PRest-Net: Multi-domain Probability Estimation Network for Robust Image Forgery Detection

JIAXIN CHEN, College of Computer Science and Electronic Engineering, Hunan University, China, and School of Computer and Communication Engineering, Changsha University of Science and Technology, China
XIN LIAO*, College of Computer Science and Electronic Engineering, Hunan University, China
ZHENXING QIAN, School of Computer Science, Fudan University, China
ZHENG QIN, College of Computer Science and Electronic Engineering, Hunan University, China

As an important carrier of information transmission in online social networks (OSNs), the authenticity protection of images is of great significance. However, the abuse of image processing technology makes its security questionable. Meantime, lossy operations adopted by OSNs will change the forgery artifacts, which brings challenges to robust image forgery detection. To address this issue, considering the suppression of lossy noise caused by transmission, a novel multi-domain probability estimation network (PRest-Net) is proposed. Firstly, we design a multi-domain probability estimation method to capture the most differentiated regional information from the spatial, residual, and wavelet domains. Since the wavelet coefficient of semantic information is larger than that of lossy noise, and the edge texture can be highlighted in the residual image, the negative effect of lossy noise would be reduced and semantic forgery traces can be exposed more easily. We further design a forgery detector composed of low-level feature extraction, high-level feature extraction, and regional edge difference learning module, which can adaptively learn rich forgery clues. Extensive experimental results are provided to validate the superiority of PRest-Net compared with existing state-of-the-art detectors in the scenarios of detecting forged images transmitted over various OSNs.

CCS Concepts: • **Security and privacy** → *Social aspects of security and privacy*.

Additional Key Words and Phrases: Image security and forensics, Forgery detection, Online social network transmission

1 INTRODUCTION

The increasing popularity of online social networks (OSNs), such as Facebook, Whatsapp, Wechat, and Weibo, has made information transmission more convenient. Meanwhile, various image sensors, such as intelligent monitoring cameras, personal computers, and smart phones, are pervasively deployed in our lives, capturing a significant number of digital images daily. As shown in Fig. 1, images have become an essential carrier for acquiring and transmitting information, whose application fields are diverse, for instance, traffic detection, journalism and communication, forensic testimony, and qualification certificate. However, the development of image processing technology makes editing images easy [1], and it is difficult to observe the image forgery from

Corresponding author: Xin Liao.
Authors' Contact Information: Jiaxin Chen, jxchen@csust.edu.cn, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China, and School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China; Xin Liao, xinliao@hnu.edu.cn, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China; Zhenxing Qian, zxqian@fudan.edu.cn, School of Computer Science, Fudan University, Shanghai, China; Zheng Qin, zqin@hnu.edu.cn, College of Computer Science and Electronic Engineering, Hunan University, Changsha, China.

ACM 1551-6865/2025/1-ART
https://doi.org/10.1145/3711930

the visual senses. There is a growing concern about ensuring the authenticity of image transmitted over online social networks [2].

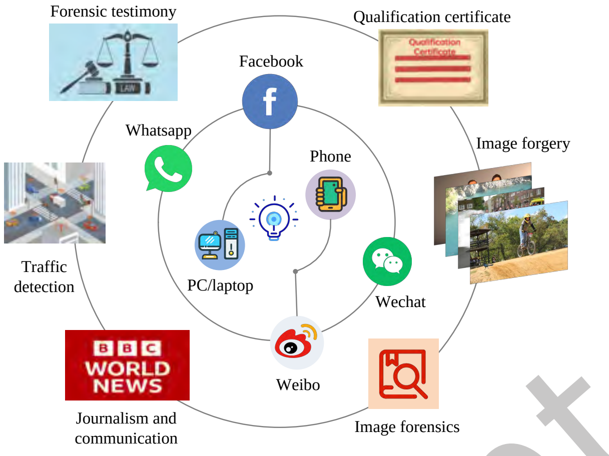

Fig. 1. Image information transmission and image forensics in various online social networks.

Previous work on forensics utilizes clues such as scale invariant feature transform (SIFT) [3], color inconsistency [4], affine transformation matrix [5], and noise inconsistency [6, 7] to determine if a specific image block has been forged and expose tampered regions [8–13]. Specifically, Amerini *et al.* [3] proposed a copy-move forgery detection method consisting of SIFT extraction, keypoint matching, and clustering. Because images with different sources have various noises introduced by the imaging equipment, splicing forgery can be exposed by detecting inconsistencies in local noise variances [7]. He *et al.* [13] proposed a deep cross-scale patchmatch framework by integrating merits from conventional and deep learning methods to detect copy-move forgery.

Most of these methods are designed to detect a specific tampering manipulation. However, actual image forgery usually involves multiple tampering operations [15, 16]. Some recently proposed methods [14, 17–20] show robustness to the localization of more general tampering manipulations. Zhang *et al.* [18] introduced a cross-layer intersection mechanism to dense u-net for image forgery localization, which adopted spatial rich model (SRM) filters to capture residual signals and used feature maps of the fully connected layer as decision information for image segmentation. Rao and Ni [19] proposed a self-supervised domain adaptation network that consists of a Siamese network and a compression approximation network for JPEG-resistant image forgery localization. In our previous work [20], the detection and localization of multiple tampering operations were achieved by separating the tampered regions from the complex real background regions and mining more boundary information with the help of the parallel atrous convolutional layers.

Nevertheless, images are often shared on online social networks and almost all OSNs will introduce several lossy operations, such as enhancement filtering, resizing, and JPEG compression [21], which would change the image content. As shown in Fig. 2, it can be found that the original forged image is different from the forged image transmitted through Facebook based on the residual image. These operations add lossy noise to the image, resulting in the loss of some high-frequency information. As a result, the image forgery detection performance would be degraded. Fig. 3 provide an example of detection for splicing and copy-move. The tampered regions from the original forgery can be accurately detected utilizing the state-of-the-art method [14], while the localization performance dropped when detecting the forged image transmitted through Facebook. Therefore, image forgery localization against online social network transmission remains an important challenge.

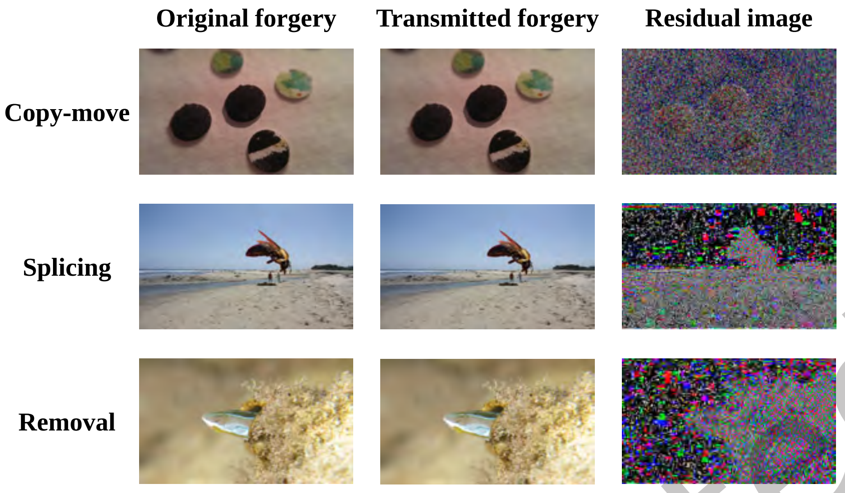

Fig. 2. Examples of the original forged image and forged image transmitted through Facebook.

