-D signal is calculated as follows:

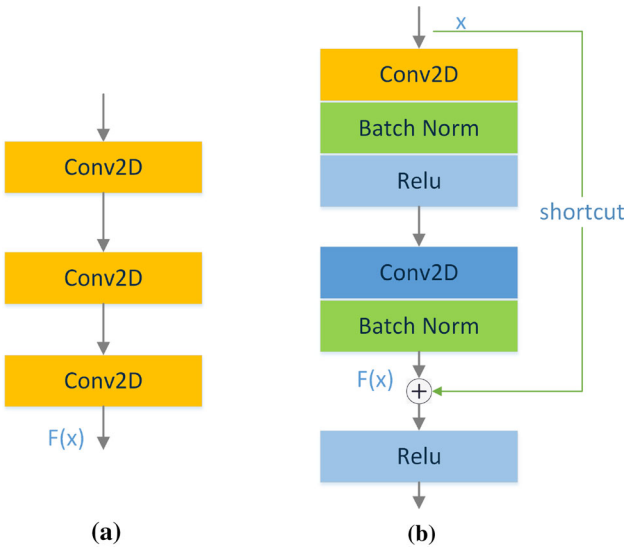

Fig. 5 Two types of convolution modules. **a** Common module. **b** Residual module

$$y[i] = \sum_{k} x[i + rk]w[k] \tag{4}$$

where x and w are input feature map and filters, respectively. r is the dilated rate, which determines the sampling step of the input signal. It is equivalent to convolute the input x with the upsampling filter. The upsampling filter is generated by inserting $r - 1$ zeros between two consecutive filter values in each spatial dimension. The schematic diagram of dilated convolution is shown in Fig. 6.

The information of context semantics is mainly determined by the size of the receptive field. If the receptive field can provide richer information, then more context information can be used. In order to extract the features of tampered targets with different scales, we use the dilated convolution operation mentioned above. The dilated convolution can enlarge the receptive field arbitrarily without introducing additional parameters and can utilize the context information of the image, so it is very suitable for multi-scale image tamper detection tasks. As shown in Fig. 7, we use three branches to receive the semantic information in the encoder module. First, set the dilated rate in dilated convolution to 1, 2, and 3 to expand the

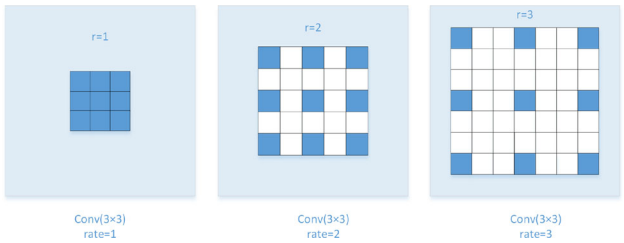

Fig. 6 Dilated convolution example graph

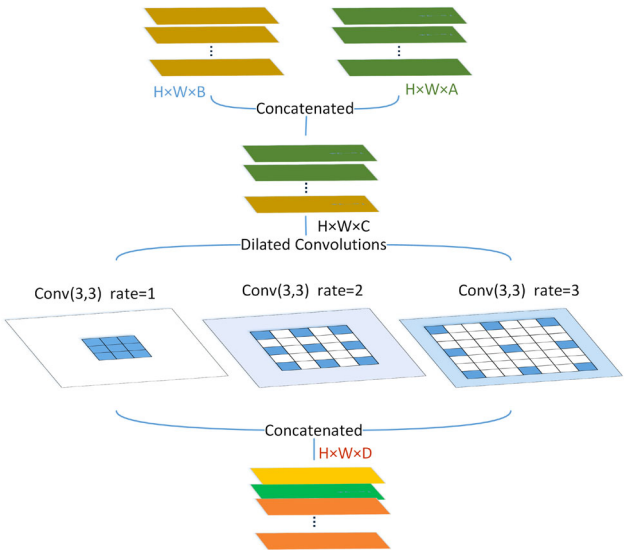

Fig. 7 Feature fusion: the first fusion is for the features extracted by two channels, and the second fusion is for the features extracted by dilation convolution

receptive field, so as to extract feature information of different scales in the encoder module. Then, the image semantic features extracted from different dilated rates are fused.

2.2.3 Feature decoder module

The feature decoder is used to recover the high-level semantic information extracted from the context semantic extraction module and feature encoder. The decoder maps the low-resolution feature image back to the size of the input image by the pixel by pixel classification. We add this mechanism in DCU-Net model according to the skip connection mode of U-Net model. Skip connection mechanism can add low-level semantic information from encoder to decoder to make up for the information loss caused by continuous convolution and pooling. As shown in Fig. 4, we only add the semantic information in the RGB feature encoder into the decoder. Because more content features are lost in the process of residual image extraction. Therefore, in order not to affect the training effect of the model, we did not add low-level residual semantic information to the decoder. In addition, in the process of decoding, the feature map dimensions may not match(because in the downsampling process, when the height or width of the image cannot be divided by 2, the feature map recovered by upsampling will not match the dimension of the corresponding down sampling feature), so we use the "cropping" method to prune the feature map extracted from the encoder. "Cropping" can prune the dimension of the feature map in the encoder according to the dimension of the feature map in the decoder. The pruned feature map

will be consistent with the dimension of feature map in decoder. In order to reduce the complexity of the model, we only perform one feature convolution for the features fused by skip connection. Through experiments, we find that the performance of DCU-Net model cannot be improved by using two convolution operations. As shown in the decoder structure in Fig. 4, each module in our decoder contains a concatenate block, a conv(3×3) convolution block, and an upsampling block. Finally, we use the "zeropadding" operation to solve the possible mismatch between the output dimension and the ground-truth dimension. Because the final result of classification is binary classification of pixels, so we use sigmoid as classifier.

