



Copy-move detection method based on Decoupled Edge Supervision and multi-domain cross correlation modeling

Niantai Jing¹ · Jie Nie¹ · Jingyu Wang¹ · Xiaodong Wang¹ · Xinyue Liang¹ · Xuesong Gao²

Received: 18 January 2024 / Revised: 14 May 2024 / Accepted: 30 May 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

Copy move forgery detection is defined as the process of moving one region in the image (the source region) to another region in the image (the tampered region). Common classification methods include copy move forgery location (CMFL, which does not distinguish between source regions and tampered regions) while existing copy move source target detection (CMSTD) use limited information to distinguish the source and region. Since the edges of the tampered region able to act as the important clues such as blurring, bringing the challenge of detecting the copy-move forgery target with the complete structure. Therefore, we propose a Copy-move Detection Method based on Decoupled Edge Supervision and Multi-domain Cross Correlation (DM-Net), including Multi-scale Similar Region Detection module (MSD), Decoupled Edge Supervision module (DEM), and Multi-domain Correlation Modeling module (MCM), which can overcome the problem that the tamper trace is fuzzy caused by the post-processing operation. Specifically, the MSD module is proposed to extract coarse similar regions by multi-scale method. The DEM module is proposed to extract the tamper region by the method of decoupling edge supervision, which avoids information redundancy while using shallow edge features. The MCM module conducts cross-correlation modeling between the tampered, source and similar region, further optimizes detection targets of similar region by mining the correlation among multiple domains. By adding edge information, we can improve the efficiency of distinguishing source and target regions by 2%. We performed experiments on USC-ISI data set, and the accuracy was improved by 0.21% compared with CNN-T GAN method, and the F1-score index was improved by 0.87% compared with DOA-GAN. The accuracy of CASIA v2.0 data set is 2.89% higher than that of Busternet method, and the precision index is 3.98% higher than that of CMSD-STRD method on source. The accuracy of CoMoFoD data set is improved by 0.93% compared with CMSD-STRD method, and the recall index is improved by 20.23% compared with CMSD-STRD target, which proves that our method is superior to the most advanced method at present.

✉ Jie Nie
niejie@ouc.edu.cn

Extended author information available on the last page of the article

Keywords Copy-move forgery detection(CMFD) · Deep matching · Edge-supervision · Self-correlation

1 Introduction

With the development of science and technology and the popularization of the Internet, people can easily use image editing tools such as Photoshop, Meitu and GIMP to process pictures. Copy-move forgery is defined as copying a part of an image to another region of the same image, and the detection of it aims to locate a pair of similar regions, that is, the source and the tampered region. In the process of tampering, to hide tampering traces, the tamper usually carries out additional post-processing operations on the tampered image, such as scaling, blurring, etc., this leads to the tampered edge in the image, which work as the important clues to detect the tampered region, are hidden that brings challenges to the detection of the structurally intact target region.

The traditional copy-move forgery detection (CMFD) methods adopt handmade features, which can be divided into two categories, one is block-based method [1–9] and the other is keypoint-based method [10–23]. In the rapidly advancing landscape of computer vision [24, 25], deep learning has experienced explosive growth in recent years [26, 27]. Numerous methods leveraging deep learning have emerged for detecting image forgery [28–32] and copy-move forgery localization (CMFL) [33–38]. However, existing traditional methods and the deep learning methods described above can only detect and locate copy-move forgery, and cannot with the realize of copy-move source/target distinguishment (CMSTD) [39]. Wu et al. [34] first proposed BusterNet network to distinguish source regions and target regions in copy-move forgery images. However, BusterNet target region detection branch feature is not fully used, resulting in the final location is not accurate. In order to improve BusterNet, Chen et al. [40] proposed a series network consisting of two subnets, namely copy-move similarity detection network (CMSDNet) and source/target region distinguishment network (STRDNet). Islam et al. [41] proposed a dual-order attention generative Adversarial network (DOA-GAN) for CMSTD. Zhang et al. [42] proposed a generative adversarial network combining convolutional neural network and transformer, and introduced Transformer into the detection of copy-move forgery for the first time. The above few methods use the tampered region as the auxiliary to assist similar region detection, and in these few methods, there is still a problem of unreasonable use of auxiliary information, because simple fusion of target region and the tampered region cannot play a complementary role in optimizing each other. In addition, the detection of the tampered region as the auxiliary is still insufficient. BusterNet only extracts tampered region by convolution and does not take advantage of the significant particularity of the tamper field edge, that is, the tamper edge is obviously different from the surrounding background. Chen et al. [43] proposed the MVSS method for detecting general tampering, which used the edge supervision scheme to supervise the tampering region detection. In order to prevent the loss of the edge features hidden in the shallow layer when detecting tampered edges, Chen et al. proposed a cascade structure, adding the edge features of the previous layer to the next layer. Although this method makes use of shallow information, it is prone to lead to redundancy and drown the effective information.

Aiming at the above problems, we propose a Copy-move Detection Method based on Decoupled Edge Supervision and Multi-domain Cross Correlation. Aiming at obtaining the reasonable features of the auxiliary region, we propose a DEM module by decoupling tampered edge layer-by-layer to supervise the detection of the tamper region. In addition, the MCM module is proposed to further detect similar regions by leveraging the tampered region

as well as the source region through cross-correlation global modeling. The contributions of this paper are as follows:

- We propose a DMNet which optimizes the coarse similar region by using the tampered region and source region as auxiliary information simultaneously through correlation modeling.
- We propose the DEM module by decoupling edge layer-by-layer to supervise the detection of the tampered region to avoid redundant shallow layer edge information which can obtain the tampered region with complete edge.
- We demonstrate the effectiveness of each module and the advancement of the whole network compared with the state-of-art on the CASIA CMFD and CoMoFoD datasets respectively.

The rest of this article is arranged as follows. Section 2 introduces the related work of copy-move image forgery detection. Section 3 provides detailed information about DMNet. The corresponding experimental results and analyses are described in Section 4. Finally, we discuss the contributions of this paper and conclude in Section 5.

2 Related work

2.1 CMFL and CMSTD

Copy-move forgery is a commonly used and easily implemented image tampering method [44, 45]. Figure 1 shows two examples of copy-move forgery. The first row is a part of the region that replicates the rubble, and the second row is the car in the parking lot. The traditional copy-move image forgery detection method can be divided into three parts: 1) Feature extraction [46, 47], extracting feature information at pixel level; 2) Information matching, searching for the most relevant matching method between feature information; 3) Post-processing operations to reduce false positives. Based on feature extraction and subsequent matching schemes, these methods can be roughly divided into two categories, one is block-based method and the other is keypoint-based method. In block-based methods, a variety of features used to describe overlapping fast and dense matching, such as DCT(discrete cosine transform) [2], DWT(discrete wavelet transform), and KPCA (Kernel Principal Component Analysis) [5], Zernike moments [7], PCT (Polar Cosine Transform) [8, 9], PCET (Polar Complex Exponential Transform) [48], LBP (Local Binary Patterns) [49], Circular Harmonic

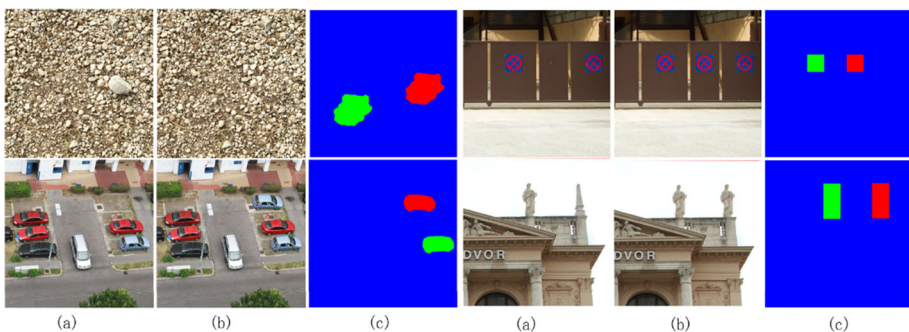


Fig. 1 Here, we give an example of image forgery detection, where image (a) is the original image, image (b) is the image after forgery, image (c) is the detection mask, where the blue region is the background part, the green region is the forgery source region, and the red region is the forgery target region

Transforms (CHT) [50]. In keypoint-based methods, the commonly used features are SIFT (Scale Invariant Feature Transform) [13, 14, 16, 18, 19, 51, 52] and SURF (Speeded-Up Robust Features) [20, 22]. Using the extracted features, the researchers investigated various feature selection and matching methods to improve recall rates. Although copy-move image tampering detection has made great progress, it is still a very challenging task to detect and locate forgery regions, which may be small or undergo complex post-processing operations such as rotation, compression, or noise addition. In addition, all of the traditional copy-move image forgery detection methods mentioned above rely on hand-crafted features, and each module is independently optimized, which makes detection less accurate and slower.