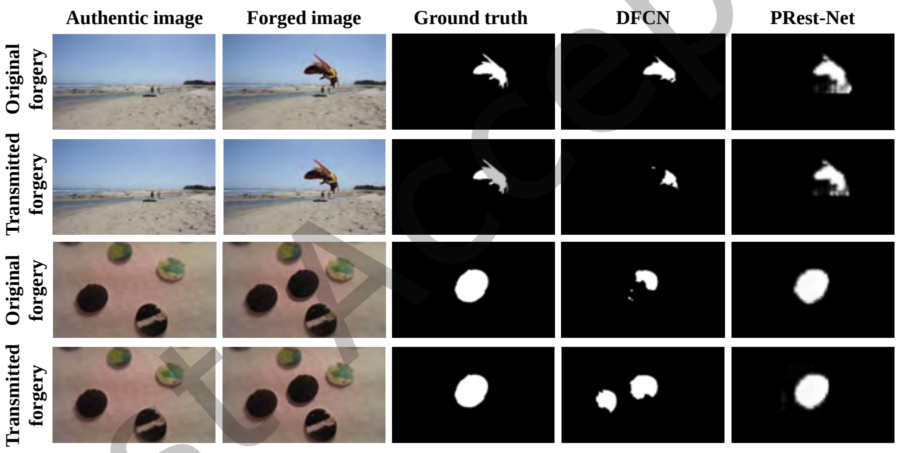

Fig. 3. The localization results of the original forged image and forged image transmitted through Facebook using **DFCN** [14] and our PRest-Net. The first two rows show images forged by splicing, and the last two rows show images forged by copy-move.

In this work, we propose a multi-domain probability estimation-based network PRest-Net to fight against the online social network shared forgery. Since the lossy operations of the online social network will introduce noise into the forged image, it is difficult to capture the traces of image forgery. From the perspective of reducing the negative influence of lossy noise on image forgery localization, we design a feature coupling-based probability estimation method that mines the most differentiated regional information from the spatial domain, residual domain, and wavelet domain, and estimates the tampering probability of the region through selective search. The fusion of multi-domain tampering probability maps would suppress lossy noise and express more complete and diversified information. As a result, the tampering artifacts will become easier to detect so that the problem of performance degradation caused by OSN transmission can be addressed. For accurately exposing the tampered regions, we further propose a forgery detector to limit the influence of complex image content on forgery localization and adaptively learn rich manipulation clues of the investigated image. The main contributions of this work can be summarized as follows:

1) We introduce multi-domain probability estimation into forgery detection of OSN-transmitted images, which brings a different viewpoint to robust forgery detection. Different from existing methods that focus on modeling lossy noise, PRest-Net analyzes the coupling relationship between multiple forensic features from different domains to suppress the lossy noise. Ultimately, the most differentiated regional information can be mined and the tampering probability of different image regions can be estimated, enabling coarse-grained forgery localization.
2) We propose a novel image forgery localization network robust to online social network transmission. PRest-Net integrates tampering probability maps estimated from multiple domains and utilizes a forgery detector based on the fusion of low-level forensic features, higher-level detection features, and more detailed regional edge differences to achieve fine-grained tampered regions prediction.
3) We conduct extensive experiments to verify the effectiveness of PRest-Net in forgery localization for online social network shared image data. Experimental results demonstrate that the proposed PRest-Net achieves better detection performance in comparison with several state-of-the-art methods.

The remainder of this paper is organized as follows. Section 2 describes related work. Section 3 illustrates the proposed multi-domain probability estimation network. Section 4 provides several experimental analyses, and the results show the effectiveness of the proposed PRest-Net. Section 5 gives the concluding remarks.

2 RELATED WORK

This section briefly reviews some related works about image forgery detection, which focus on detecting tampered regions. According to the principle of these forgery detection methods, they can be divided into statistical characteristics-based and deep learning-based approaches.

2.1 Detection based on Statistical Characteristics

Many image forgery detection methods have been proposed to expose the tampered regions through statistical characteristics analysis. A method [4] based on gray level co-occurrence matrix of thresholded edge image of image chroma was proposed for color image splicing detection. Li *et al.* [5] first segmented the testing image into semantically independent patches and then estimated an affine transform matrix to confirm the existence of copy-move forgery. Since the amount of noise in the entire authentic image is uniform, Mahdian *et al.* [6] introduced a segmentation method to detect the changes in noise level. If various noise levels are caught in a specific region, it means that the region has been forged. Peng *et al.* [12] try the first attempt to use the contact constraint of standing objects as a new clue for splicing detection. Qi *et al.* [22] studied the forensic-oriented image representation based on stable description with mathematical guarantees, and explored copy-move forgery detection and perceptual hashing. The hybrid method [23] that compares triangles rather than blocks or single points was used to detect copy-move forgery, which was robust to geometric transformations. To deal with the cases where copy-move forgeries only involve small or smooth regions, Li and Zhou [24] first generated a sufficient number of keypoints by lowering the contrast threshold and rescaling the given image, and then adopted a hierarchical matching strategy to address the keypoint matching problem. Finally, they obtained tampered regions by using the proposed iterative localization method.

2.2 Detection based on Deep Learning

Except for detecting specific tampering operations, several deep learning-based methods have been proposed to detect more general operations. Ye *et al.* [11] proposed a feature pyramid deep matching and localization network that exploited the correlation information between forensic features and integrated two pathways into just one simple pathway to localize small tampered regions. Rao *et al.* [25] designed a splicing and copy-move detection network to automatically learn hierarchical representations from RGB images, which initialized the weights at the first network layer with the basic high-pass filter set used in the calculation of residual maps in SRM. To suppress the scene content and enhance model-related artifacts, Cozzolino *et al.* [26] extracted a camera model fingerprint and applied a Siamese network to detect the tampered regions. In [27], a variational auto-encoder model based on vision transform was proposed to perform run-time learning of noise inconsistency, high-pass residual inconsistency, and edge discontinuity from the tampered images.

However, when completing an image forgery, a post-processing operation may be utilized to weaken the tampering artifacts of the image content. In this case, the forgery detection performance of most forensic methods will be degraded. By converting the forgery localization problem to a local anomaly detection problem, a fully convolutional network [17] was proposed to perform both detection and localization on many forgery types such as splicing, copy-move, and removal without additional processing. Besides, the network can learn robust image tampering traces from classifying 385 types of image operations and thus be robust to some post-processing operations. In [14], an encoder-decoder architecture based on dense connections and dilated convolutions can obtain excellent localization performance. Zhuo *et al.* [28] presented a self-adversarial training incorporating forgery attention to localize forged regions based on a channel-wise high pass filter block to extract contextual dependencies of intrinsic inconsistency from forged areas. Mayer *et al.* [29] proposed a two-part deep-learning system to determine whether two image patches contain different forensic traces.

Another more pervasive but complex situation is when forgers spread maliciously forged images on online social networks. These online social networks will perform a series of lossy operations on the uploaded image [21, 30]. In order to enhance the robustness of online social network transmission, Wu *et al.* [31, 32] modeled the noise introduced by online social networks and designed a new training scheme for robust image forgery detection. Unlike these existing methods that feed lossy noise into a training framework to improve the robustness of OSN transmission, we consider reducing the impact of lossy noise on image forensics.

3 THE PROPOSED NETWORK

In this section, we first illustrate an overview of the proposed PRest-Net. Then, the details of the proposed multi-domain probability estimation method and a forgery detector are introduced, respectively.

3.1 Overview of the Proposed Network

Image content forgery will change image pixels, and the continuity between adjacent pixels will be destroyed. For example, suppose an image is forged by splicing. In that case, there is a significant difference between the pixels of the tampered area and the real area because the tampered area of the fake image comes from another image. In addition, if the image is forged by copy-move or removal, although the tampered region comes from the source image, the continuity of adjacent pixels at the edge where the tampered region intersects the real region will not exist. Thus, the differential information of different regions is used as effective evidence for image forgery detection in our proposed network.