2.2.4 Loss function

As shown in Fig. 1, the proposed DCU-Net detection framework is an end-to-end deep learning system. We need to train the framework to determine the tampered and non-tampered regions, that is, to classify pixels into foreground (tampered) and background, which is essentially a binary classification of pixels. The most common loss function in deep learning framework is cross-entropy loss function, so we apply binary cross-entropy loss function to calculate the loss in the training process. The formula of binary cross-entropy loss function is as follows:

$$L_{\text{bce}} = -\frac{1}{N}\sum_{i}^{N} g(x)\log p(x) + (1-g(x))\log(1-p(x)) \tag{5}$$

where N is the number of pixels, $g(\cdot)$ is the expected output, that is, the real data label, the value is $g(\cdot) \in \{0,1\}$; $p(\cdot)$ is the actual output, the value is $p(\cdot) \in [0,1]$. However, in the image tampering, the size of the tampered object is uncertain. Sometimes the tampered object only occupies a very small part of the whole image. When the sample is unbalanced, it is not optimal to use the cross-entropy loss function only. Because dice loss function directly takes the evaluation index of segmentation effect as the loss to supervise the model training, and ignores a large number of background pixels in the calculation, so it has a good effect on the sample imbalance problem. Therefore, in this study, we combine the binary cross-entropy loss with dice loss. The formula of dice loss function is as follows:

$$L_{\text{dice}} = 1 - \frac{2\sum_{i}^{N} p(x)g(x)}{\sum_{i}^{N} p^2(x) + \sum_{i}^{N} g^2(x)} \tag{6}$$

The final loss function is defined as:

$$L_{\text{loss}} = w * L_{\text{bce}} + L_{\text{dice}} \tag{7}$$

where w is the weight applied to the binary cross-entropy loss function. In addition, in order to improve the prediction accuracy and speed up the training, we use adaptive moment estimation (Adam) with nesterov momentum as the optimization algorithm for model training. Compared with the traditional optimization algorithm, Adam has the advantages of high computational efficiency, small memory consumption, and adaptive adjustment of learning rate.

2.3 Post-processing

Using DCU-Net model for tamper detection can obtain relatively accurate pixel-level tamper location. However, the obtained image still contains some under segmented and transitional segmented localization images. Based on this, we choose full connection conditional random field to further refine the detection results. FCRFs can further process the results of deep learning prediction combined with the relationship between all pixels in the tampered image. FCRFs can optimize the rough and uncertain marks in the classification image, correct the small misclassification region, and get more detailed segmentation boundary. After FCRF processing, there may be outliers in the image. We use opening operation to deal with it.

2.3.1 Fully connected conditional random field

For image I and label X, I is defined as the random field containing the observation set $\{I_1, I_2, \cdots, I_k\}$, and X is the random field containing the hidden state set $\{X_1, X_2, \cdots, X_k\}$. I_k is the eigenvector of pixel k ; X_k is the label of pixel k. X and I can be modeled as conditional random fields, which can be described as follows with Gibbs distribution:

$$P(X = x|I) = \frac{1}{Z(I)}\exp(-E(x|I)) \tag{8}$$

where $P(X = x|I)$ is the posterior probability of label distribution X when the pixel distribution is I ; $Z(I)$ is the conditional probability normalization factor; $E(x|I)$ is the energy function. The Gibbs energy function $E(x)$ can be expressed as follows:

$$E(x) = \sum_{i} \psi_u(x_i) + \sum_{i<j} \psi_p(x_i, x_j) \tag{9}$$

In the formula, the unary potential function $\psi_u = -\log P(x_i)$ represents the probability that the global coarse classification pixel i belongs to a certain class. It can use the shape, structure, color, and texture information in the image. $\psi(x_i, x_j)$ is a pairwise potential function, which is expressed as follows:

$$\psi_p(x_i, x_j) = \mu(x_i, x_j) \sum_{m=1}^{M} w^{(m)} k_G^{(m)}(f_i, f_j) \tag{10}$$

where f_i and f_j are the eigenvectors of pixels i and j ; $k_G^{(m)}$ is the Gaussian kernel function acting on the eigenvector; m is the number of Gaussian kernels; and $w^{(m)}$ is the weight corresponding to the Gaussian kernel function. $\mu(x_i, x_j)$ is the error penalty term, which can be expressed as follows:

$$\mu(x_i, x_j) = \begin{cases} 1 & x_i \neq x_j \\ 0 & x_i = x_j \end{cases} \tag{11}$$

From the error penalty term, we can see that only when the pixel i and j labels are different, the pairwise potential function has a value.

2.3.2 Morphological opening

After FCRF processing, the predicted image may have broken boundaries and small isolated pixels. Morphological opening operation can be used to eliminate small objects, smooth shape boundaries, and do not change their area. Therefore, we further do the image processing. In mathematical morphology, the opening operation is defined as corrosion first and then dilation. The structure element B is used to perform morphological opening operation on target image A, which is represented by symbol $A \circ B$. It is defined as:

$$A \circ B = (A \ominus B) \oplus B \tag{12}$$

where $\ominus$ is the corrosion operation and $\oplus$ is the dilation operation. After the morphological opening operation, the predicted image is shown in Fig. 8.

3 Experiment

This section mainly introduces a variety of comparative experiments. Section 3.1 describes the details of the experiment, including experimental parameters, datasets and evaluation criteria. Section 3.2 shows the comparison results between the proposed method and other methods. Section 3.3 introduces the results of attack experiments, including Gaussian noise attack and JPEG compression attack.

3.1 Experimental details

3.1.1 Experimental parameters

The environment used in this study is based on Linux system. Keras and tensor flow framework are used to build and test the model. The GPU on the server is configured as Tesla V100 and the memory is 16GB. The epoch of model training is 300 and the batch size is 16. The initial learning rate is set to 0.0001, and the learning rate is set to 0.00001 after 100 iterations. The weight used by in the loss function is set to 0.01. During the experiment, the proportion of validation set is 0.2. When the fully connected random field is used for post-processing, the number of FCRF inferences is 5 times.

3.1.2 Dataset

In this paper, we use two public datasets as experimental data, which are Casia v2.0 [59] and Columbia [60] uncompressed data. Casia includes two types of tampering: copy–move and cut–paste. We selected 1062 groups of splicing tampered images from Casia dataset as experimental data, including 967 training data and 95 test data. In order to prevent overfitting due to less data samples and improve the training effect of the model, we expand the training data. We have expanded the training data by using three times flipping operation, and expanded training data is 3868 groups. Similarly, we selected 90 groups of cut–paste tampered images from the Columbia uncompressed dataset and expanded them. After three times of flipping, the data volume of Columbia was 360 groups. We selected 340 groups as training data and 20 groups as test data.

3.1.3 Evaluating indicator

In order to make a quantitative evaluation of the experimental results of this study, we select the precision, recall, F-measure, and accuracy as the evaluation index.