In recent years, copy-move image tamper detection based on end-to-end deep learning has attracted wide attention. Wu et al. proposed an end-to-end deep neural network to predict copy-move forgery masks. They first constructed convolutional neural networks (CNNs) for feature extraction, then calculated the correlation of features by self-correlation, and finally forged masks by deconvolution reconstruction. This method only extracts features at single scale, and does not extract features at multiple scales. A multi-scale feature extraction method is used, in [44], a network AR-Net with adaptive attention and residual refinement is proposed to improve the accuracy of the judgment results by refining the prediction mask of the network. Although this method can improve the accuracy of the network judgment, the residual module only plays an auxiliary role, and the network prediction is not accurate, and the refinement has a very weak effect on the performance improvement. AR-Net only uses series structure, in [45], a two-stage image forgery detection network is proposed. One branch extracts features and calculates correlations to get preliminary prediction results. The other branch scores features by invoking many existing image judgment models and locates the two regions with the highest similarity to the forgery region. Although the method adopts a double-branch structure, most of the judgment methods are based on the existing methods, and misjudgment will occur if there is no forged region in the image but two regions are similar. CMFL attempts to determine whether there is a forged region in the queried image and localize the forged region. However, these methods can only detect similar regions in the image, and cannot distinguish the source region and the target region, as shown in the third column of Fig. 1.

The purpose of CMSTD [53] is to obtain the masks of the source region (shown in green Fig. 1), target region (shown in red Fig. 1), and background based on the input image, as shown in column c in Fig. 1. At present, the mainstream method [35, 36] is to use convolutional neural network to process the image. BusterNet, proposed by Wu et al. [54], is an end-to-end model that can locate and distinguish source region and target region. In order to solve the problem of inadequate feature fusion of BusterNet parallel structure, Chen et al. [40] proposed a series structure copy-move image forgery detection network, which connected the copy-move similarity localization network with the source region and target region differentiation network in series to solve the problem of unreasonable and inadequate feature fusion of parallel network. Although the series structure increases the fusion of feature information, there is still a certain gap between the series structure and GAN network in distinguishing the source region and the target region. In [41], introduce a dual-order attentive generative adversarial network, where the generator incorporates both first-order and second-order attention mechanisms. The first-order attention is specifically crafted to capture information related to copy-move locations, while the second-order attention leverages more discriminative features pertaining to patch co-occurrence. Although DOA-GAN can distinguish between the forged source region and the target region, the judgment error will still occur when the gap between the source region and the target region is large or when there are many target regions. Although the above methods have achieved certain results in

CMFD tasks, the following problems still exist: 1) Only similar features and tamper region features are used, and source region features are not used. 2) The features of similar regions and tampered regions are only simple and crude splicing, which cannot give full play to the value of feature information. Therefore, we increase the use of source region features in the feature fusion stage, and feature fusion is not a simple splicing, but a correlation calculation, and feature selection at the pixel level through the gating mechanism.

2.2 Auxiliary branch

In most of the existing copy-move image forgery detection methods, a two-branch parallel network structure is adopted. One main branch conducts tamper detection and location, and the other branch acts as a helper to accurately locate the tamper region by collecting information about the tamper region, thus helping the main branch better locate the forgery region. The auxiliary branch can be considered as a special segmentation network whose function is to segment the input image. For input images, VGG or ResNet networks are generally used for initial feature extraction, and then segmentation features of the tamper region are obtained according to noise features or other post-processing methods. In [45], Liu et al. also adopted a two-branch network architecture, firstly obtaining the similar features of the tampered images, and then gradually determining the copy-move image forgery region with the help of auxiliary branches. Although the method can judge the forgery region, when there are multiple forgery regions or the gap between the source region and the target region is large, the judgment error will still occur. BusterNet proposed by Wu et al. [54] also adopted auxiliary branches, but the use of auxiliary branches was too simple, and only simple concatenation was used when fusing with similar features, and the auxiliary branches only used the features of the forged region, without using the information of the source region. Since copy-move image forgery is to move one part of the image to another part of the image, the derived feature and the target region feature have great similarity. Obtaining the characteristics of the tamper region through the auxiliary branch can greatly improve the location of the main branch to the forgery region, and even play a great role in distinguishing the source region from the target region. Therefore, our network structure adopts parallel structure and adds auxiliary branches. It not only refines the auxiliary branch feature fusion method, but also increases the use of source region information.

2.3 Edge supervision

Since the existence of falsified region boundary artifacts, it provides a necessary basis for the falsified region detection [55]. In the field of image segmentation [56], Hu et al. [57] proposed a boundary-sensing segmentation method based on graph convolution. On the basis of the existing types of segmentation, this method adds the category of segmentation object edge. The edge is used to continuously refine the details of the segmented object, so as to achieve the purpose of accurate segmentation. Edge supervision plays an important role in the field of forged image detection and image segmentation, which can improve the accuracy and precision of object segmentation, and the effect is more obvious when the object edge is complex. In [43], Chen et al. proposed an edge supervision method, which put the feature information into ResNet, and put the output of different ResNet blocks into Sobel and edge residual modules, and then add layer by layer to finally output edge information.

Edge supervision plays an important role in the field of forged image detection and image segmentation, which can improve the accuracy and precision of object segmentation, and the effect is more obvious when the object edge is complex. Although this method can obtain edge features, this layer-by-layer accumulation method will cause a lot of information redundancy, which will not only increase the amount of computation, but also make useful information submerged in a lot of useless information, which is difficult to obtain. Therefore, we propose the Decoupled Edge Supervision module (DEM), it not only uses edge supervision, but also reduces the low-level features contained in the high-level features by subtracting the low-level features, thus reducing the information redundancy.

3 Proposed method

In the field of copy-move image forgery detection, in order to improve the accuracy of similar branch detection, auxiliary branches are introduced to locate the tamper region. However, in tamper region localization, with the increase of network depth, a lot of information redundancy will be caused, which will drown out useful features. At the same time, after the feature extraction of tampered region is directly spliced with similar branch features, the value of auxiliary branches cannot be fully utilized. In order to solve the above problems, we proposed DMNet to locate the tamper source region and target region at the pixel level, and the overall network diagram is shown in Fig. 2.

Firstly, MSD generates the similar features of the source region and the target region, and then DEM obtains the features of the target region, and obtains the source region features by subtracting the similar features and the target region features, so as to solve the problem of difficult feature acquisition. Then, the obtained similar features, target region features, and source region features are passed through MCM to get the final prediction graph. We will go into detail about each module next.

3.1 Multi-scale similar region detection module

3.1.1 Feature extractor

Due to the significant difference in scale between the source and target regions in copy move forged images, it will be difficult to obtain accurate pixel level results by directly detecting the correlation between the source and target regions. In order to solve this problem, we use three convolutional kernels of different sizes for feature extraction and use them to extract features at different scales, thus achieving multi-scale detection [58]. Atrous convolution, as outlined in [59] and [60], possesses the ability to generalize standard convolution. It allows for the adjustment of the filter's field-of-view and provides control over the resolution of convolutional features. Let $y(i_c, j_c)$ represent the output of atrous convolution applied to a 2D input signal $x(i_c, j_c)$. The computation of atrous convolution is expressed as follows:

$$y(i_c, j_c) = \sum_{k_1, k_2} w(k_1, k_2) \times x(i_c + r_{ac}k_1, j_c + r_{ac}k_2), \quad (1)$$

where $k_1, k_2 \in [-fl(\frac{K}{2}), fl(\frac{K}{2})]$ ($fl(\cdot)$ is a floor function), $w(k_1, k_2)$ denotes a $K \times K$ filter, atrous rate r_{ac} determines the stride with which we sample the input signal. In the fourth block of our basic architecture, atrous rate r_{ac} is set to 2. The extracted block module is shown in Fig. 3.