Fig. 4 shows the proposed multi-domain probability estimation network for online social network shared image forgery localization. PRest-Net is mainly composed of multi-domain probability estimation and forgery detector. Since the lossy noise introduced by online social network transmission will change the tampering traces of semantic tampering operations (such as splicing, copy-move, and removal), in order to reduce the negative

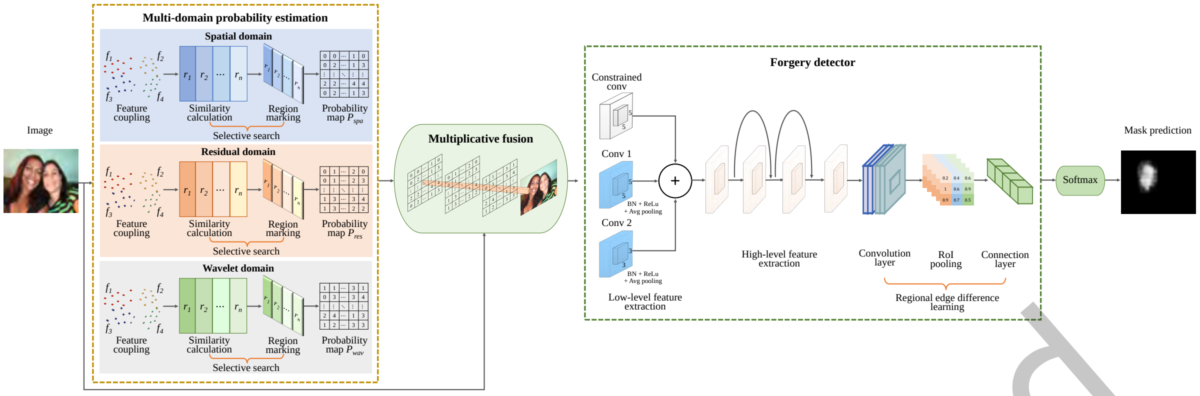

Fig. 4. Overview of the proposed PRest-Net for OSNs shared image forgery detection. The multi-domain probability estimation module constructs new coupled features to capture diversified forensic information by analyzing the coupling relationships between forensic features $\{f_1, f_2, f_3, f_4\}$. Then, based on selective search, the similarity of different regions is calculated, and the suspected tampered regions are marked to obtain the tampering probability map. The probability maps obtained from the spatial domain, residual domain, and wavelet domain are fused with the given image and fed to the forgery detector. The detector consists of low-level feature extraction, high-level feature extraction, and regional edge difference learning module, which adaptively learns more detailed local edge information and achieves fine-grained tampered region prediction. By optimizing the model, PRest-Net can accurately detect OSN-shared image forgery. BN: Batch-Normalization Layer; ReLu: Rectified Linear Unit Layer.

impact of lossy noise on forgery detection, we adopt probability estimation to mine the most differentiated regional information from the spatial, residual, and wavelet domain. More specifically, we learn pixel-wise forensic information from the spatial domain. Based on the fact that the wavelet coefficients corresponding to the effective image information are larger than those corresponding to lossy noise, and residual processing can enhance the detail of image edge structure weakened by lossy noise, we convert the images into the residual domain and wavelet domain respectively to suppress lossy noise and extract the most differentiated regional traces. Then, for the given image, we can obtain the tampering probability of different areas by calculating the similarity of adjacent regions and marking the suspicious tampered areas. As a result, coarse-grained image forgery detection could be realized.

Furthermore, the forgery detector consists of low-level feature extraction, high-level feature extraction, and regional edge difference learning module is designed to capture inconsistencies between the boundaries of the tampered and real regions. Concretely, low-level feature extraction utilizes three parallel convolutions to limit the complex background content of the image and learn shallow forensic information of the tampering probability map obtained by multi-domain probability estimation. Then, high-level feature extraction is employed to capture deeper semantic tampering information and optimize the learning of boundary details in tampered regions. In the regional edge difference learning module, the regions with significant local edge differences are selected as the candidate tampered regions through computations such as convolution, pooling, and connection. Finally, the softmax layer is used to estimate the tampering probability of each region within the candidate regions, thereby predicting tampered regions and achieving fine-grained forgery localization.

3.2 Multi-domain Probability Estimation

The multi-domain probability estimation method is composed of feature coupling and selective search. In fact, image forgery is usually performed in texture regions, and content editing will cause inconsistent tones in

Algorithm 1 The multi-domain probability estimation algorithm.

```
Input: $I_{spa}$: a spatial domain image (i.e., the investigated image)
Output: $P_{spa}$: spatial probability map, $P_{res}$: residual probability map, $P_{wav}$: wavelet probability map
Convert $I_{spa}$ into the residual domain image $I_{res}$ and wavelet domain image $I_{wav}$
while $I_i$ is contained in $\mathbb{I}$={$I_{spa}$,$I_{res}$,$I_{wav}$} do
  Initialize the region $R = \{r_1, r_2, ..., r_n\}$ for $I_i$
  for each neighbouring region pair $(r_i, r_j)$ do
    Extract features $f_1$, $f_2$, $f_3$ and $f_4$ by Eqs. (3)-(6)
    for $i = 1 \longrightarrow 4$ do
      $\mathbf{f}_i = [f_i, f_i^2, f_i^3]$
    end for
    for $i = 1 \longrightarrow 4$ do
      for $j = 1 \longrightarrow 4$ do
        $dis(\mathbf{f}_i, \mathbf{f}_j) = \sqrt{\sum_{k=1}^{m} (\mathbf{f}_{i_k} - \mathbf{f}_{j_k})^2}$
      end for
      $\overrightarrow{f_i} = f_i \times min(dis(\mathbf{f}_i, \mathbf{f}_j))$
    end for
    Calculate similarity $Sim(r_i, r_j) = \sum_{i=1}^{4} \overrightarrow{f_i}$
  end for
  Extract suspicious tampering regions in $R$
  Obtain tampering probability map $P_i$
end while
```

different regions. In addition, the peak signal to noise ratio can expose the difference between the tampered areas and the real area, and the size feature can fit the boundary information of each region. Hence, we first extract the texture pixels, color histograms, peak signal to noise ratio, and size features of different regions from the spatial, residual, and wavelet domains to obtain the most differentiated regional information and reduce the negative impact of lossy noise generated by OSN transmission. Then, we analyze the coupling relationships between these forensic features based on distance calculation and generate new coupled features. By calculating the coupled feature similarity between different image blocks, the suspicious tampered regions can be marked, so as to obtain the tampering probability map. Algorithm 1 gives pseudocode outlining this process, and the more detailed steps are demonstrated as follows.

Step 1: To limit the masking effect of lossy operations on forensic clues, we mine forgery traces from multiple domains. The spatial domain image I_{spa} is firstly converted into the residual domain image I_{res}, namely, the image I_{spa} is box filtered to obtain I_{filt}, and then the residual image I_{res} is computed as

$$I_{res} = |I_{spa} - I_{filt}|, \tag{1}$$

where $|\cdot|$ is calculating the absolute value.

Step 2: Convert the spatial domain image I_{spa} into the wavelet domain image I_{wav}, that is, the image I_{spa} is transformed by Haar wavelet to obtain I_{wav}.

$$I_{wav} = \sum I_{spa} \cdot \frac{1}{\sqrt{2}} \cdot \psi(t), \tag{2}$$

where $\psi(t)$ represents the wavelet basis function. '$\cdot$' denotes the convolution operation.

Step 3: Create the initial region $R = \{r_1, r_2, ..., r_n\}$ for I_i by image segmentation [33], respectively.

Step 4: Extract the texture pixels feature f_1 from neighbouring region pair (r_i, r_j).

$$f_1 = \sum_{k=1}^{N} min(t_i^k, t_j^k), \tag{3}$$

where t_i and t_j are texture histograms of regions r_i and r_j. To extract the histogram, we take the Gaussian derivative in eight directions for r_i and r_j, and set a bin of size 10. Due to the given image includes three color channels, $N = 3 \times 8 \times 10 = 240$.

Step 5: Extract the color histograms feature f_2 from neighbouring region pair (r_i, r_j).

$$f_2 = \sum_{k=1}^{N} min(c_i^k, c_j^k), \tag{4}$$

where c_i and c_j are color histograms of regions r_i and r_j. We use 25 bins to extract the color histogram feature. Thus, $N = 3 \times 25 = 75$.