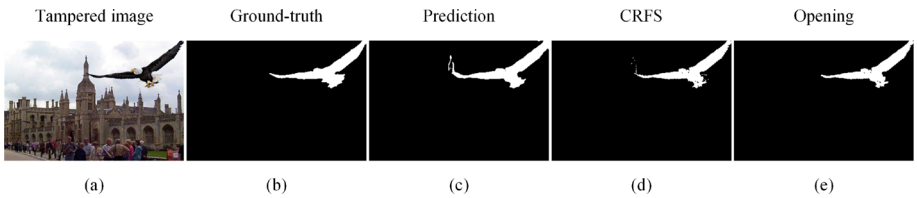

Fig. 8 Image of post-processing results. In the figure **a–e** are original tampered image, ground-truth image, prediction image, FCRF processing image and opening operation processing image

$$\text{Recall} = \frac{T_p}{T_P + F_N} \tag{13}$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{14}$$

$$F-\text{measure} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{15}$$

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{16}$$

where T_P, T_N, F_P, and F_N are true positive, true negative, false positive, and false negative, respectively.

3.2 Comparative experiment and analysis

This section introduces the comparison and analysis of the experimental results of DCU-Net model and other detection models. We have carried on the contrast experiment from two aspects: the intuitionistic result and the quantitative index evaluation.

3.2.1 Detection result

DCU-Net model intuitive detection results: we have done experiments on Casia and Columbia datasets. Several groups of test result images are randomly selected from the test results. Figure 9 shows the detection results of this study. It can be seen from the detection results that the DCU-Net model proposed in this study has achieved good results, and can carry out accurate pixel-level tamper location.

Baseline method test results: in order to compare the effectiveness of the proposed methods, we will carry out experiments on some baseline methods. We compare the DCU-Net model proposed in this paper with these baseline methods, and give an intuitive image of the detection effect (as shown in Fig. 10). Each baseline method is described as follows:

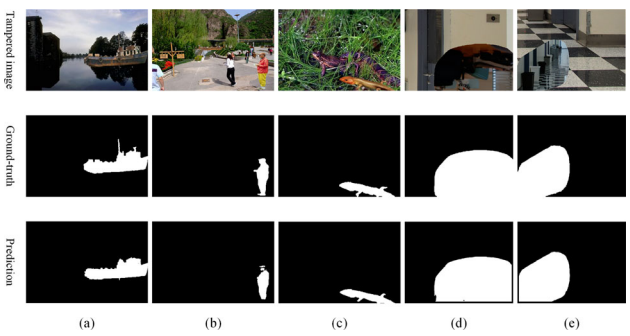

Fig. 9 Tamper location detection result example image. **a**–**c** are the detection result of images in Casia dataset; **d** and **e** show the detection results of images in Columbia dataset

- NADQ [61]: based on the derivation of a unified statistical model that characterizes DCT coefficients when aligned or misaligned double JPEG compression is applied; the statistical model is used to generate a likelihood map that shows the probability of each 8×8 image block being double compressed.
- CFA [62]: the techniques are based on computing a single feature and a simple threshold based classifier. By interpreting the locally nonexistent CFA artifacts as tampering evidence, the proposed scheme takes the forgery graph of the credible probability of each small pixel block as the output.
- FCN [63]: this method is used to classify images at the pixel level, thus solving the problem of image segmentation at the semantic level. FCN is applied to image splicing tamper detection, which can automatically learn image tampering features.
- U-Net [55]: as one of the models of full convolution network structure, it can also be used for pixel-level classification. Compared with FCN, the most important improvement is to add skip connection structure to reduce the loss of information.
- U-Net-RGB: based on the dual-channel model structure proposed in this study, the U-Net-RGB model only uses RGB channel as input. We remove the channel where the residual image is located, and only retain the channel of RGB image as the input, so as to verify the detection effect of using only RGB image.
- U-Net-Rimg: U-Net-Rimg model only uses residual image as input. We apply the residual image to the input channel of the original RGB image to test the effect of the residual image used in image splicing detection.

As shown in Fig. 10, in order to compare the performance of the DCU-Net model proposed in this study, we compare it with the six methods. Among all the comparison methods, there are two traditional detection methods based on feature extraction; two detection methods based on semantic segmentation; and the other two are from the branches of DCU-Net model. We randomly selected seven groups of data from the test data as examples, in which (a)–(d) these four groups are from the Casia dataset, and (e)–(g) the three groups are from the Columbia dataset. For Casia data, the splicing region forgery is more fine, and after processing operation; for Columbia data, the splicing region forgery is relatively rough. The first line and the second line are the original tampered image and the corresponding ground truth, and the rest are the detection results of seven detection methods. From the perspective of subjective vision, we can see that NADQ algorithm has a certain detection effect for splicing tampering in Casia dataset. Since NADQ is suitable for detecting JPEG