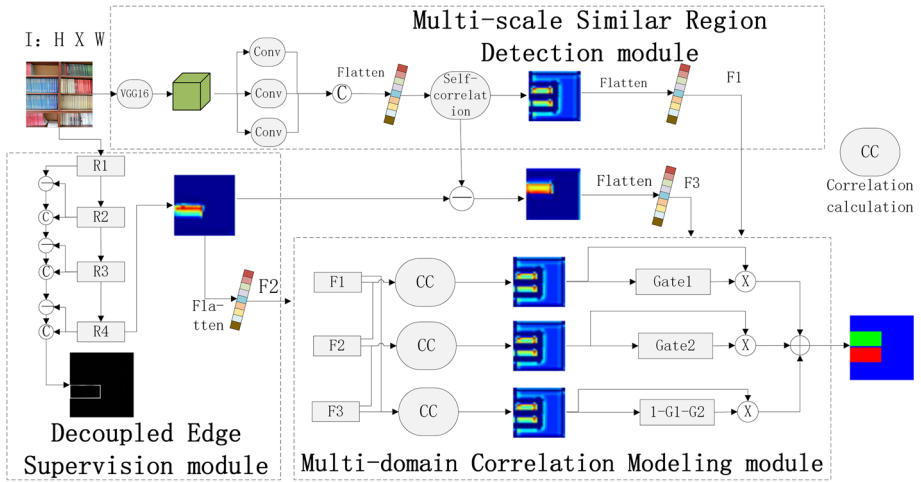


Fig. 2 The overall structure of DMNet is mainly composed of three parts:1)The MSD uses multi-scale feature extraction and self-correlation calculation to obtain the rough similar region of features;2)The DEM obtains the tamper domain features through edge supervision and feature decoupling;3)The MCM further optimizes the features through correlation calculation and gated selection mechanism. Among them, the source domain feature is obtained by the difference between MSD feature and DEM feature

3.1.2 Self-correlation

In this section, we will discuss self-correlation in detail. Let F_l denote the l -th block feature maps, and $F_l(i, j)$ denotes a c -dimensional descriptor at (i, j) . Note that $F_l \in \mathbb{R}^{H \times W \times C}$, $i \in [1, h]$, $j \in [1, w]$, h and w indicate the height and width of the feature maps, and $h = w$ in our work. Before the correlation computation, L2-normalization is conducted, $\hat{F}_l(i, j) = L2_norm F_l(i, j) = F_l(i, j) / \|F_l(i, j)\|_2$. Self-correlation aims to compute the similarity between every two locations in the convolutional feature maps. Scalar product is commonly used:

$$c_l^{(m,n)} = (\hat{F}_l^{(m)})^T \hat{F}_l^{(n)} \tag{2}$$

Thus, we can get a raw correlation map tensor $C_l = \{c_l^{(m,n)} | m, n \in [1, h \times w]\} \in R^{h \times w \times (h \times w)}$. In fact, only a small fraction of features has close relations, and the majority of features are dissimilar. Consequently, C_l is sorted along the $(h \times w)$ channels and top- T values are selected:

$$\hat{C}_l(i, j, 1 : T) = Top_T(Sort(C_l(i, j, :))) \tag{3}$$

Then, we get a correlation mapping tensor \bar{C} , which, since \bar{C} is computed from three layers of grouping features, contains a wealth of coarse-to-fine correlation information.

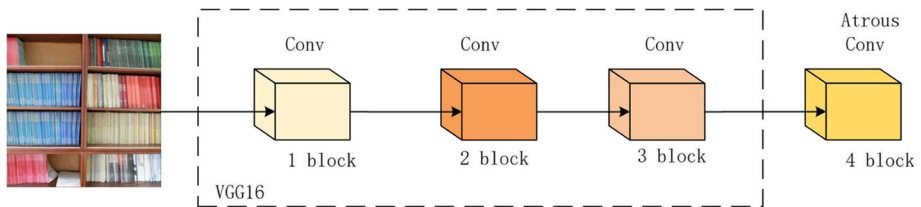


Fig. 3 Feature extraction diagram

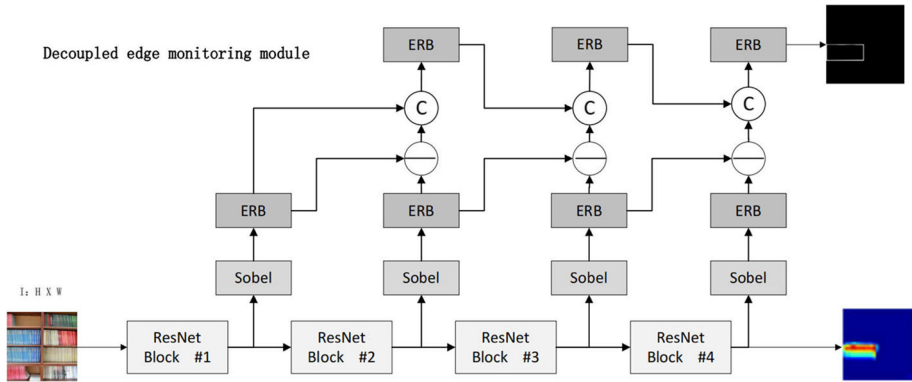


Fig. 4 Decoupled edge supervision module detailed diagram

3.2 Decoupled edge supervision module

Because the forged image is obtained by copying a part of the image, there will be a significant difference between the forged region in the image and the other regions, even after optimization, this problem still exists. With edge supervision, we want the response regions of the network to be more concentrated in the regions that have been tampered with. In the existing tamper region localization networks [61], there exists the phenomenon of feature redundancy, which leads to inaccurate tamper region localization [62]. Therefore, we use the decoupled edge supervision module to locate the tamper region more easily. In [61], Chen et al. processed the output of each layer of Resnet through Edge Residual Block(ERB), and then added layer by layer to process the bottom layer, so as to locate the edge. However, layer upon layer of feature information leads to a lot of information redundancy and useful information drowning, so we propose a simplified decoupled edge supervision module.

As shown in Figs. 4 and 5, the input image I enters the ResNet for feature extraction and tamper region localization. For the output features of each layer, the Sobel layer is first entered to enhance the edge-related pattern, and then the edge residual module (ERB) is entered. Since the deep feature will contain the shallow feature, after the deep ERB, we will carry out the difference processing with the features of the previous layer to eliminate redundancy, and the result of the difference processing will be merged with the shallow feature. This ensures that shallow features can be fully integrated and utilized without containing too much redundant information.

$$Block_{i=1,2,3,4} = ResNet(I), \tag{4}$$

$$R_j = ERB(S(Block_{j+1})) - ERB(S(Block_j)), \tag{5}$$

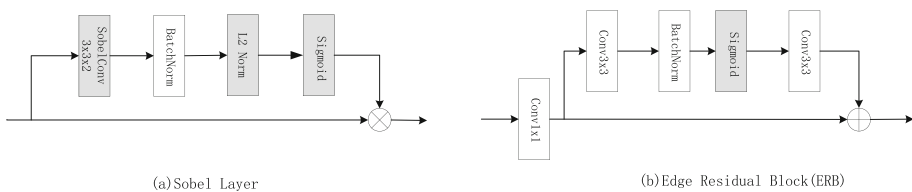


Fig. 5 Diagrams of (a) Sobel layer and (b) edge residual block, used in EDM for manipulation edge detection

$$P_j = ERB (Concat (R_j, ERB (S (Block_j)))) , \tag{6}$$

where Block represents the output of different layers of ResNet, S represents the Sobel layer, R represents the feature after eliminating the redundancy, and p represents the feature after the front and back layers are joined.

DEM output has two parts. One part is the feature mapping of the last ResNet block, expressed as $\{f_{dem,1}, \dots, f_{dem,k}\}$, used to predict the feature of the target domain, and the prediction edge map obtained by sigmoid layer transformation of the output of the last ERB is recorded as follows $\{G_{edge}(I_i)\}$.

3.3 Multi-domain correlation modeling module

In order to further refine the MSD module and detect the refined source region target region, we propose a Multi-domain Correlation Modeling module(MCM). After obtaining similar features and target region features, the conventional network directly concatenates to get the final prediction result. Unlike other networks, we not only use tamper region features, we also increase the utilization of source region features.

We use similar features $F1$, source region features $F3$ and target region features $F2$ for pairwise correlation calculation, three similar features $Z1, Z2$ and $Z3$ are obtained respectively, so that the features we get can be fully and reasonably used. For details, see Formula 7.

$$Z1 = Corr (F1, F2) , Z2 = Corr (F1, F3) , Z3 = Corr (F2, F3) , \tag{7}$$

$Corr$ stands for correlation calculation. After the fusion, the features $Z1, Z2$ and $Z3$ are filtered through the gating selection mechanism, and the filtered features are fused with the original features to obtain $M1, M2$ and $M3$ respectively. Then add $M1, M2, M3$ into the activation function to get the final result. Refer to the MCM section in Fig. 2.

$$M1 = Gate1 (Z1) \otimes Z1 \tag{8}$$

$$M2 = Gate2 (Z2) \otimes Z2 \tag{9}$$

$$M3 = (1 - Gate1 - Gate2) (Z3) \otimes Z3 \tag{10}$$

In this way, the rationality of feature selection can be ensured without the loss of important information. Among them, Gate1 and Gate2 are two learnable parameters. With the optimization of neural network, the fused feature information is optimized. The three gates cooperate with each other and share pixel weights in the same location. With the continuous updating of parameters in the gating mechanism, our feature selectivity will become better and better, and this will in turn adjust the front network, so that the performance of the entire network will continue to improve. The specific calculation method is shown in Algorithm 1.