Step 6: Extract the peak signal to noise ratio feature f_3 from neighbouring region pair (r_i, r_j).

$$f_3 = 10 \times (\log_{10} \frac{255}{MSE(t_i, t_j)} + \log_{10} \frac{255}{MSE(c_i, c_j)}), \tag{5}$$

where $MSE(t_i, t_j)$ is the mean square error of the texture histograms of regions r_i and r_j. $MSE(c_i, c_j)$ is the mean square error of the color histograms of regions r_i and r_j.

Step 7: Extract the size feature f_4 from neighbouring region pair (r_i, r_j).

$$f_4 = 1 - \frac{size(r_i) + size(r_j)}{size(image)}, \tag{6}$$

where $size(image)$ represents the size of the investigated image in pixels.

Step 8: To expose the intrinsic linear and non-linear coupling relationships among the texture pixels feature f_1, color histograms feature f_2, peak signal to noise ratio feature f_3, and size feature f_4, we first expand these features with their powers.

$$\mathbf{f}_i = [f_i, f_i^2, f_i^3]. \tag{7}$$

Moreover, the coupling relationship between these features is measured with the Euclidean distance. Given the features $\mathbf{f}_i$ and $\mathbf{f}_j$, the Euclidean distance is calculated as

$$dis(\mathbf{f}_i, \mathbf{f}_j) = \sqrt{\sum_{k=1}^{m} (\mathbf{f}_{i_k} - \mathbf{f}_{j_k})^2}, \tag{8}$$

where m is the feature dimension.

Let $\mathbb{F} = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4\}$, the paired coupling relationships of features in $\mathbb{F}$ can be expressed as a Feature Coupling Matrix (FCM),

$$FCM = \begin{bmatrix} dis(\mathbf{f}_1, \mathbf{f}_1) & dis(\mathbf{f}_1, \mathbf{f}_2) & dis(\mathbf{f}_1, \mathbf{f}_3) & dis(\mathbf{f}_1, \mathbf{f}_4) \\ dis(\mathbf{f}_2, \mathbf{f}_1) & dis(\mathbf{f}_2, \mathbf{f}_2) & dis(\mathbf{f}_2, \mathbf{f}_3) & dis(\mathbf{f}_2, \mathbf{f}_4) \\ dis(\mathbf{f}_3, \mathbf{f}_1) & dis(\mathbf{f}_3, \mathbf{f}_2) & dis(\mathbf{f}_3, \mathbf{f}_3) & dis(\mathbf{f}_3, \mathbf{f}_4) \\ dis(\mathbf{f}_4, \mathbf{f}_1) & dis(\mathbf{f}_4, \mathbf{f}_2) & dis(\mathbf{f}_4, \mathbf{f}_3) & dis(\mathbf{f}_4, \mathbf{f}_4) \end{bmatrix}. \tag{9}$$

The smaller the calculated distance value, the greater the coupling degree of the two features, reflecting the higher support for the judgment of whether the two regions r_i and r_j are similar. The features with high coupling are assigned greater weight. Based on this analysis, the new coupled features are expressed as

$$\overrightarrow{f_i} = f_i \times min(dis(\mathbf{f}_i, \mathbf{f}_j)), \tag{10}$$

where $i, j = 1, 2, 3, 4$ and $j \neq i$. The dimension of feature $\overrightarrow{f_i}$ is 1.

Through feature coupling, it can be found which feature plays a significant role in image forgery localization. By optimizing the learning weights of features with high coupling degree, the complementarity of forensic information can be adaptive strengthened and the localization accuracy can be improved.

Step 9: Coupled features $\{\overrightarrow{f_1}, \overrightarrow{f_2}, \overrightarrow{f_3}, \overrightarrow{f_4}\}$ are used to calculate the similarity of adjacent regions. The final similarity can be calculated as

$$Sim(r_i, r_j) = \sum_{i=1}^{4} \overrightarrow{f_i}. \tag{11}$$

Similar regions are merged and the suspicious tampered regions are obtained via selective search [34]. Then, the pixel values in the suspicious regions are marked 1. In contrast, those that are not in the suspicious region are marked 0. In addition, the pixel values in the overlapping suspicious regions are accumulated with a step size of 1. As a result, the tampering probability map P_i is estimated.

Multiplicative fusion module: In order to obtain more diversified forensic clues, we merge the tampering probability maps $\{P_{spa}, P_{res}, P_{wav}\}$ generated from the spatial, residual, and wavelet domain with the investigated image. The multiplicative fusion can be calculated as follows.

$$I_{fuse} = P_{spa} \times P_{res} \times P_{wav} + I_{spa}. \tag{12}$$

Through the utilization of multi-domain probability estimation, the masking impact of lossy noise on semantic forgery traces is mitigated and coarse-grained forgery detection is realized.

3.3 Forgery Detector

Since several online social networks modify the uploaded images by using many loss operations, the transmission over OSNs would affect the effectiveness of forgery detection methods. For example, when an image is uploaded to Facebook, it will experience four operations: format conversion, resizing, enhancement filtering, and JPEG compression [30]. These operations will introduce lossy noise into the uploaded images. We realize the suppression of lossy noise through the proposed multi-domain probability estimation and then mine richer forensic information by using the forgery detector to achieve accurate tampered areas localization.

As illustrated in Fig. 4, the forgery detector is composed of low-level feature extraction, high-level feature extraction, and regional edge difference learning module. In the low-level feature extraction, three parallel convolutions are adopted. Precisely, we first utilize the constrained convolution [35] to suppress the complex image content and adaptively learn the prediction error of probability estimation and the changes in local pixel relationships introduced by different tampering operations. Besides, to learn shallow edge features, we use two regular convolutions, namely Conv 1 of size 5 × 5 and Conv 2 of size 3 × 3, each of which is followed by a batch normalization layer, rectified linear unit layer, and average pooling. The Conv 1 with smaller convolution kernel is used to capture local edge details in the given image, while the Conv 2 with 5 × 5 convolution kernel is used to capture larger scale image texture structures. Finally, the output features of three parallel convolutions are fused by element-wise addition to achieve the information gain of local edge.

Due to the vulnerability of the shallow features, richer semantic forgery traces are learned via the high-level feature extraction, which adopts the residual block [36]. By stacking residual blocks, the image texture details can be enhanced and local edge around the tampered regions and global forgery information can be better captured. Since the multi-domain probability estimation module has assigned tampering probability weights to different regions of the image, the combined utilization of low and high-level features allows the proposed PRest-Net to extract forensic evidence from both local edge and macro perspectives, leading to a more robust representative feature.

In the regional edge difference learning module, to highlight the boundary inconsistency between the real area and the tampered area, the deep features generated from the high-level feature extraction are fed into the convolution layer [37] followed by the RoI pooling and connection layer. By exploring the extracted local boundary difference features in conjunction with the softmax layer, PRest-Net would predict the tampered regions at a fine-grained level.

4 EXPERIMENTAL EVALUATIONS

Following the experimental settings, this section provides the simulation of multi-domain probability estimation. Secondly, ablation studies are performed to show the effectiveness of multi-domain probability estimation. Then, we compare PRest-Net with state-of-the-art methods to verify the forgery detection performance of forged images transmitted over online social networks. Besides, we present the detection performance of forged images without OSN transmission. Finally, we provide some qualitative comparisons of PRest-Net and some state-of-the-art methods.

4.1 Experimental Settings

Implement Details. PRest-Net is implemented with the TensorFlow framework and adopt the SGD with default hyperparameters to optimize model. The training loss function contains cross entropy loss and smooth L_1 loss. The batch size of PRest-Net is 256 for training and 300 for testing. The initial learning rate is set to 1e-3 and then is decreased to 1e-4 after 40K steps. ALL experiments are carried out with a single NVIDIA 2080 Ti GPU server.

Datasets. Wu *et al.* [31] used the Dresden dataset [38] as the source of the original images and created forged images by splicing the original images with the objects from the MS-COCO dataset [39]. We adopt the same dataset as [31] to train our model, which contains 9984 forged images. Then, we randomly divide this dataset into training and validation sets with a ratio of 9:1.