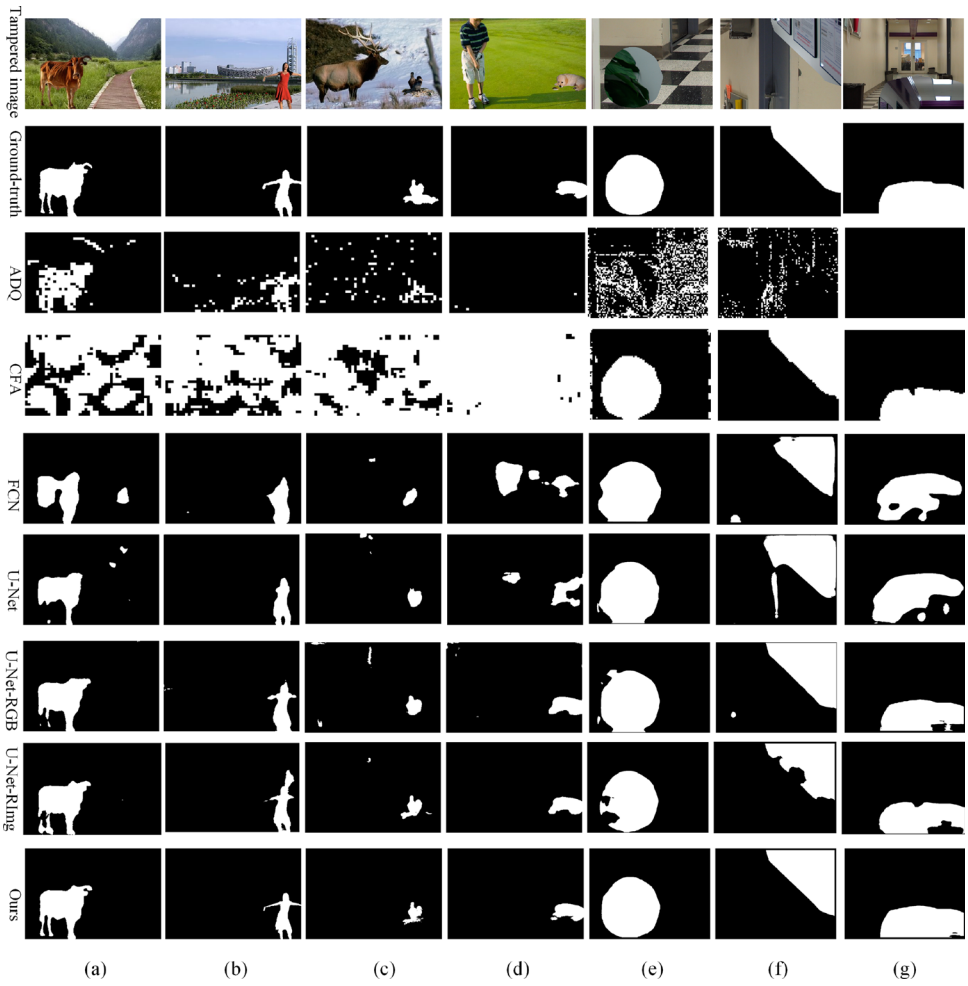

Fig. 10 Comparison between DCU-Net and other methods for splicing tamper detection. NADQ and CFA are traditional methods, and the rest are deep learning methods. **a–d** from Casia dataset, **e–g** from Columbia dataset

compressed data, it is invalid to detect tampered images in Columbia uncompressed data. CFA algorithm has a good detection result for tampering in Columbia data, but it is invalid for image detection in Casia. Although FCN and U-Net can effectively locate the tampered area, there are many problems of mis-segmentation and over segmentation. Compared with the traditional semantic segmentation method, the DCU-Net based branching method can improve the detection of tampered regions, but there are still a few mis-segmentation and over segmentation. Finally, our proposed method combines the RGB feature of the original tampered image with the residual image feature and performs a two-step post-processing operation, so it can locate the tampered area more accurately.

3.2.2 Index evaluation

In order to make a more objective and accurate evaluation for the method proposed in this paper, we use the four evaluation indexes mentioned above to evaluate DCU-Net more objectively. Compared with subjective evaluation, three algorithms are added for image splicing detection, including 11 comparison algorithms. Tables 1 and 2 show the evaluation results on Casia and Columbia data, respectively. From the evaluation results of pixel-level classification in Tables 1 and 2, it can be seen that the results of four evaluation indexes of NADQ method, F-measure, precision, recall, and accuracy, are relatively low. CFA algorithm has a certain detection effect for simple tampering in Colombia data, and achieves higher scores than NADQ in four evaluation indexes. On the whole, the image tamper detection based on deep learning is better than the traditional detection algorithm. C2RNet [44] proposed a coarse to fine splicing detection method and finally used clustering algorithm to post-processing the detection results. Although the tamper image can be located roughly, the tampered area cannot be accurately located after clustering. Therefore, the detection effect is poor in precision. RRU-Net [46] adds residual learning and feedback process to the traditional U-Net algorithm, which improves the detection effect of the model greatly. DU-DC-EC Net [47] proposed a cross-layer crossover mechanism and added SRM filter to capture the residual signal in the image, which effectively improved the F-measure and

Table 1 Casia dataset evaluation results

Method	F-measure	Precision	Recall	Accuracy
CFA [62]	0.2026	0.1439	0.6973	0.3488
NADQ [61]	0.2847	0.2777	0.4555	0.7274
FCN [63]	0.5470	0.6654	0.5308	0.9369
U-Net [55]	0.5978	0.6869	0.6121	0.9462
C2RNet [44]	0.6758	0.5810	0.8080	–
RRU-Net [46]	0.8410	0.8480	0.8340	–
DU-DC-EC Net [47]	0.6830	–	–	0.9782
D-Unet [50]	0.8590	0.8660	0.8520	–
DCU-Net-RImg	0.8209	**0.8912**	0.8175	0.9661
DCU-Net-RGB	0.8335	0.8620	0.8591	0.9693
DCU-Net-NFF	0.8525	0.8600	0.8735	0.9742
DCU-Net	**0.8667**	0.8772	**0.8893**	**0.9793**

Bold text in the table indicates the highest value in each column

Table 2 Columbia dataset evaluation results

Method	F-measure	Precision	Recall	Accuracy
CFA [62]	0.5836	0.7472	0.5994	0.8646
NADQ [61]	0.2378	0.3292	0.2254	0.6557
FCN [63]	0.6885	0.9001	0.6126	0.8847
U-Net [55]	0.7779	0.9850	0.6987	0.9134
C2RNet [44]	0.6950	0.8040	0.6120	–
RRU-Net [46]	0.9150	0.9610	0.8703	–
DU-DC-EC Net [47]	0.9307	–	–	0.9663
D-Unet [50]	0.9300	0.9600	0.9010	–
DCU-Net-RImg	0.8858	0.9965	0.8252	0.9407
DCU-Net-RGB	0.9175	**0.9981**	0.8637	0.9545
DCU-Net-NFF	0.9216	0.9971	0.9004	0.9647
DCU-Net	**0.9498**	0.9871	**0.9176**	**0.9727**

Bold text in the table indicates the highest value in each column

accuracy of the model. DCU-Net-Rimg and DCU-Net-RGB, respectively, use the single channel of our model. Tables 1 and 2 shows that the detection effect of the two models has been greatly improved. DCU-Net-NFF means that it only contains dual-channels and does not include multi-scale feature extraction and fusion. Finally, the model proposed in this paper is a dual-channel input model, which combines the RGB features of the original tampered image and the residual image features of the tampered edge, making the results of the four evaluation indicators all get the optimal.

3.3 Comparative analysis of attack experiments

In order to further evaluate the effectiveness and robustness of the algorithm, we have carried out attack experiments on Casia and Columbia datasets. The two attacks used in this article are Gaussian noise attacks and JPEG compression attacks, which are common in image data. The experimental results and evaluation indexes are shown in Figs. 11, 12, 13 and 14.