3.4 Loss function

We use edge detection loss, source and forgery region detection loss to optimize the network parameters. The overall loss function is as follows:

$$L_{total} = \alpha L_{edg} + \beta L_{det} , \tag{11}$$

Algorithm 1 Multi-domain correlation modeling.

```

Input : Similar features  $F_{sim} \in R^{H \times W \times C}$ , Source feature  $F_{sou} \in R^{H \times W \times C}$ , Target feature
           $F_{tar} \in R^{H \times W \times C}$ 
Output: Multi-domain correlation features  $F_z \in R^{H \times W \times C}$ 
1 Function Correlation( $F_{sim}, F_{sou}, F_{tar}$ ):
2   setCorrelationCal(Feature1, Feature2)
   // Specifies a function for two-feature computation
3
    $V \leftarrow$  setCorrelationCal( $F_{sim}, F_{sou}$ )
    $M \leftarrow$  setCorrelationCal( $F_{sim}, F_{tar}$ )
    $W \leftarrow$  setCorrelationCal( $F_{tar}, F_{sou}$ )
   Gate(Features)
   // Gated operation
4
    $V_g \leftarrow$  Gate( $V$ )
    $W_g \leftarrow$  Gate( $W$ )
    $M_g \leftarrow$  Gate( $M$ )
5   for  $l = 1 : HW$  do
6      $F_v = VV_g$ 
      $F_v$  in length dimension
7   end
   //  $M$  and  $W$  operate in the same way as  $V$ 
8
    $F_z = F_v + F_w + F_m$ 
   return  $F_z$ 

```

where L_{edg} and L_{det} denote the edge loss and detection loss. The α and β is the learnable parameters. In our network, we use Dice Loss for edge supervision optimization parameters. The edge loss function is shown as follows:

$$L_{edg} = 1 - \frac{2 \times \sum_{i=1}^{W \times H} \hat{p}_i \times y_i}{\sum_{i=1}^{W \times H} \hat{p}_i^2 + \sum_{i=1}^{W \times H} y_i^2}, \quad (12)$$

where $y_i \in \{0, 1\}$ indicates whether the i -th pixel in the image is an edge pixel of the tampered region, \hat{p}_i denotes the probability of this pixel being predicted as the tampered edge. The spatial cross-entropy loss L_{det} is used as the loss of source region and target region. The spatial cross-entropy function is as follows:

$$L_{det} = -\frac{1}{H \times W} \sum_{i=1}^{H \times W} w_1 \times \hat{p}_i \times \log(p_i) + w_2 \times (1 - \hat{p}_i) \times \log(1 - p_i), \quad (13)$$

where $p_i \in \{0, 1\}$ denotes whether the i -th pixel belongs to the regions we aim to detect, \hat{p} represents the probability of this pixel is predicted correctly. w_1 and w_2 are the hyperparameter to balance the contribution of the two class pixels which are set to 0.8 and 0.2 respectively.

4 Experimental results and analysis

To evaluate the effectiveness of this method, we conducted a number of experiments. In this section, we first briefly describe the data sets used, evaluation metrics, and experimental

details. After that, we compared this method with other advanced methods, followed by some ablation studies to verify the effectiveness of each module.

4.1 Datasets

Because the number of publicly available copy-move forgery datasets is limited, about a few thousand, but it is impossible to distinguish the source region of the replicated move from the tampering region, Wu et al. created the USC-ISI CMFD dataset. We took the USC-ISI CMFD data set as the training set and divided it into the training set, the validation set and the test set according to the ratio of 8:1:1. At the same time, we also verify the generalization ability of our model on CASIA v2.0 and CoMoFoD [63] datasets. The details of these three datasets are shown in Table 1.

- **USC-ISI**: USCISI dataset contains 100k samples, each of which has a binary mask for copy-move forgery detection, and a three-class mask that can distinguish the source and target regions for copy-move source-target distinguishment. In our experiments, 80k, 10k, and 10k samples are randomly selected from the USCISI dataset for training, validation, and testing, respectively.
- **CASIA v2.0**: The CASIA v2.0 dataset contains 7491 true samples and 5123 falsified samples. However, the pictures are not only obtained by means of copy-move, but also by means of slicing, removing, etc., and some pictures have no ground truth. Therefore, 1313 copy-move samples were selected according to the Busternet method for the experiment.
- **CoMoFoD**: The CoMoFoD dataset contains a total of 200 basic tampered images, each of which provides 25 post-processing methods, resulting in 5,000 forged images.

4.2 Evaluation metrics

In order to evaluate the effectiveness of our network, we use precision, recall and F1-score as our evaluation metrics to evaluate the performance of our network at the pixel level. The formulas for precision and recall are as follows:

$$Precision = \frac{T_p}{T_p + F_p}, \quad (14)$$

$$Recall = \frac{T_p}{T_p + F_n}, \quad (15)$$

T_p indicates the number of correctly detected pixels, which is divided into three parts: source region, forged region, and background region. F_p indicates that the predicted result of the

Table 1 Details of three forgery datasets

Datasets	Copy-move images	Transformation	Train or test
USC-ISI	100k	Rotation,scaling, translation	Train and test
CASIA v2.0	1313	Rotation,deformation, resize	Test
CoMoFoD	200	Translation,rotation,scaling, distortion,combination	Test

network is a positive sample, and the actual sample is a negative sample, that is, the number of negative samples of false positives. FP indicates that the predicted result of the network is negative sample, but the actual sample is positive, that is, the number of positive samples missed.

Therefore, Precision represents the accuracy of predicting correct positive samples, and Recall represents the coverage rate of predicting correct positive samples. Finally, F1-score was used to conduct a comprehensive evaluation of Precision and Recall. The formula for the F1-score is as follows:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (16)$$

Accuracy is a metric used to evaluate a classification model. Specifically, the proportion of the total quantity that the model predicts is correct, where we take the average of the pristine, source, and target.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

In our data set, pristine, source, and target in an image can all be taken separately, and if we want to measure pristine, it becomes a binary problem of pristine and other. Our model can perform three classification judgments of RGB. We use R to represent target, G to represent source, and B to represent pristine. We take the maximum value of a pixel on the three channels of RGB and consider the pixel to be judged as target, source or pristine. In this way, each pixel in an image can be uniquely classified as pristine, source, and target, and then compared with the mask of pristine, source, and target respectively, pristine, source, and target can be used to calculate TP, FP, FN, TN on the whole picture in turn. Then the precision, recall and fscore indexes of pristine, source, and target are obtained respectively.

4.3 Implementation details

We utilized PyTorch deep learning framework to structure and train DMNet, and use PyTorch default function to initialize network parameters. We used Adam as the optimizer and set the learning rate to $1e-4$ and we set batch size to 16 for training optimization. In order to reduce network parameters and accelerate network optimization, we resize the image size in the data set to 256×256 for network training. All experiments were performed on a single 16GB Tesla V100, 2.60GHz Intel(R) Xeon(R) Gold 6126 CPU, and 187 GB RAM.

4.4 Contrast methods

We chose to evaluate the superiority of our approach by comparing it with some typical traditional methods and deep learning methods. The traditional methods including the segmentation-based image copy-move forgery detection scheme (Seg-based) [52], the dense field (DenseFiled) [50], and the adaptive segmentation approach (Adaptive-Seg) [51].

In addition, we compare several of the most classic deep learning methods in the field, including BusterNet [54], DOA-GAN [41], and others.

BusterNet [54]: BusterNet is a two-branch parallel network, with one main branch for feature extraction and location, and the other branch as an auxiliary to help the main branch locate more accurately. And the network can distinguish between forgery source region and tampering region.

DOA-GAN [41]: DOA-GAN adopts the mode of generating adversarial network and uses dual-order attention scheme to extract similar regions. And DOA-GAN can also distinguish between forged source region and tampered region.

Multi-branch CMSTD [64]: Multi-branch CMSTD obtains a pair of similar regions through feature extraction of the conventional network, and distinguishes the tampering source region from the tampering target region by detecting the boundary artifacts around the tampering region.

CNN-T GAN [42]: CNN-T GAN is a generative adversarial network. The generator of this method uses a combination of CNN and Transformer to distinguish the forged source region and tampered region.

4.5 Comparison with state-of-the-art methods

To evaluate DMNet's superiority in the field of copy-move tamper detection, we compared it to the most advanced tamper detection networks and conducted experiments on different datasets.