For performance evaluations, we compare PRest-Net with current state-of-the-art methods on the following three widely-used datasets.

- NIST [40]: This dataset contains 564 tampered images, which are forged by commonly used semantic tampering operations, i.e., splicing, copy-move, and removal. Meanwhile, these images are post-processed with unknown operations. The resolutions of the forged images range from 500 × 500 to 5616 × 3744.
- DSO [41]: This dataset consists of 100 expertly forged images. To cover up the forgery traces, the forged images have experienced a series of post-processing operations, such as adjustments of color and illumination. The resolutions of the forgeries are 2048 × 1536.
- Columbia [42]: This dataset contains 160 spliced images without compression operation. The resolutions of the spliced images range from 757 × 568 to 1152 × 768.

Besides, to evaluate the robustness of PRest-Net against OSN sharing, we use the same OSN-transmitted dataset as [31]. That is, transmitting the NIST, DSO, and Columbia datasets through the four most popular online social networks: Facebook, Whatsapp, Weibo, and Wechat.

SOTA Models. We compare our method with the following state-of-the-art methods:

- **MT-Net**: A fully convolutional network using Z-score feature and long short-term memory solution to capture local anomaly [17].
- **NoiPri**: A Siamese network using a camera model fingerprint to detect tampered region [26].
- **ForSim**: A two-part network using the forensic similarity to determine whether two image patches contain the same or different tampering traces [29].
- **DFCN**: An encoder-decoder network using dense connections and dilated convolutions to learn forgery traces [14].

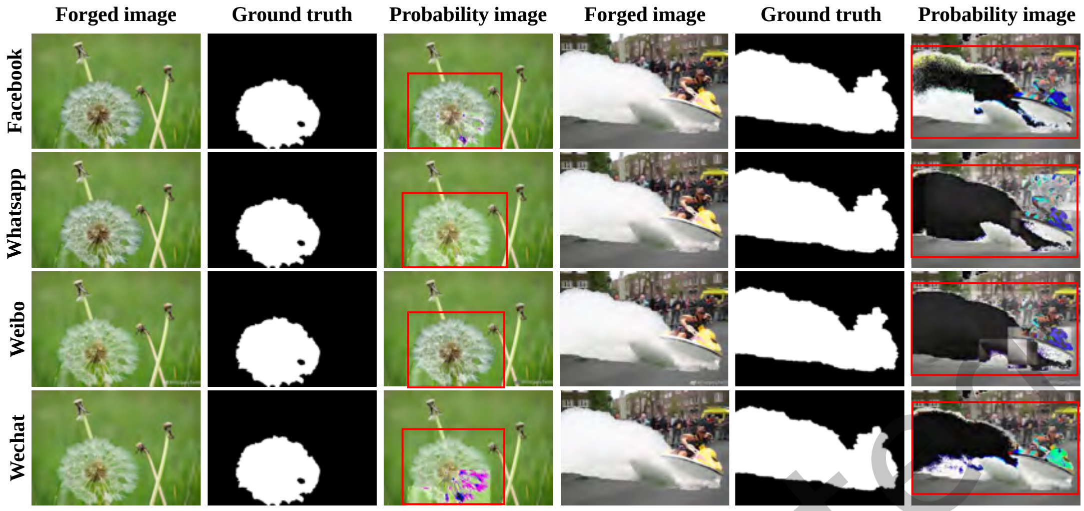

Fig. 5. Visualization results of multi-domain probability estimation derived from the forged image transmitted over Facebook, Whatsapp, Weibo, and Wechat. Images are selected from the NIST dataset. The rectangular box marks the estimated tampered region, whose average coincidence error with the actually tampered region in the ground truth is 0.091.

- **RIFD**: A robust forgery detection network for online social network shared image, which models the lossy noise generated by OSNs transmission and adds it to the training of forgery detector [31].
- **EITLNet**: A two-branch enhanced transformer encoder with attention-based feature fusion for image tampering localization [43].

We compare PRest-Net with **DFCN** and **EITLNet** retrained on our training dataset for fairness. For the other four competitors, we adopt their officially released models because their training mechanism requires specific training samples. Besides, the released model provided by **RIFD** was trained on two datasets, one is an unreleased dataset, and the other is the training set used in our experiment.

Evaluation Metric. We use the pixel-level F_1 score and Area Under the receiver operating characteristic Curve (AUC) to measure the forgery localization performance. For calculating the F_1 score, threshold values are necessary because the direct output of the proposed network is the probability value. Similar to [31], we set the threshold to 0.5.

4.2 Simulation of Multi-domain Probability Estimation

To intuitively demonstrate the usefulness of the proposed multi-domain probability estimation method for OSN transmitted image forgery detection, we perform a visual simulation. As illustrated in Fig. 5, the forged image is transmitted over Facebook, Whatsapp, Weibo, and Wechat. The tampering probability of different image regions can be calculated by using probability estimation. Then, through the multiplicative fusion by Eq. (12), the tampering probability map would be obtained. Specifically, through feature coupling and selective search, the suspected tampered regions can be identified and marked. The pixel values within the tampered regions are designated as 1, while those in authentic regions are set to 0. In instances where multiple suspected tampered regions overlap, their corresponding pixel values are aggregated in step 1. A higher pixel value indicates a greater likelihood of tampering. As a result, the contrast between suspected tampered regions and authentic regions in the generated probability map is enhanced, highlighting significant boundary differences. The rectangular

Table 1. Comparison of ablation study: detection evaluation with and without (w/out) multi-domain probability estimation (PRest). "with spatial", "with residual", and "with wavelet" indicate that only the corresponding single image domain is used to perform probability estimation. The highest value is **bold**.

Metric	Method	Facebook	Whatsapp	Weibo	Wechat
F_1	w/out PRest	0.417	0.407	0.416	0.419
	with spatial	0.414	0.382	0.415	0.407
	with residual	0.419	0.416	0.420	0.416
	with wavelet	0.416	0.416	0.414	0.412
	with PRest	**0.420**	**0.418**	**0.423**	**0.422**
AUC	w/out PRest	0.732	0.708	0.725	0.725
	with spatial	0.726	0.693	0.716	0.717
	with residual	0.735	0.725	0.739	0.737
	with wavelet	0.729	0.709	0.718	0.725
	with PRest	**0.746**	**0.742**	**0.740**	**0.744**

box indicated in the probability map represents the corresponding predicted mask, which is shown in the third and sixth columns of Fig. 5. The first three columns denote the forgery localization of small target objects in the image, and the last three columns are the localization of large target objects. By comparing the ground truth and probability map, we can find that the tampered region predicted by multi-domain probability estimation has a high coincidence degree with ground truth.

We further calculate the mean absolute percentage error *MAPE* between the tampered region obtained by multi-domain probability estimation and ground truth to quantify the coincidence degree of the predicted tampered region and the actually tampered region,

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}|\frac{coor_{real} - coor_{pred}}{coor_{real}}|, \tag{13}$$

where $coor_{real} = \{x, y, w, h\}$, (x, y) is the starting coordinate of the actually tampered region, w is the width of the actually tampered region, and h is the height of the actually tampered region. $coor_{pred} = \{x, y, w, h\}$ is the value of the predicted tampered region. $m = 4$ is the dimension of the real sample and the predicted sample. $|\cdot|$ is calculating the absolute value. The smaller the value of *MAPE*, the higher the degree of coincidence.

The average coincidence error *MAPE* on Facebook, Whatsapp, Weibo, and Wechat is 0.072, 0.082, 0.126, and 0.082 respectively, which indicates that the coarse-grained forgery detection for images shared over Facebook, Whatsapp, Weibo, and Wechat can be realized, although different OSNs will introduce diverse lossy operations. The reason is that we comprehensively consider the multi-domain information fusion and the coupling relationships between multiple forensics features. Because the wavelet transform can strengthen the useful information in the image and weaken the noise information, while the residual transform can enhance the boundary texture details of the image, the inconsistency between the tampered regions and the real regions can be reflected. As a result, the lossy noise caused by OSN transmission can be suppressed, and the forgery traces will become easier to detect.