3.3.1 Comparison of experimental results of Gaussian noise attack

This part discusses the influence of Gaussian noise on image tamper detection. a–c in Fig. 11 show the experimental results of Casia data under different variance noise attacks, and a–c in Fig. 12 show the experimental results of Columbia data under different variance noise attacks. From the experimental results, we can see that with the increase in the Gaussian noise, all the evaluation indexes show different degrees of decrease. It shows that Gaussian noise has a great influence on image tamper detection. As shown in Figs. 11 and 12, when the variance is 0.002, the F-measure and precision values of CFA and ADQ have almost reached the lowest point. With the increase in the Gaussian noise, the recall value of the CFA method in the two datasets is basically unchanged. This is because the CFA method determines almost all detection areas as tampered areas, which leads to detection failure. As shown in Fig. 15, when the variance is 0.004, the CFA method recognizes almost all areas as tampered parts. As the variance of Gaussian noise gradually increases from 0.002 to 0.01, the three evaluation indicators of the detection method based on deep learning also show a gradual decline. It can be seen that the noise attack also has a great impact on the deep learning method, but there is no detection failure like the traditional method. In general, the method based on deep learning is still better than the traditional detection method in the final results. In addition, it can be seen from the detection index that with the increase in the Gaussian noise, the detection index of the proposed model is least affected. With the increase in the Gaussian noise, the proposed method still achieves the best results in F-measure and precision.

3.3.2 Comparison of experimental results of JPEG compression attack

This section discusses the impact of JPEG compression attack on detection results. a–c in Fig. 13 show the results of three evaluation indexes when JPEG compression attack is added to Casia data, and a–c in Fig. 14 show the results of three evaluation indexes when JPEG compression attack is added to Columbia data. From the overall detection index results, JPEG compression has a certain impact on the detection of all models, but the influence of deep learning based method is less than that of traditional

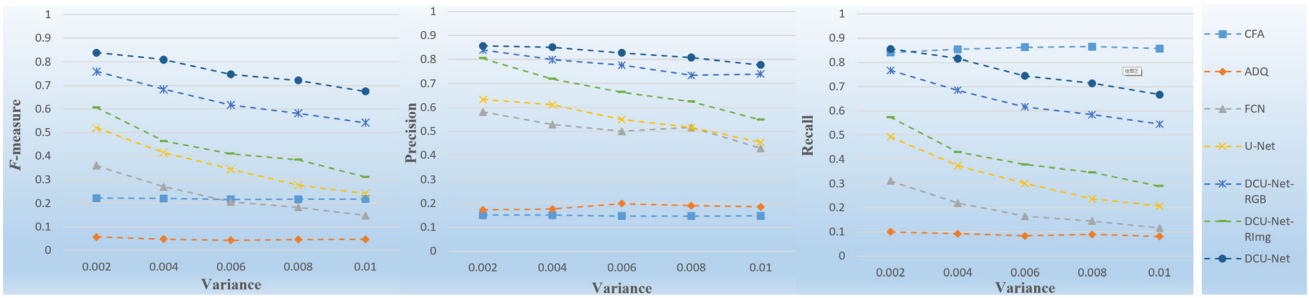

Fig. 11 Casia data detection results. **a–c** Show the experimental results of adding Gaussian noise to Casia dataset

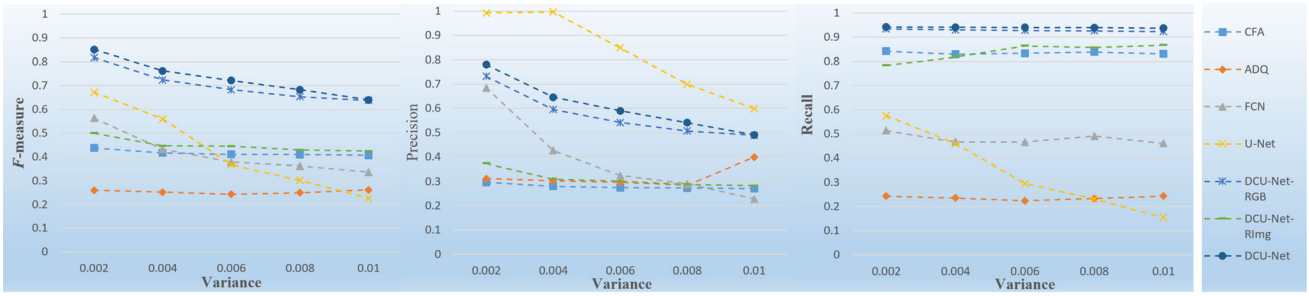

Fig. 12 Columbia data detection results. **a–c** Show the experimental results of adding Gaussian noise to Columbia dataset

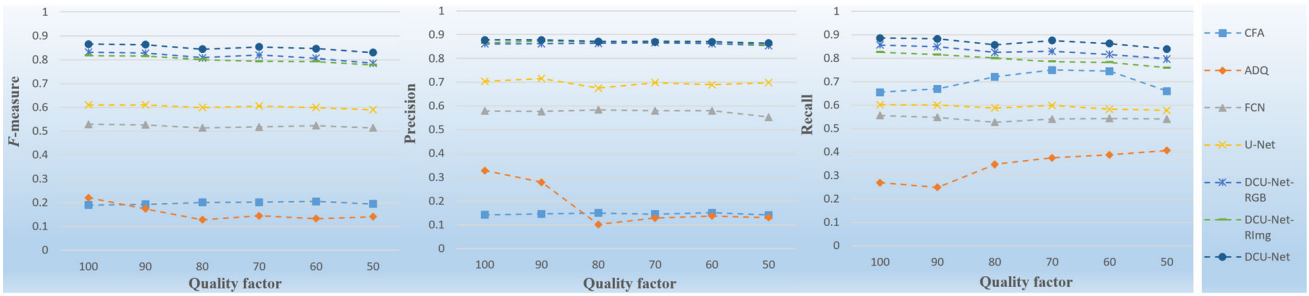

Fig. 13 JPEG compression attack in Casia data. **a–c** Show the experimental results of JPEG compression attack in Casia dataset

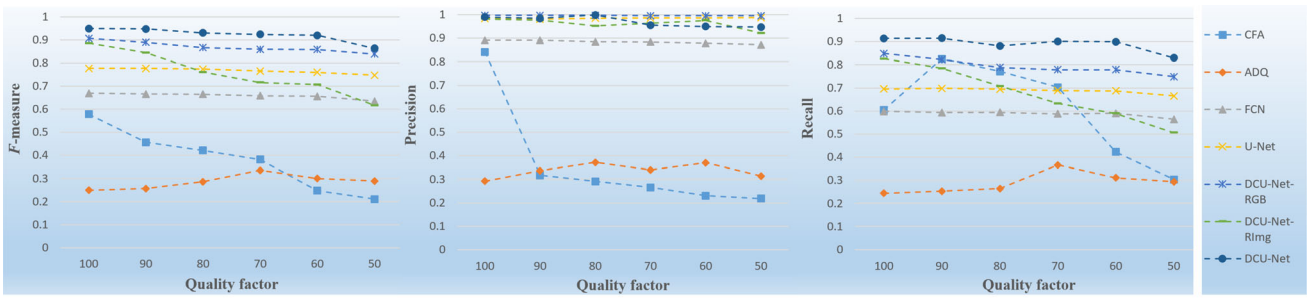

Fig. 14 JPEG compression attack in Columbia data. **a–c** Show the experimental results of JPEG compression attack in Columbia dataset

detection method. With the two traditional detection methods of CFA and ADQ, as the quality factor drops from 100 to 50, the values of the three evaluation indicators are already in a very low position. It can be seen from the