In the field of copy-move image forgery detection, most of the methods can only detect the forged region, and cannot distinguish the source region from the forged region. Therefore, the methods that can compare the background of the source region and the target region are limited, which also explains the challenge of this research. As shown in Table 2, we evaluated the precision, recall, and F-score of different methods for source region, tamper region, and background on USC-ISI, CASIA v2.0, and CoMoFoD datasets respectively. On the USC-ISI dataset, we made comparisons with BueterNet, CMSD STRD, DOA-GAN, Multi-branch CMSTD [64], and CNN-T GAN [42]. On CASIA v2.0 and CoMoFoD datasets, we compared with BueterNet, CMSD STRD, DOA-GAN, Multi-branch CMSTD.

On the USC-ISI dataset, the experimental results show that our DMNet has the best performance. In terms of background judgment, whether Precision, Recall or F1-score, our method is superior to all other methods. It can be seen that the gap between BueterNet and DMNet in background judgment is not particularly large, but in the judgment of source region and target region, the gap between BueterNet and DMNet is very obvious. This is because BueterNet simply adds the features of the target region and similar regions, and the fusion method is too simple. At the same time, when using auxiliary branches, the features of the source region are not used, but only the features of the target region are used. However, our network has also improved on sources and targets of varying magnitude. On the CASIA v2.0 and CoMoFoD datasets, our method comprehensively outperforms other compared methods. Especially for the judgment of source domain, the accuracy of Precision and F1-score improved by 3.48% and 5.47% respectively. The effect is also evident on the CoMoFoD dataset, which can be detailed in Table 2. Pristine scores are high and pristine scores are low. It is relatively easy to detect similar and background areas, but it is difficult to distinguish source and target, because there is no discriminative clue between source and target. That's why pristine scores are high and source and target scores are low. It is for this reason that we added DEM to help us detect the target part, which also facilitates the distinction between source and target, and finally makes our method more effective than other detection methods.

Figure 6. shows the visualized results. We can see that the existing methods are not good at locating the target region, especially the edge processing can not be well segmented and there will be misjudgment, which leads to poor effect of the source region and the target region. DMNet is closest to the real mask and has better structural integrity than other networks.

Table 2 Results of precision, recall, F1-score, accuracy(%) on pristine, source, and target regions of contrast approaches on three different datasets

Datasets	Methods	Precision		Recall		F1-score		Accuracy			
		Pristine	Source	Target	Pristine	Source	Target	Pristine	Source	Target	
USC-ISI	BusterNet	91.01	23.83	10.19	98.72	11.44	2.91	94.51	12.84	3.99	93.04
	CMSTD_STRD	97.02	50.68	26.24	97.84	45.48	43.32	97.37	45.35	60.40	95.03
	DOA-GAN	96.68	68.65	74.64	98.33	58.89	78.65	97.46	60.29	74.79	97.04
	Multi-branch CMSTD	95.55	31.84	24.58	90.50	43.57	30.69	92.64	33.62	25.41	90.39
	CNN-TGAN	98.71	86.12	87.10	99.14	94.37	98.92	84.45	86.36	98.66	98.66
CASIA v2.0	DMNet	98.77	80.01	87.11	99.90	94.46	85.97	98.33	76.43	83.25	98.87
	BusterNet	94.45	20.93	6.36	98.01	22.51	1.64	96.00	18.58	2.14	94.21
	CMSTD_STRD	96.43	31.88	22.86	97.39	36.74	26.91	96.75	30.86	22.71	94.09
	DOA-GAN	91.02	9.86	12.52	99.56	4.76	6.16	94.76	5.30	6.92	93.76
	Multi-branch CMSTD	92.50	14.58	12.32	88.03	24.93	10.29	89.36	14.07	9.12	87.63
CoMoFoD	DMNet	97.21	35.86	31.68	99.67	56.11	48.21	96.98	36.33	30.17	97.10
	BusterNet	96.92	15.88	5.81	94.87	25.03	1.64	95.64	14.38	2.18	94.42
	CMSTD_STRD	97.95	32.05	27.52	98.35	29.53	29.63	98.13	28.15	26.58	96.35
	DOA-GAN	94.22	15.47	12.60	99.63	6.09	4.71	96.69	7.25	5.63	95.71
	Multi-branch CMSTD	94.84	17.99	20.21	92.03	15.30	5.65	92.84	9.56	7.72	91.04
DMNet	98.11	33.62	28.26	99.65	48.39	49.83	98.18	28.33	26.98	97.28	

Pristine, source, and target represent the background(remove the source region and the rest of the target region from the image), source region(forgery source region), and target region(forgery target region) respectively. For example, the blue(background), green(source region), and red(target region) areas in Fig. 1

The bold text indicate the maximum value of an indicator in the same data set

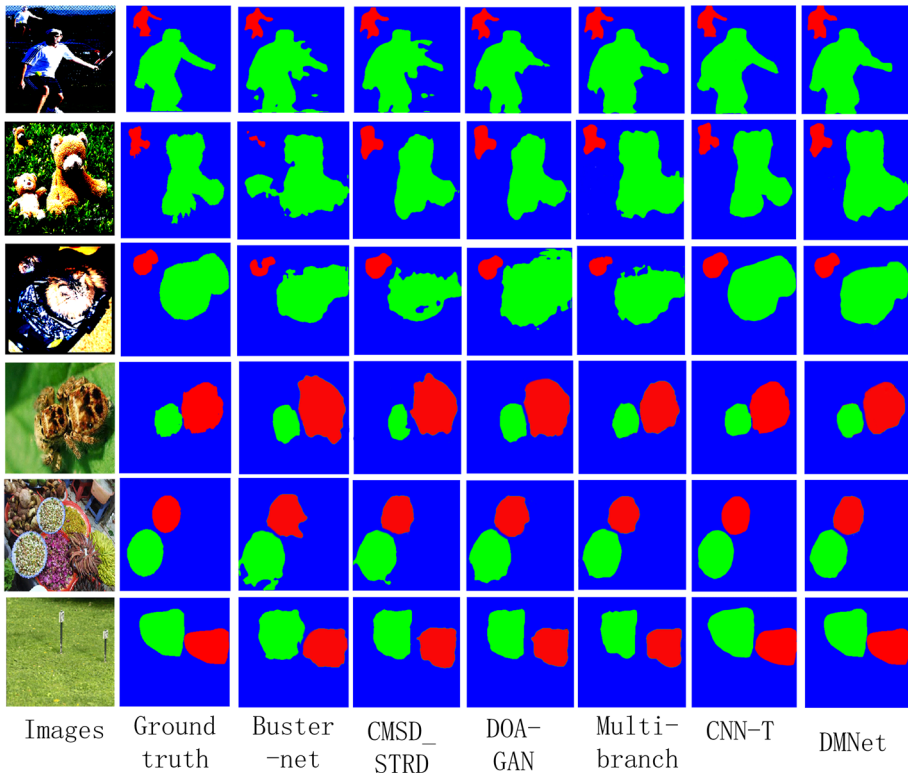


Fig. 6 The first column is the image after copy-move forgery, and the second column is the Ground Truth of the mask detection, where red is the forgery region, green is the source region, and blue is the background region. 3 to 6 are listed for comparison of other method detection results, and the last column is our network DMNet

After a large number of experiments, the overall prediction time of the model is maintained at about 8 seconds.

4.6 Module ablation experiments

In order to verify the role of each module, we conducted ablation experiments on three datasets respectively and presented the data in Table 3. The experimental design is as follows:

- **Single-scale:** After the initial feature extraction of the input image is carried out using the VGG network, the further feature extraction is carried out using the atrous convolution of a single convolution kernel. The network's prediction of the source region, target region and background of the forged image is then output by calculating the features through self-correlation, and CELoss is added at the end.
- **Multi-scale:** After using the VGG network for initial feature extraction of the input image, further feature extraction is carried out using three different sized convolutional kernels of the atrous convolution. Then, through self-correlation calculation of features, the network predicts the source region, target region, and background of the forged image, and adds CELoss at the end.