4.3 Ablation Study

In the ablation study, we evaluate the effectiveness of the multi-domain probability estimation method (PRest), which is designed to reduce the negative impact of lossy noise on image forgery detection. Specifically, we evaluate the forgery detection performance of our network without and with multi-domain probability estimation.

Table 2. Comparison against SOTA methods by using F_1 and AUC as criteria. The forged images are transmitted over Facebook, Whatsapp, Weibo, and Wechat. The highest value is **bold**.

OSNs	Method	NIST		DSO		Columbia	
		F_1	AUC	F_1	AUC	F_1	AUC
Facebook	**MT-Net** [17]	0.132	0.547	0.073	0.526	0.119	0.529
	NoiPri [26]	0.057	0.583	0.238	0.499	0.412	0.500
	ForSim [29]	0.140	0.580	0.356	0.689	0.450	0.607
	DFCN [14]	0.214	0.633	0.218	0.549	0.525	0.700
	RIFD [31]	0.326	0.717	0.459	**0.720**	**0.716**	**0.818**
	EITLNet [43]	0.320	0.694	0.348	0.702	0.619	0.707
	PRest-Net	**0.420**	**0.746**	**0.460**	**0.720**	0.633	0.711
Whatsapp	**MT-Net** [17]	0.113	0.528	0.070	0.526	0.113	0.528
	NoiPri [26]	0.073	0.579	0.240	0.499	0.412	0.499
	ForSim [29]	0.137	0.586	0.233	0.542	0.436	0.595
	DFCN [14]	0.226	0.645	0.314	0.649	0.522	0.700
	RIFD [31]	0.313	0.717	0.342	0.670	**0.727**	**0.828**
	EITLNet [43]	0.281	0.729	0.345	0.600	0.622	0.708
	PRest-Net	**0.418**	**0.742**	**0.416**	**0.672**	0.626	0.709
Weibo	**MT-Net** [17]	0.132	0.549	0.069	0.522	0.102	0.524
	NoiPri [26]	0.054	0.580	0.239	0.500	0.411	0.498
	ForSim [29]	0.150	0.581	0.260	0.568	0.453	0.610
	DFCN [14]	0.217	0.641	0.299	0.636	0.519	0.699
	RIFD [31]	0.289	0.707	0.376	0.691	**0.726**	**0.826**
	EITLNet [43]	0.347	0.734	0.349	0.659	0.619	0.711
	PRest-Net	**0.423**	**0.740**	**0.441**	**0.701**	0.633	0.718
Wechat	**MT-Net** [17]	0.113	0.542	0.073	0.528	0.149	0.540
	NoiPri [26]	0.041	0.575	0.240	0.500	0.412	0.499
	ForSim [29]	0.136	0.581	0.247	0.564	0.496	0.650
	DFCN [14]	0.225	0.635	0.302	0.646	0.522	0.698
	RIFD [31]	0.280	0.693	0.368	0.680	**0.732**	**0.829**
	EITLNet [43]	0.284	0.733	0.341	0.697	0.614	0.700
	PRest-Net	**0.422**	**0.744**	**0.451**	**0.708**	0.624	0.703

The network without probability estimation means that the input image is fed directly into our proposed forgery detector. Besides, we evaluate the performance improvement effects of probability estimation based on a single domain (i.e., spatial domain, residual domain, or wavelet domain) and probability estimation based on multi-domain fusion. The comparison results of forgery detection of images transmitted on Facebook, Whatsapp, Weibo and Wechat are shown in Table 1. The testing forged images are from the NIST dataset. We can find that the forgery localization performance of the proposed network with multi-domain probability estimation (with PRest) outperforms that without multi-domain probability estimation (w/out PRest).

Since there exists continuity of image pixels in the spatial domain, lossy operations will introduce lossy noise into the image and destroy its continuity. When the image is converted from the spatial domain to the wavelet domain, the wavelet coefficients corresponding to the lossy noise are smaller than those corresponding

Table 3. Comparison of detection time efficiency (Unit: seconds) between PRest-Net, **MT-Net**, and **RIFD** on different datasets. The smallest value is **bold**.

Method	OSNs	NIST	DSO	Columbia
MT-Net [17]	Facebook	7329.31	1255.76	2123.41
	Whatsapp	7382.37	1470.20	2074.83
	Weibo	7400.02	1270.28	2096.09
	Wechat	7368.40	1256.79	2150.54
RIFD [31]	Facebook	2566.47	539.42	110.54
	Whatsapp	555.19	113.54	105.87
	Weibo	1357.53	257.91	104.99
	Wechat	1349.10	260.38	111.82
PRest-Net	Facebook	**633.94**	**155.40**	**47.04**
	Whatsapp	**176.53**	**39.80**	**38.56**
	Weibo	**342.35**	**74.32**	**40.32**
	Wechat	**356.45**	**71.70**	**42.88**

to the valuable image information, which can effectively suppress the negative effect of lossy noise on forensic information extraction. Meanwhile, the lossy noise may cover the weak edge information of the forged image, when the image is transformed from the spatial domain to the residual domain, the structural details of the forged image can be enhanced. The spatial domain-based probability estimation (with spatial) can effectively reflect the pixel discontinuity caused by image forgery, the residual domain based-probability estimation (with residual) can strengthen the edge difference between the tampered region and the real region, and the wavelet domain-based probability estimation (with wavelet) can extract the difference between useful information and lossy noise information. As a result, the expression of forensic knowledge can be strengthened and the differential information of different regions can be learned through the fusion of multi-domain tampering probability maps. Hence, the lossy noise is suppressed and the detection performance of images transmitted over online social networks is further improved by utilizing the probability estimation based on multi-domain fusion (with PRest).

4.4 Comparisons With SOTA Methods

In this experiment, we validate the robustness of PRest-Net and compare it with **MT-Net** [17], **NoiPri** [26], **ForSim** [29], **DFCN** [14], **RIFD** [31], and **EITLNet** [43]. Table 2 provides the comparison of F_1 score and AUC between PRest-Net and the SOTA methods for forged images transmitted over different online social networks. As seen from this table, the forgery detection performance of PRest-Net is superior to that of SOTA methods on the NIST and DSO datasets.

In the validation of the Columbia dataset, although the tampered regions localization performance of our PRest-Net is lower than **RIFD**, compared with **MT-Net**, **NoiPri**, **ForSim** and **DFCN**, PRest-Net can achieve more outstanding performance. **RIFD** can obtain better detection results on the Columbia dataset that may be related to the other training dataset it used. The images in the Columbia dataset are uncompressed spliced images of high quality. The images still have high quality after being transmitted through online social networks. For the training of **RIFD**, a dataset containing original images and their version transmitted over Facebook was adopted. This training process may make **RIFD** overfit so that the detection performance of high-quality spliced images is improved. Conversely, the results on the NIST dataset showed that the forensic performance of **RIFD** decreased for images that experienced different tampering operations.

Table 4. Comparison against SOTA methods on forged images without online social network transmission. F_1 and AUC are used as evaluation criteria. The highest value is **bold**.

Method	NIST		DSO		Columbia	
	F_1	AUC	F_1	AUC	F_1	AUC
MT-Net [17]	0.157	0.560	0.088	0.528	0.352	0.620
NoiPri [26]	0.119	0.672	0.238	0.500	0.420	0.510
ForSim [29]	0.188	0.642	**0.487**	**0.796**	0.604	0.731
DFCN [14]	0.170	0.596	0.220	0.564	0.546	0.710
RIFD [31]	0.334	0.686	0.446	0.723	**0.712**	**0.815**
EITLNet [43]	0.246	0.704	0.320	0.694	0.622	0.706
PRest-Net	**0.419**	**0.747**	0.451	0.724	0.627	0.710

In addition, considering the detection time efficiency, Table 3 shows the comparison results between **MT-Net**, **RIFD**, and the proposed PRest-Net. We can find that PRest-Net requires less detection time than **MT-Net** and **RIFD**, especially, the detection time required by PRest-Net is at least half less than that of **RIFD**. Namely, it is more efficient to use PRest-Net for image forgery detection over online social networks.