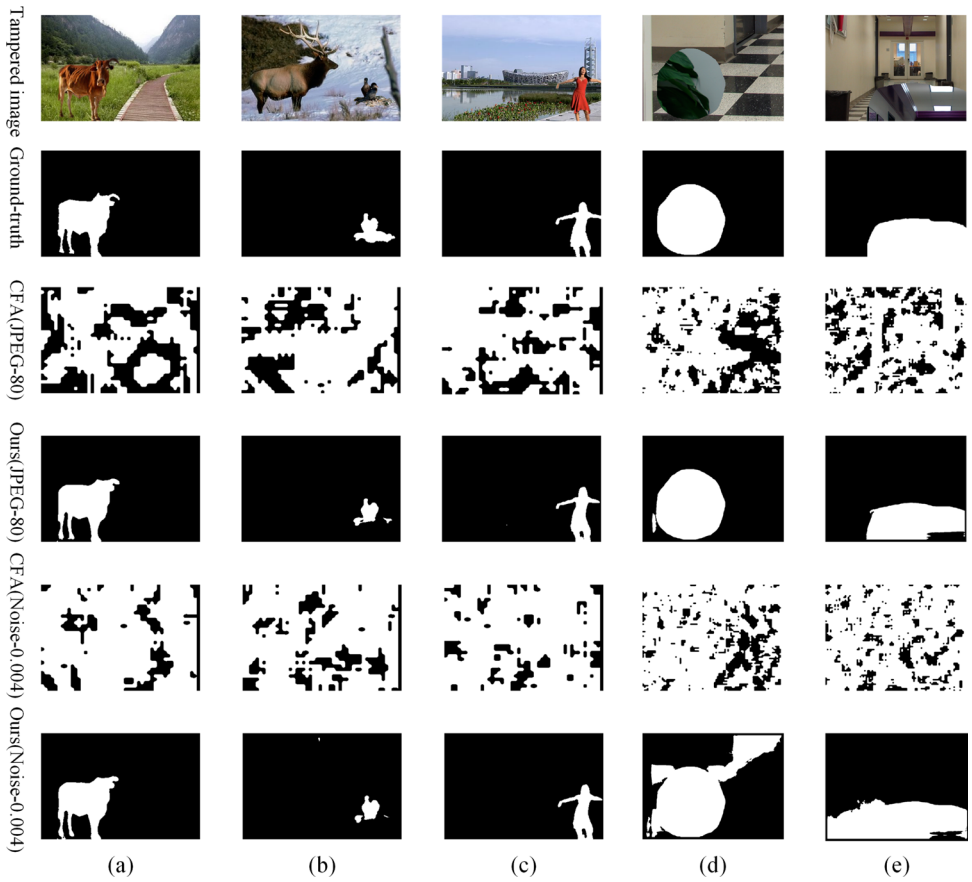

Fig. 15 Detection results of DCU-Net and CFA methods under two kinds of attacks. **a**–**c** Show the test results of Casia dataset, and **d** and **e** show the test results of Columbia dataset. The first and second lines show the tampered image and the ground-truth image. The third and fourth lines show the detection results of CFA method and DCU-Net method after adding JPEG compression attack; the fifth and sixth lines show the detection results of CFA method and DCU-Net method after adding noise

detection results in Fig. 15 that the two methods can no longer perform effective detection when the quality factor is 80. Deep learning method is a detection method based on image content features, which has good resistance to compression attacks, so it has little impact. In addition, from the three evaluation indicators, the method in this paper still has a good detection effect when the quality factor drops to 50.

3.4 Discussions

In this section, we conduct a comprehensive discussion on the above experimental results. The experimental results show that the traditional method only has a certain detection effect for tampered images with certain attributes, and its generalization ability is weak, and it is easy to be interfered by external attacks. The deep learning method can automatically learn the tamper features of the image, without manual extraction, so it can automatically detect the tampered image with various attributes, which are also the advantage of deep learning method. The DCU-Net model itself has an excellent detection effect; combined with the corresponding post-processing, it can achieve more accurate detection. First of all, from the subjective and objective index detection results, the methods we proposed have achieved better results. Subjectively: DCU-Net is superior to other methods in pixel-level positioning accuracy, with smaller pixel false positives and missing positives. Objective index: DCU-Net has achieved good results on three experimental indexes, which indicates that the method can not only locate effectively, but also locate accurately. Secondly, from the results of attack experiments, DCU-Net had better robustness. For Gaussian noise attacks, from the detection results of the Casia dataset and Columbia dataset, the detection method based on deep learning is generally better than the traditional detection method, and the phenomenon of detection failure will not occur as the noise variance increases. DCU-Net further reduces the influence of Gaussian noise, so as to obtain the best results compared to other methods. Regarding the JPEG compression attack, from the experimental results of the two types of datasets, JPEG compression did not have a significant impact on the deep learning method. The DCU-Net model maintains high detection performance even when the quality factor is minimized. Compared with other deep learning models, the main advantage of DCU-Net model is multi-feature fusion, which not only uses the RGB features of the image, but also applies the tamper boundary features of the tampered object in the residual image. Specifically, DCU-Net has designed a dual-channel

structure, namely RGB feature extraction channel and residual image feature extraction channel, and combined with the dilated convolution, effectively using the context semantic information. To sum up, the DCU-Net model proposed in this paper combined with the corresponding post-processing operations can not only effectively improve the accuracy of splicing forgery detection, but also improve the robustness of detection.