Table 3 Results of precision, recall, F1-score (%) on pristine, source, and target regions of ablation experiments on three different datasets

Datasets	Methods	Precision		Recall		F1-score		
		Pristine	Source	Pristine	Source	Pristine	Source	
USC-ISI	Single-scale	98.01	76.99	93.69	80.73	83.43	72.77	79.78
	Multi-scale	98.26	77.36	84.11	81.26	83.89	73.38	80.18
	Edge-MVSS	98.58	77.62	85.52	82.34	84.21	74.62	81.68
	MAE	98.62	78.33	85.98	82.86	84.99	75.39	81.99
	NOS	98.71	79.55	86.37	93.45	85.72	76.12	82.46
CASIA v2.0	DMNet	98.77	80.01	87.11	94.46	85.97	76.43	83.25
	Single-scale	96.78	31.46	26.85	53.96	46.07	33.92	22.62
	Multi-scale	96.93	32.28	27.25	54.21	46.83	34.15	23.34
	Edge-MVSS	96.97	33.68	28.46	54.67	47.23	35.02	24.32
	MAE	97.05	34.33	29.51	55.01	47.89	35.86	25.11
CoMoFoD	NOS	97.16	34.87	30.86	55.92	48.09	36.17	25.84
	DMNet	97.21	35.36	31.68	56.11	48.21	36.33	26.98
	Single-scale	97.80	26.35	24.23	44.55	46.53	25.48	22.62
	Multi-scale	97.91	27.46	25.63	45.26	47.07	26.25	23.34
	Edge-MVSS	97.78	28.86	26.74	46.34	48.31	27.86	24.32
NOS	MAE	98.05	29.73	27.06	47.85	48.92	28.15	25.11
	NOS	98.09	31.35	27.98	48.01	49.46	28.27	25.84
	DMNet	98.11	33.62	28.26	48.39	49.83	28.33	26.98

Pristine, source, and target represent the background(remove the source region and the rest of the target region from the image), source region(forgery source region), and target region(forgery target region) respectively. For example, the blue(background), green(source region), and red(target region) areas in Fig. 1

The bold text indicate the maximum value of an indicator in the same data set

- Edge-MVSS [43]: The edge features of auxiliary branches are fused directly with multi-scale similar features. The auxiliary branch does not eliminate redundant information. After the features pass through the ResNet Block, they enter Sobel layers and then pass through the edge residual block. After the features are added layer by layer, they enter an edge residual block. The detailed process is shown in Figs. 7 and 8.
- MAE : After eliminating the redundant information, the auxiliary branch is spliced with multi-scale features, and CELoss is added to the last output. After the features pass through the ResNet Block, they enter the Sobel layers, and then pass through the edge residual block. After the difference between the high-level features and the bottom features is made to eliminate the redundancy, and then they enter an edge residual block after splicing. The specific redundancy elimination operations are shown in Fig. 4.
- NOS : Without using the source region features, the correlation between the obtained multi-scale similar features and the target region features of the auxiliary branches is calculated, and the final prediction graph is output through a gating mechanism.
- DMNet : Multi-scale convolution kernel is used to extract similar features, and auxiliary branches are used to eliminate redundant information. In the use of auxiliary branch features, it is no longer a simple splicing operation, but also increases the use of source region features, and after the correlation calculation of features can be fully integrated, a gating mechanism is also added for feature selection.

The experimental results in Table 3 show that the accuracy of multi-scale judgment is higher than that of single-scale judgment, because after using convolution kernels of different sizes, we can make full use of image features, especially when there is a large gap between the source region and the target region. This also proves the correctness of our use of multiple scales. From Table 3, it can be seen that compared to Edge MVSS, network performance has been improved to varying degrees. This also indicates that our redundant elimination operation is effective, allowing important features to be extracted while removing useless information. After adding source domain feature information and MCM module, the perfor-

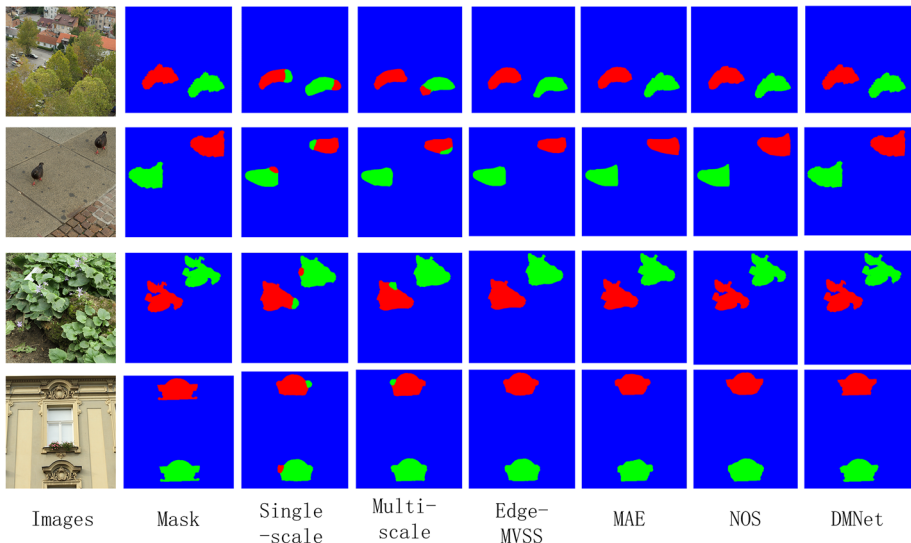


Fig. 7 The results of ablation experiment were visualized

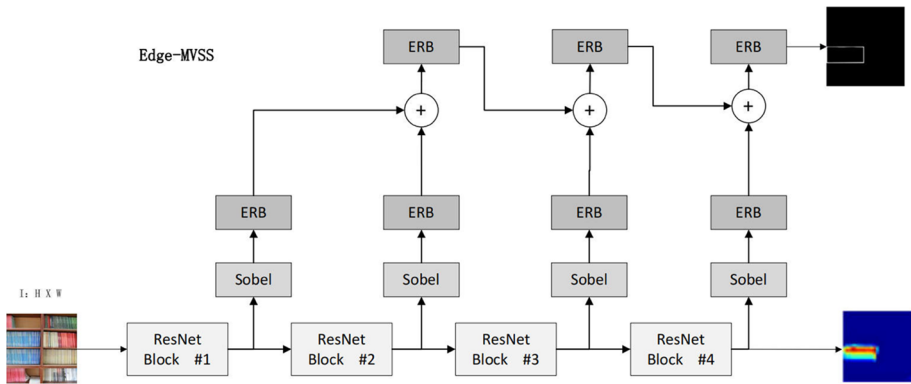


Fig. 8 Edges-MVSS structure flowchart

mance of network prediction is improved again, which fully demonstrates the importance of source domain feature and the effectiveness of the feature fusion method designed by us. The experimental results in Table 3 show that with the increase of modules, the network’s judgment of source domain, target domain and background in forged images becomes more and more accurate, and the network performance is constantly improved, which also proves that each module in our network plays an important role.

4.7 Robustness analysis

Under normal circumstances, in order to avoid detection, the forged image will go through a series of post-processing operations such as noise addition, color contrast change and so on. Therefore, it is particularly important to verify the robustness of our proposed network.

The CoMoFoD data set is a 512 x 512 copy-move image forgery detection data set of 5000 images obtained from 200 basic images by transforming, distorting, scaling, rotating, and combining in 5 different ways. To make CoMoFoD more challenging, researchers will use contrast adjustment (CA), JPEG compression (JC), noise adding (NA), brightness change (BC), color reproduction (CR), and image blurring (IB) to hide the traces of copy-move

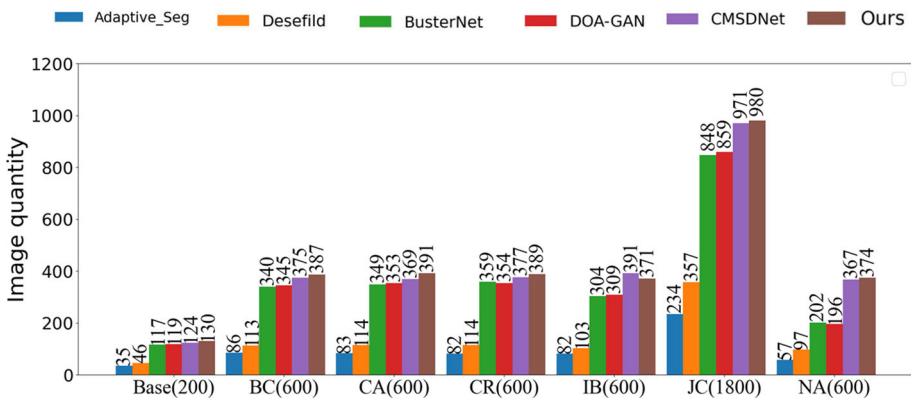


Fig. 9 The number of images under each attack correctly detected through DMNet and other comparison networks

image forgery [63]. As can be seen from Table 2, our method can still judge the forged source domain, target domain and background well under image transformation.

To evaluate the robustness of our model against different post-processing methods, Fig. 9 shows the number of correctly detected images on the CoMoFoD dataset for different attack types (if the image has a pixel-level F1-score greater than 0.5, we consider the image to be correctly detected). From the figure, we can see that our method is superior to all other methods except Image Blurring processing method, which strongly proves the robustness of our method. Figure 10 shows the F1-score of our network under different attacks, demonstrating the robustness of our model.