The reason why PRest-Net can achieve better forgery localization performance may be that PRest-Net captures forgery clues from the spatial, residual, and wavelet domain, respectively, and utilizes feature coupling to increase the weight of useful forensics features, so as to obtain the differential information of different regions. Furthermore, our proposed forgery detector considers both local boundary difference learning and macroscopic general features learning, which enhances the representation ability of PRest-Net and thus improves the localization performance.

4.5 Performance Comparison on Forged Images without OSNs Transmission

In this experiment, we evaluate the effectiveness of PRest-Net in the case of original forgery. That is, the forged images have not been transmitted over online social networks. The results in Table 4 demonstrate that PRest-Net achieves competitive forgery localization performance than these SOTA methods on the NIST dataset, which contains three kinds of forged images: spliced images, copy-moved images, and removed images. That is, our PRest-Net has better generalization for detecting and locating different semantic tampering operations.

On the DSO dataset, in addition to **ForSim**, PRest-Net and state-of-the-art methods achieve comparable results, especially compared to **MT-Net**, **NoiPri**, **DFCN** and **EITLNet**, PRest-Net shows a more desirable performance. The F_1 and AUC values of **ForSim** are higher than that of other methods, probably because **ForSim** captures the camera model features to determine whether image patches contain the same or different forensic traces. Camera model features are beneficial features for identifying different camera sources. Therefore, **ForSim** could obtain great detection performance in spliced datasets, like the DSO and Columbia datasets, in which the source of tampered regions and original real images are different. However, the image forgery detection performance on the NIST dataset is worse than our PRest-Net. Because the tampered regions of some images (such as copy-moved images) are from the same source as the real regions and the camera model features will be invalid.

On the Columbia dataset, the performance of PRest-Net is only weaker than that of **RIFD**. The reason may be that **RIFD** feeds a large number of original images when training the detector. By learning more image information, better detection results for images without compression could be achieved.

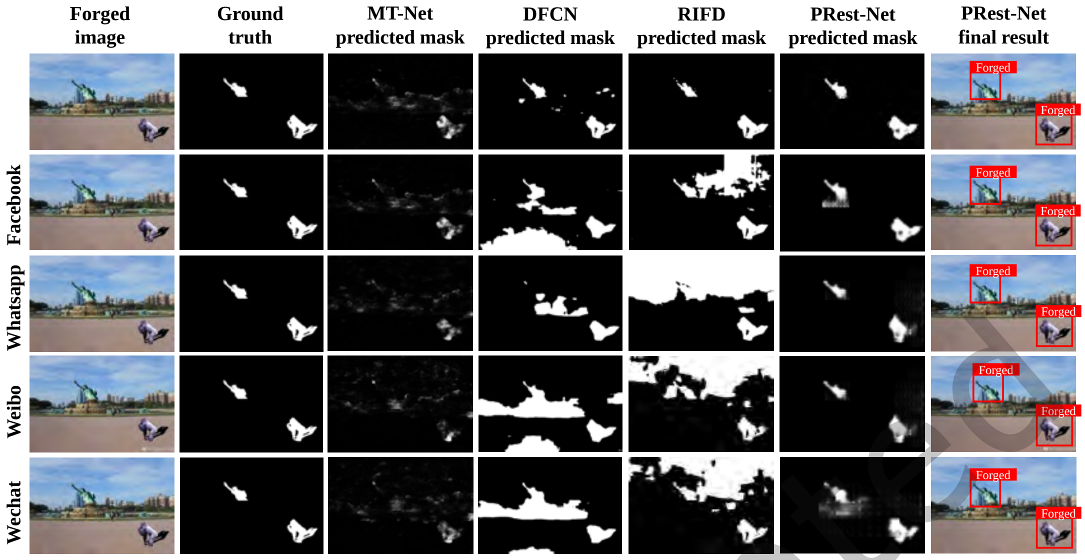

Fig. 6. Qualitative comparisons for the different OSNs transmitted image forgery localization on the NIST dataset. The images from left to right are the image forged by splicing, ground truth, predicted mask obtained using **MT-Net** [17], **DFCN** [14], **RIFD** [31] and our PRest-Net, and the final result obtained using our PRest-Net. From top to bottom, it shows the cases without OSN transmission, and with Facebook, Whatsapp, Weibo and Wechat transmissions, respectively.

4.6 Qualitative Comparison

Figs. 6-8 provide some qualitative results for comparison of the proposed PRest-Net architecture, **MT-Net**, **DFCN**, and **RIFD**. To show that our PRest-Net is effective for multiple tampering operations and different datasets, Figs. 6-7 give two examples of splicing forgery detection and removal forgery detection, and the forged images are selected from the NIST dataset. Additionally, Fig. 3 provides the example of copy-move forgery detection. Fig. 8 gives an example of splicing forgery detection, where the forged image is selected from the Columbia dataset.

As shown in Fig. 6, we can find that the state-of-the-art methods perform well in cases where forged images are not uploaded to online social networks. However, in cases of OSN-transmitted images, their detection performance decreased. In contrast, the results provided in Figs. 6-8 demonstrate that our proposed PRest-Net obtains more precise localization results over various OSN transmission. This is because PRest-Net using the multi-domain probability estimation to limit the confusing effect of lossy operations on content forgery traces so that relatively clean forensic clues can be captured.

Moreover, when altering images through various tampering operations, the state-of-the-art methods may encounter missed or false detection, such as the inability to accurately identify the removed area using **DFCN**. Besides, for falsified images exhibiting significant discrepancies between the authentic background and the forged region, SOTA methods demonstrates superior localization performance. Conversely, when the edges of the tampered region are seamlessly integrated with the real background, the localization performance of SOTA methods degrades. For instance, **DFCN** shows relatively good performance in Fig. 8 but performs poorly in Fig. 6. This discrepancy stems from the utilization of dilated convolution, which offers larger receptive fields and enables the capture of more pronounced edge differences. Nevertheless, **DFCN** also overlooks some subtle tampering trace, thereby hindering its ability to precisely delineate tampering boundaries in regions that closely resemble their backgrounds. In contrast, our PRest-Net is generalized to detect different operations and is capable of accurately

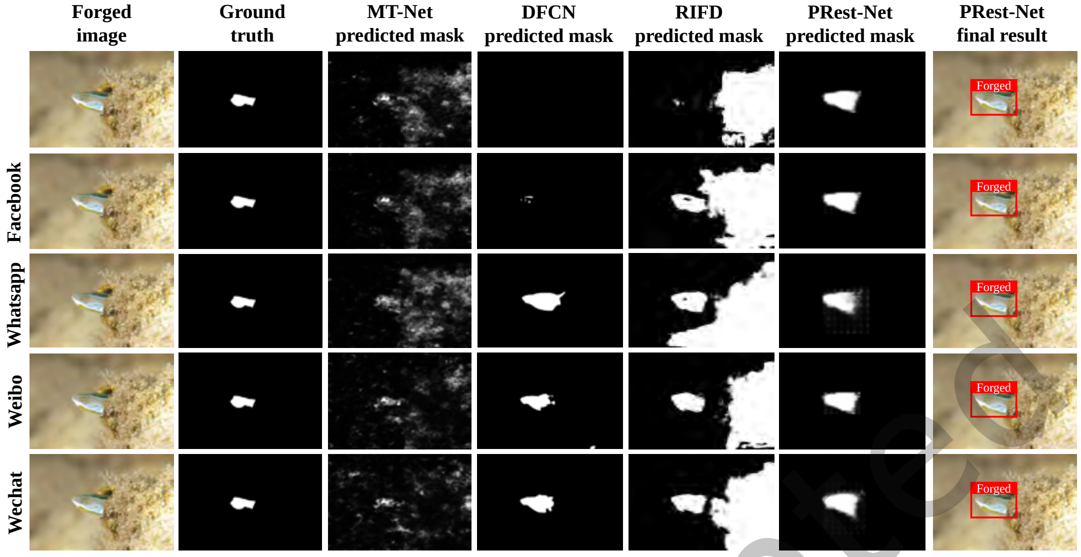

Fig. 7. Qualitative comparisons for the different OSNs transmitted image forgery localization on the NIST dataset. The images from left to right are the image forged by removal, ground truth, predicted mask obtained using **MT-Net** [17], **DFCN** [14], **RIFD** [31] and our PRest-Net, and the final result obtained using our PRest-Net. From top to bottom, it shows the cases without OSN transmission, and with Facebook, Whatsapp, Weibo and Wechat transmissions, respectively.

identifying the tampered regions in well-forged images. This capability is attributed to the implementation of a parallel convolution architecture that enhances local edge features across multiple scales, allowing PRest-Net to effectively capture variations in different image areas.