4 Conclusion

We propose an end-to-end dual-channel U-Net model for image splicing forgery detection, which can accurately locate the splicing forgery region. The final detection result of DCU-Net method proposed in this study is to locate the tampered area at pixel level. In order to achieve more accurate positioning, we use high-pass filters to extract the residual image of the tampered image, and add the residual image to the DCU-Net model. Based on this, we design a dual-channel input model, in which one channel is used to input the RGB features of the original tampered image, and the other channel is used to input the residual image features. In our DCU-Net model, the encoder of RGB channel adopts the model structure of vgg16, which is used to extract the deep-seated tampering features; the encoder of the residual image channel adopts the residual structure to better retain the edge features of the tampered area in the residual image. In the transition region of encoding and decoding, we design two feature fusions, which can better fuse the semantic information of context and provide different scale feature information. In the final detection and location, we use full connection conditional random field and opening operation to locate the tampered area detected by DCU-Net model more precisely. Through the test on Casia and Columbia data, our method has achieved better results among the four evaluation indexes. The F-measure, precision, recall, and accuracy of DCU-Net model on Casia are 0.8667, 0.8772, 0.8893, and 0.9793, respectively, and those of F-measure, precision, recall, and accuracy on Colombia dataset are 0.9498, 0.9871, 0.9176, and 0.9727, respectively. Finally, we test the robustness of the model by adding Gaussian noise and JPEG compression attack into the test image. The test results show that DCU-Net model has better anti-noise and anti-compression ability than other detection methods. In summary, the improved methods proposed in this paper are helpful to improve the accuracy and robustness of splicing forgery detection. These methods can provide a new research idea for image splicing forgery detection.

Although our method shows good performance, it still has some disadvantages. For example, (1) our method can only process images of a fixed size; (2)the detection ability is relatively single, which is only suitable for the detection of splicing images. Based on this, we will make further improvements from the above two shortcomings in future work. Specifically, (1) combine the method in this article with target detection models (such as Faster R-CNN and YOLO models) to solve the problem of only processing fixed-size images; (2) by further exploring the potential distinguishing features between the tampering/non-tampering areas of more types of forged images, a more general detection model can be designed; and (3) combined with the idea of transfer learning, an effective detection model suitable for cross-dataset training and detection can be proposed to improve the generalization ability of the model.

Acknowledgements This work has been supported by National Key Research and Development Program of China. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Zampoglou M, Papadopoulos S, Kompatsiaris Y (2017) Large-scale evaluation of splicing localization algorithms for web images. Multimedia Tools Appl 76(4):4801–4834
2. Joseph RM, Chithra AS (2015) Literature survey on image manipulation detection. Int Res J Eng Technol (IRJET) 2(04):2395–0056
3. Cozzolino D, Poggi G, Verdoliva L (2015) Efficient dense-field copy-move forgery detection. IEEE Trans Inf Forensics Secur 10(11):2284–2297
4. Wu Y, Abd-Almageed W, Natarajan P (2018) Busternet: detecting copy-move image forgery with source/target localization. In: Proceedings of the European conference on computer vision (ECCV). pp 168–184
5. Wu Y, Abd-Almageed W, Natarajan P (2017) Deep matching and validation network: an end-to-end solution to constrained image splicing localization and detection. In: Proceedings of the 25th ACM international conference on multimedia. pp 1480–1502
6. Huh M, Liu A, Owens A, Efros AA (2018) Fighting fake news: Image splice detection via learned self-consistency. In: Proceedings of the European conference on computer vision (ECCV). pp 101–117
7. Zhu X, Qian Y, Zhao X, Sun B, Sun Y (2018) A deep learning approach to patch-based image inpainting forensics. Signal Process Image Commun 67:90–99
8. Zhou P, Han X, Morariu VI, Davis LS (2018) Learning rich features for image manipulation detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 1053–1061
9. Bappy JH, Simons C, Nataraj L, Manjunath BS, Roy-Chowdhury AK (2019) Hybrid lstm and encoder-decoder architecture for detection of image forgeries. IEEE Trans Image Process 28(7):3286–3300