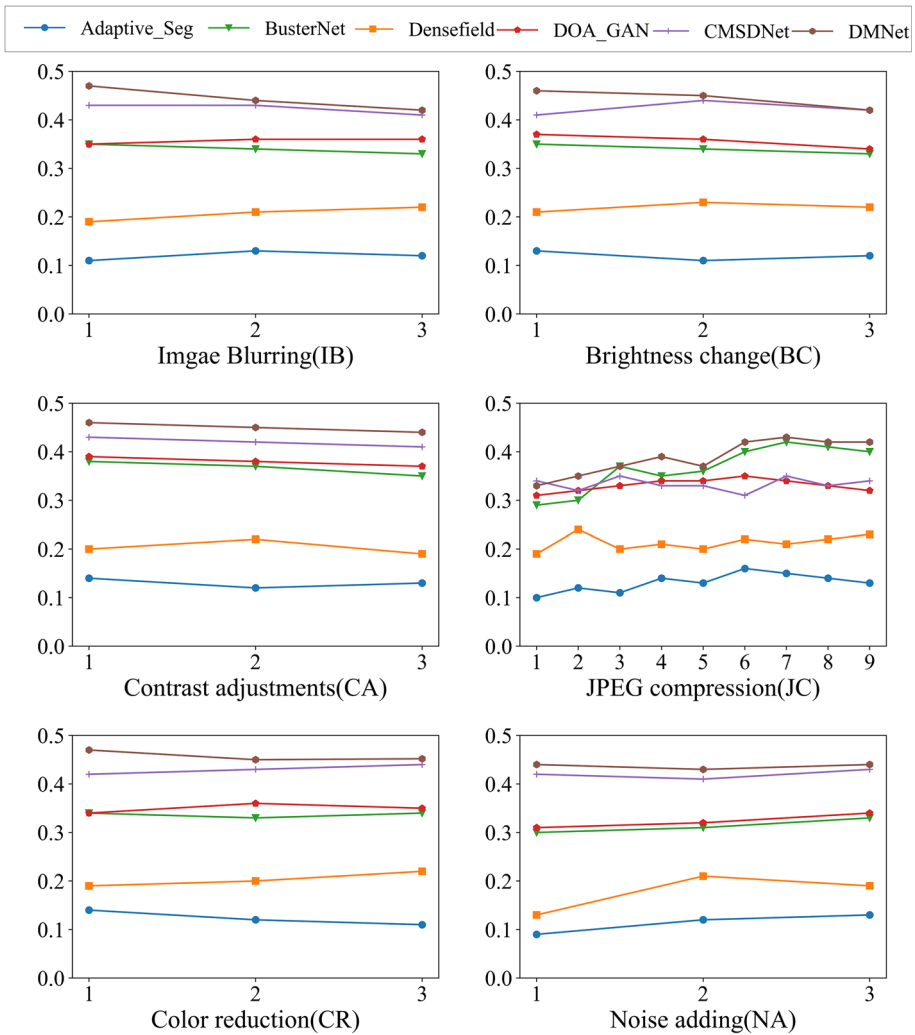


Fig. 10 Comparison of DMNet with other methods on pixel level F1-score(Y-axis) for different attacks (X-axis) on CoMoFoD dataset

5 Conclusion

In this paper, a copy-move image forgery detection method based on DMNet is proposed. In order to avoid information redundancy when using shallow edge features, we design an decoupled edge supervision module to optimize auxiliary network parameters while increasing edge positioning, improve the accuracy of tamper region positioning. To solve the problem that the value of auxiliary region features cannot be fully utilized, we propose the multi-domain correlation modeling module, which uses both tamper region features and source region features. After fusion of multi-scale similarity features, tamper region features and source region features, a gating mechanism is added to optimize parameters, and the detection targets of similar regions are further optimized by mining the correlation between multiple domains. Ablation experiments have verified the effectiveness of the designing of network architecture, the feature selection, and loss function. For feature extraction and utilization, the method of CNN is currently limited to the use of CNN, and the combination of CNN and Transformer may be tried in the next stage. For tampering edges, after multiple convolution operations, there will be a certain loss of information, which will be the next step we will improve. We performed experiments on USC-ISI data set, and the accuracy was improved by 0.21% compared with CNN-T GAN method, and the F1-score index was improved by 0.87% compared with DOA-GAN. The accuracy of CASIA v2.0 data set is 2.89% higher than that of Busternet method, and the precision index is 3.98% higher than that of CMSD-STRD method on source. The accuracy of CoMoFoD data set is improved by 0.93% compared with CMSD-STRD method, and the recall index is improved by 20.23% compared with CMSD-STRD target, which proves that our method is superior to the most advanced method at present. This fully proves the effectiveness of our proposed method and provides a new approach to CMSTD.

Acknowledgements This work is supported by Shandong Provincial Central Guiding Local Science and Technology Development Fund Project (YDZX2022028) and National Natural Science Foundation Regional Innovation and Development Fund (U22A2068).

Data Availability The dataset involved in this article is the same as the dataset used in BusterNet.

Declarations

Conflicts of interest The authors have no conflict of interest.

References

1. Muhammad G, Hussain M, Bebis G (2012) Passive copy move image forgery detection using undecimated dyadic wavelet transform. *Digit Investig* 9(1):49–57
2. Mahmood T, Nawaz T, Irtaza A, Ashraf R, Shah M, Mahmood MT et al (2016) Copy-move forgery detection technique for forensic analysis in digital images. *Math Prob Eng* 2016
3. Lowe DG (1999) Object recognition from local scale-invariant features. In: *Proceedings of the seventh IEEE international conference on computer vision*, vol 2, pp 1150–1157. Ieee
4. Gani G, Qadir F (2021) Copy move forgery detection using dct, patchmatch and cellular automata. *Multimed Tools App* 80:32219–32243
5. Bashar M, Noda K, Ohnishi N, Mori K (2010) Exploring duplicated regions in natural images. *IEEE Trans Image Process*
6. Ryu S-J, Lee M-J, Lee H-K (2010) Detection of copy-rotate-move forgery using zernike moments. In: *Information hiding: 12th international conference, IH 2010, Calgary, AB, Canada, June 28–30, 2010, Revised Selected Papers* 12, pp 51–65. Springer

7. Ryu S-J, Kirchner M, Lee M-J, Lee H-K (2013) Rotation invariant localization of duplicated image regions based on zernike moments. *IEEE Trans Inf Forensics Secur* 8(8):1355–1370
8. Li Y (2013) Image copy-move forgery detection based on polar cosine transform and approximate nearest neighbor searching. *Forensic Sci Int* 224(1–3):59–67
9. Yap P-T, Jiang X, Kot AC (2009) Two-dimensional polar harmonic transforms for invariant image representation. *IEEE Trans Pattern Anal Mach Intell* 32(7):1259–1270
10. Su L, Li C, Lai Y, Yang J (2017) A fast forgery detection algorithm based on exponential-fourier moments for video region duplication. *IEEE Trans Multimed* 20(4):825–840
11. Zhong J, Gan Y (2016) Detection of copy–move forgery using discrete analytical fourier–mellin transform. *Nonlinear Dyn* 84(1):189–202
12. Huang D-Y, Huang C-N, Hu W-C, Chou C-H (2017) Robustness of copy-move forgery detection under high jpeg compression artifacts. *Multimed Tools App* 76:1509–1530
13. Soni B, Das PK, Thounaojam DM (2018) multicmfd: fast and efficient system for multiple copy-move forgeries detection in image. In: *Proceedings of the 2018 international conference on image and graphics processing*, pp 53–58
14. Soni B, Das PK, Thounaojam DM (2018) Keypoints based enhanced multiple copy-move forgeries detection system using density-based spatial clustering of application with noise clustering algorithm. *IET Image Process* 12(11):2092–2099
15. Yang B, Sun X, Guo H, Xia Z, Chen X (2018) A copy-move forgery detection method based on cmfd-sift. *Multimed Tools App* 77:837–855
16. Amerini I, Ballan L, Caldelli R, Del Bimbo A, Serra G (2011) A sift-based forensic method for copy–move attack detection and transformation recovery. *IEEE Trans Inf Forensics Secur* 6(3):1099–1110
17. Costanzo A, Amerini I, Caldelli R, Barni M (2014) Forensic analysis of sift keypoint removal and injection. *IEEE Trans Inf Forensics Secur* 9(9):1450–1464
18. Pan X, Lyu S (2010) Region duplication detection using image feature matching. *IEEE Trans Inf Forensics Secur* 5(4):857–867
19. Li Y, Zhou J (2018) Fast and effective image copy-move forgery detection via hierarchical feature point matching. *IEEE Trans Inf Forensics Secur* 14(5):1307–1322
20. Silva E, Carvalho T, Ferreira A, Rocha A (2015) Going deeper into copy-move forgery detection: exploring image telltales via multi-scale analysis and voting processes. *J Vis Commun Image Represent* 29:16–32
21. Manu V, Mehre BM (2016) Detection of copy-move forgery in images using segmentation and surf. In: *Advances in signal processing and intelligent recognition systems: proceedings of second international symposium on signal processing and intelligent recognition systems (SIRS-2015) December 16–19, 2015, Trivandrum, India*, pp 645–654. Springer
22. Ardizzone E, Bruno A, Mazzola G (2015) Copy–move forgery detection by matching triangles of keypoints. *IEEE Trans Inf Forensics Secur* 10(10):2084–2094
23. Zhu Y, Shen X, Chen H (2016) Copy-move forgery detection based on scaled orb. *Multimed Tools App* 75:3221–3233
24. Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. In: *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, pp 234–241. Springer
25. Chen L-C, Papandreou G, Schroff F, Adam H (2017) Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*
26. Omar A, Abd El-Hafeez T (2024) Optimizing epileptic seizure recognition performance with feature scaling and dropout layers. *Neural Comput App* 36(6):2835–2852
27. Eliwa EHI, El Koshiry AM, Abd El-Hafeez T, Farghaly HM (2023) Utilizing convolutional neural networks to classify monkeypox skin lesions. *Sci Rep* 13(1):14495
28. Yang J, Xiao S, Li A, Lu W, Gao X, Li Y (2021) Msta-net: forgery detection by generating manipulation trace based on multi-scale self-texture attention. *IEEE Trans Circ Syst Vid Tech* 32(7):4854–4866
29. Zhuang P, Li H, Tan S, Li B, Huang J (2021) Image tampering localization using a dense fully convolutional network. *IEEE Trans Inf Forensics Secur* 16:2986–2999
30. Zhang Y, Zhu G, Wu L, Kwong S, Zhang H, Zhou Y (2021) Multi-task se-network for image splicing localization. *IEEE Trans Circ Syst Vid Tech* 32(7):4828–4840
31. Ding F, Zhu G, Alazab M, Li X, Yu K (2020) Deep-learning-empowered digital forensics for edge consumer electronics in 5g hetnets. *IEEE Consum Electro Mag* 11(2):42–50
32. Wu H, Zhou J (2021) Iid-net: Image inpainting detection network via neural architecture search and attention. *IEEE Trans Circ Syst Vid Tech* 32(3):1172–1185
33. Rao Y, Ni J (2016) A deep learning approach to detection of splicing and copy-move forgeries in images. In: *2016 IEEE International workshop on information forensics and security (WIFS)* pp 1–6. IEEE