10. Zeng H, Zhan Y, Kang X, Lin X (2017) Image splicing localization using pca-based noise level estimation. Multimed Tools Appl 76(4):4783–4799
11. Benrhouma O, Hermassi H, El-Latif Ahmed AA, Belghith S (2016) Chaotic watermark for blind forgery detection in images. Multimed Tools Appl 75(14):8695–8718
12. Yao H, Wang S, Zhang X, Qin C, Wang J (2017) Detecting image splicing based on noise level inconsistency. Multimed Tools Appl 76(10):12457–12479
13. Liu B, Pun CM (2017) Multi-object splicing forgery detection using noise level difference. In: 2017 IEEE conference on dependable and secure computing. IEEE, pp 533–534
14. Zhu N, Li Z (2018) Blind image splicing detection via noise level function. Signal Process Image Commun 68:181–192
15. Zhang Y, Zhao C, Pi Y, Li S (2012) Revealing image splicing forgery using local binary patterns of dct coefficients. In: Communications, signal processing, and systems. Springer, New York, pp 181–189
16. Alahmadi A, Hussain M, Aboalsamh H, Muhammad G, Bebis G, Mathkour H (2017) Passive detection of image forgery using dct and local binary pattern. Signal Image Video Process 11(1):81–88
17. Zhang Q, Wei L, Weng J (2016) Joint image splicing detection in dct and contourlet transform domain. J Vis Commun Image Represent 40:449–458
18. Han Jong G, Park Tae H, Moon YH, Eom Il K (2018) Quantization-based markov feature extraction method for image splicing detection. Mach Vis Appl 29(3):543–552
19. Zhao X, Li J, Li S, Wang S (2010) Detecting digital image splicing in chroma spaces. In: International workshop on digital watermarking. Springer, pp 12–22
20. Chen B, Qi X, Sun X, Shi YQ (2017) Quaternion pseudo-zernike moments combining both of rgb information and depth information for color image splicing detection. J Vis Commun Image Represent 49:283–290
21. Liu Q, Cooper Peter A, Chen L, Cho H, Chen Z, Qiao M, Yuting S, Wei M, Sung AH (2013) Detection of jpeg double compression and identification of smartphone image source and post-capture manipulation. Appl Intell 39(4):705–726
22. Mire Archana V, Dhok Sanjay B, Mistry Narendra J, Porey Prakash D (2018) Automated approach for splicing detection using first digit probability distribution features. EURASIP J Image Video Process 2018(1):1–11
23. Amerini I, Becarelli R, Caldelli R, Mastio AD (2014) Splicing forgeries localization through the use of first digit features. In: 2014 IEEE International workshop on information forensics and security (WIFS). IEEE, pp 143–148
24. Iakovidou C, Zampoglou M, Papadopoulos S, Kompatsiaris Y (2018) Content-aware detection of jpeg grid inconsistencies for intuitive image forensics. J Vis Commun Image Represent 54:155–170
25. Zhao Y, Wang S, Zhang X, Yao H (2012) Robust hashing for image authentication using zernike moments and local features. IEEE Trans Inf Forensics Secur 8(1):55–63
26. Tagliasacchi M, Valenzise G, Tubaro S (2009) Hash-based identification of sparse image tampering. IEEE Trans Image Process 18(11):2491–2504
27. Wang W, Dong J, Tan T (2009) Effective image splicing detection based on image chroma. In: 2009 16th IEEE international conference on image processing (ICIP). IEEE, pp 1257–1260
28. Ye S, Sun Q, Chang EC (2007) Detecting digital image forgeries by measuring inconsistencies of blocking artifact. In: 2007 IEEE international conference on multimedia and expo. IEEE, pp 12–15
29. Wang X, Pang K, Zhou X, Zhou Y, Li L, Xue J (2015) A visual model-based perceptual image hash for content authentication. IEEE Trans Inf Forensics Secur 10(7):1336–1349
30. Vasan D, Alazab M, Wassan S, Naeem H, Safaei B, Zheng Q (2020) Imcfn: image-based malware classification using fine-tuned convolutional neural network architecture. Comput Netw 171:107138
31. Gadekallu TR, Rajput DS, Reddy MPK, Lakshmanna K, Bhattacharya S, Singh S, Jolfaei A, Alazab M (2020) A novel pca–whale optimization-based deep neural network model for classification of tomato plant diseases using gpu. J Real-Time Image Process. pp 1–14
32. Li W, Ding W, Sadasivam Ra, Cui X, Chen P (2019) His-gan: a histogram-based gan model to improve data generation quality. Neural Netw 119:31–45
33. Li W, Linchuan X, Liang Z, Wang S, Cao J, Lam TC, Cui X (2021) Jdgan: enhancing generator on extremely limited data via joint distribution. Neurocomputing 431:148–162
34. Bhattacharya S, Maddikunta PKR, Pham QV, Gadekallu TR, Chowdhary CL, Alazab M, Piran MJ et al (2020) Deep learning and medical image processing for coronavirus (covid-19) pandemic: a survey. Sustain Cities Soc 65:102589
35. Sedik A, Hammad M, El-Samie FEA, Gupta BB, El-Latif AAA (2021) Efficient deep learning approach for augmented detection of coronavirus disease. Neural Comput Appl 1–18
36. Alghamdi A, Hammad M, Ugail H, Abdel-Raheem A, Muhammad K, Khalifa HS, El-Latif AAA (2020) Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. Multimed Tools Applications. pp 1–22
37. Hammad M, Iliyasu AM Subasi A, Ho Edmond SL, El-Latif Ahmed AA (2020) A multitier deep learning model for arrhythmia detection. IEEE Trans Instrum Meas 70:1–9
38. Li W, Liu X, Liu J, Chen P, Wan S, Cui X (2019) On improving the accuracy with auto-encoder on conjunctivitis. Appl Soft Comput 81:105489
39. Wu Y, AbdAlmageed W, Natarajan P (2019) Mantra-net: manipulation tracing network for detection and localization of image forgeries with anomalous features. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 9543–9552
40. Horváth J, Montserrat DM, Hao H, Delp EJ (2020) Manipulation detection in satellite images using deep belief networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. pp 664–665
41. Shan W, Yi Y, Qiu J, Yin A (2019) Robust median filtering forensics using image deblocking and filtered residual fusion. IEEE Access 7:17174–17183
42. Bappy Jawadul H, Roy-Chowdhury AK, Bunk J, Nataraj L, Manjunath BS (2017) Exploiting spatial structure for localizing manipulated image regions. In: Proceedings of the IEEE international conference on computer vision. pp 4970–4979
43. Yang C, Li H, Lin F, Jiang B, Zhao H (2020) Constrained r-cnn: a general image manipulation detection model. In: 2020 IEEE International conference on multimedia and expo (ICME). IEEE, pp 1–6
44. Xiao B, Wei Y, Bi X, Li W, Ma J (2020) Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering. Inf Sci 511:172–191
45. Cun X, Pun CM (2018) Image splicing localization via semi-global network and fully connected conditional random fields. In: Proceedings of the European conference on computer vision (ECCV)
46. Bi X, Wei Y, Xiao B, Li W (2019) Rru-net: the ringed residual u-net for image splicing forgery detection. In: Proceedings of the

IEEE conference on computer vision and pattern recognition workshops
47. Zhang R, Ni J (2020) A dense u-net with cross-layer intersection for detection and localization of image forgery. In: ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, pp 2982–2986
48. El Biach FZ, Iala I, Laanaya H, Minaoui K (2021) Encoder-decoder based convolutional neural networks for image forgery detection. Multimed Tools Appl 1–18
49. Rao Y, Ni J, Xie H (2021) Multi-semantic crf-based attention model for image forgery detection and localization. Signal Process 108051
50. Bi X, Liu Y, Xiao B, Li W, Pun CM, Wang G, Gao X (2020) D-unet: a dual-encoder u-net for image splicing forgery detection and localization. arXiv preprint arXiv:2012.01821
51. Wu Y, AbdAlmageed W, Natarajan P (2019) ManTra-Net: manipulation tracing network for detection and localization of image forgeries with anomalous features. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)
52. Krähenbühl P, Koltun V (2011) Efficient inference in fully connected crfs with gaussian edge potentials. In: Advances in neural information processing systems. pp 109–117
53. Cozzolino D, Verdoliva L (2017) Single-image splicing localization through autoencoder-based anomaly detection. In: 2016 IEEE International workshop on information forensics and security (WIFS)
54. Verdoliva L, Cozzolino D, Poggi G (2015) A feature-based approach for image tampering detection and localization. In: IEEE workshop on information forensics and security
55. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted interventio. Springer, New York, pp 234–241
56. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556
57. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778
58. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal Mach Intell 40(4), 834–848
59. Dong J, Wang W, Tan T (2013) Casia image tampering detection evaluation database. In: 2013 IEEE China summit and international conference on signal and information processing. IEEE, pp 422–426
60. Hsu YF, Chang SF (2006) Detecting image splicing using geometry invariants and camera characteristics consistency. In: 2006 IEEE international conference on multimedia and expo. IEEE, pp 549–552
61. Bianchi T, Piva A (2012) Image forgery localization via block-grained analysis of jpeg artifacts. IEEE Trans Inf Forensics Secur 7(3):1003–1017
62. Dirik AE, Memon N (2009) Image tamper detection based on demosaicing artifacts. In: 2009 16th IEEE international conference on image processing (ICIP). IEEE, pp 1497–1500
63. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 3431–3440