34. Wu Y, Abd-Almageed W, Natarajan P (2018) Image copy-move forgery detection via an end-to-end deep neural network. In: 2018 IEEE Winter conference on applications of computer vision (WACV) pp 1907–1915. IEEE
35. 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/CVF conference on computer vision and pattern recognition pp 9543–9552
36. Zhong J-L, Pun C-M (2019) An end-to-end dense-inceptionnet for image copy-move forgery detection. *IEEE Trans Inf Forensics Secur* 15:2134–2146
37. Wang J, Gao X, Nie J, Wang X, Huang L, Nie W, Jiang M, Wei Z (2024) Strong robust copy-move forgery detection network based on layer-by-layer decoupling refinement. *Inf Process Manage* 61(3):103685
38. Wang J, Jing N, Liu Z, Nie J, Qi Y, Chi C-H, Lam K-Y (2024) Object-level copy-move forgery image detection based on inconsistency mining. arXiv preprint [arXiv:2404.00611](https://arxiv.org/abs/2404.00611)
39. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
40. Chen B, Tan W, Coatrieux G, Zheng Y, Shi Y-Q (2020) A serial image copy-move forgery localization scheme with source/target distinguishment. *IEEE Trans Multimed* 23:3506–3517
41. Islam A, Long C, Basharat A, Hoogs A (2020) Doa-gan: dual-order attentive generative adversarial network for image copy-move forgery detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition pp 4676–4685
42. Zhang Y, Zhu G, Wang X, Luo X, Zhou Y, Zhang H, Wu L (2022) Cnn-transformer based generative adversarial network for copy-move source/target distinguishment. *IEEE Trans Circ Syst Vid Tech*
43. Dong C, Chen X, Hu R, Cao J, Li X (2022) Mvss-net: multi-view multi-scale supervised networks for image manipulation detection. *IEEE Trans Pattern Anal Mach Intell* 45(3):3539–3553
44. Zhu Y, Chen C, Yan G, Guo Y, Dong Y (2020) Ar-net: adaptive attention and residual refinement network for copy-move forgery detection. *IEEE Trans Ind Inf* 16(10):6714–6723
45. Liu Y, Xia C, Zhu X, Xu S (2021) Two-stage copy-move forgery detection with self deep matching and proposal superglue. *IEEE Trans Image Process* 31:541–555
46. Saabia AA-B, El-Hafeez T, Zaki AM (2019) Face recognition based on grey wolf optimization for feature selection. In: Proceedings of the international conference on advanced intelligent systems and informatics 2018 4 pp 273–283. Springer
47. Abd El-Hafeez T (2010) A new system for extracting and detecting skin color regions from pdf documents. *Int J Comput Sci Eng (IJCSSE)* 9(2):2838–2846
48. Bi X, Pun C-M (2018) Fast copy-move forgery detection using local bidirectional coherency error refinement. *Pattern Recognit* 81:161–175
49. Li L, Li S, Zhu H, Chu S-C, Roddick JF, Pan J-S (2013) An efficient scheme for detecting copy-move forged images by local binary patterns. *J Inf Hiding Multim Signal Process* 4(1):46–56
50. Cozzolino D, Poggi G, Verdoliva L (2015) Efficient dense-field copy-move forgery detection. *IEEE Trans Inf Forensics Secur* 10(11):2284–2297
51. Pun C-M, Yuan X-C, Bi X-L (2015) Image forgery detection using adaptive oversegmentation and feature point matching. *IEEE Trans Inf Forensics Secur* 10(8):1705–1716
52. Li J, Li X, Yang B, Sun X (2014) Segmentation-based image copy-move forgery detection scheme. *IEEE Trans Inf Forensics Secur* 10(3):507–518
53. Chang S (2023) Can deep network balance copy-move forgery detection and distinguishment? arXiv preprint [arXiv:2305.10247](https://arxiv.org/abs/2305.10247)
54. 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
55. El-Sayed MA, Hafeez TA-E (2012) New edge detection technique based on the shannon entropy in gray level images. arXiv preprint [arXiv:1211.2502](https://arxiv.org/abs/1211.2502)
56. Eman M, Mahmoud TM, Ibrahim MM, Abd El-Hafeez T (2023) Innovative hybrid approach for masked face recognition using pretrained mask detection and segmentation robust pca and knn classifier. *Sensors* 23(15):6727
57. Hu H, Cui J, Zha H (2021) Boundary-aware graph convolution for semantic segmentation. In: 2020 25th International conference on pattern recognition (ICPR) pp 1828–1835. IEEE
58. Gan Y, Zhong J, Vong C (2022) A novel copy-move forgery detection algorithm via feature label matching and hierarchical segmentation filtering. *Inf Process Manage* 59(1):102783
59. Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the european conference on computer vision (ECCV) pp 801–818

60. Chen L-C, 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
61. Chen X, Dong C, Ji J, Cao J, Li X (2021) Image manipulation detection by multi-view multi-scale supervision. In: *Proceedings of the IEEE/CVF international conference on computer vision* pp 14185–14193
62. Wang W, Dong J, Tan T (2011) Tampered region localization of digital color images based on jpeg compression noise. In: *Digital watermarking: 9th international workshop IWDW 2010 Seoul Korea October 1-3 2010 Revised Selected Papers* 9 pp 120–133. Springer
63. Tralic D, Zupancic I, Grgic S, Grgic M (2013) Comofod—new database for copy-move forgery detection. In: *Proceedings ELMAR-2013* pp 49–54. IEEE
64. Barni M, Phan Q-T, Tondi B (2020) Copy move source-target disambiguation through multi-branch cnns. *IEEE Trans Inf Forensics Secur* 16:1825–1840

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Niantai Jing¹ · Jie Nie¹  · Jingyu Wang¹ · Xiaodong Wang¹ · Xinyue Liang¹ · Xuesong Gao²

Niantai Jing
jingniantai@stu.ouc.edu.cn

Jingyu Wang
wangjingyu3186@stu.ouc.edu.cn

Xiaodong Wang
wangxiaodong@ouc.edu.cn

Xinyue Liang
liangxinyue@ouc.edu.cn

Xuesong Gao
xuesong@outlook.com

¹ Faculty of Information Science and Engineering, Ocean University of China, QingDao, China

² Qingdao Hisense State Key Laboratory of Digital Multimedia Technology, QingDao, China