5 CONCLUSION

In this paper, we propose a novel multi-domain probability estimation network for robust image forgery detection against online social network transmission. The detection issue is solved from the perspective of multi-domain information learning, which introduces a probability estimation idea to suppress the lossy noise caused by OSN transmission and capture the most differentiated regional features. The proposed PRest-Net allows for the combination of low-level features with higher-level features and focuses on the overall and local detail changes caused by various tampering operations, significantly improving the forgery detection performance. Experimental results demonstrated that PRest-Net achieves competitive and robust localization performance in different online social network transmission scenarios. As part of our future efforts, we will design more forensic features to learn richer forgery information and extend PRest-Net to improve the forgery detection performance for images shared over different online social networks.

6 ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China under Grants 62402062, U22A2030, and U20B2051, the National Key R&D Program of China under Grants 2024YFF0618800 and 2022YFB3103500, Hunan Provincial Funds for Distinguished Young Scholars under Grant 2024JJ2025, Hunan Provincial Key Research and Development Program under Grant 2024AQ2027, Natural Science Foundation of Changsha City, China

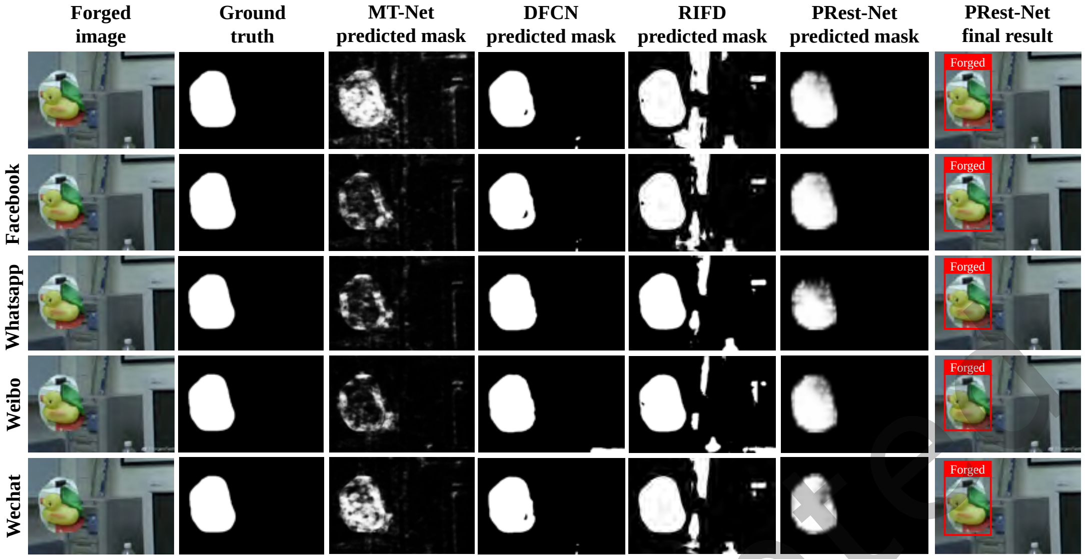

Fig. 8. Qualitative comparisons for the different OSNs transmitted image forgery localization on the Columbia dataset. The images from left to right are the image forged by splicing, ground truth, predicted mask obtained using **MT-Net** [17], **DFCN** [14], **RIFD** [31] and our PRest-Net, and the final result obtained using our PRest-Net. From top to bottom, it shows the cases without OSN transmission, and with Facebook, Whatsapp, Weibo and Wechat transmissions, respectively.

under Grant kq2402031, Key Laboratory of Forensic Science and Technology at College of Sichuan Province under Grant 2024ZD03.

REFERENCES

[1] Feng Ding, Zhangyi Shen, Guopu Zhu, Sam Kwong, Yicong Zhou, and Siwei Lyu. 2023. ExS-GAN: Synthesizing Anti-Forensics Images via Extra Supervised GAN. *IEEE Transactions on Cybernetics* 53, 11 (2023), 7162–7173.

[2] Ritesh Vyas, Michele Nappi, Alberto del Bimbo, and Sambit Bakshi. 2024. Introduction to Special Issue on "Recent trends in Multimedia Forensics". *ACM Transactions on Multimedia Computing, Communications, and Applications* (2024), 1–7.

[3] Irene Amerini, Lamberto Ballan, Roberto Caldelli, Alberto Del Bimbo, and Giuseppe Serra. 2011. A SIFT-Based Forensic Method for Copy-Move Attack Detection and Transformation Recovery. *IEEE Transactions on Information Forensics and Security* 6, 3 (2011), 1099–1110.

[4] Wei Wang, Jing Dong, and Tieniu Tan. 2009. Effective image splicing detection based on image chroma. In *2009 IEEE International Conference on Image Processing (ICIP)*. 1257–1260.

[5] Jian Li, Xiaolong Li, Bin Yang, and Xingming Sun. 2015. Segmentation-Based Image Copy-Move Forgery Detection Scheme. *IEEE Transactions on Information Forensics and Security* 10, 3 (2015), 507–518.

[6] Babak Mahdian and Stanislav Saic. 2009. Using Noise Inconsistencies for Blind Image Forensics. *Image and Vision Computing* 27, 10 (2009), 1497–1503.

[7] Xunyu Pan, Xing Zhang, and Siwei Lyu. 2012. Exposing Image Splicing with Inconsistent Local Noise Variances. In *2012 IEEE International Conference on Computational Photography (ICCP)*. 1–10.

[8] Fengyong Li, Huajun Zhai, Teng Liu, Xinpeng Zhang, and Chuan Qin. 2024. Learning Compressed Artifact for JPEG Manipulation Localization Using Wide-Receptive-Field Network. *ACM Transactions on Multimedia Computing, Communications, and Applications* (2024), 1–22.

[9] Yuanman Li, Lanhao Ye, Haokun Cao, Wei Wang, and Zhongyun Hua. 2024. Cascaded Adaptive Graph Representation Learning for Image Copy-Move Forgery Detection. *ACM Transactions on Multimedia Computing, Communications, and Applications* (2024), 1–23.

[10] Chao Wang, Zhiqiu Huang, Shuren Qi, Yaoshen Yu, Guohua Shen, and Yushu Zhang. 2023. Shrinking the Semantic Gap: Spatial Pooling of Local Moment Invariants for Copy-Move Forgery Detection. *IEEE Transactions on Information Forensics and Security* 18 (2023),

1064–1079.

[11] Kui Ye, Jing Dong, Wei Wang, Bo Peng, and Tieniu Tan. 2018. Feature Pyramid Deep Matching and Localization Network for Image Forensics. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. 1796–1802.

[12] Bo Peng, Wei Wang, Jing Dong, and Tieniu Tan. 2018. Image Forensics Based on Planar Contact Constraints of 3D Objects. *IEEE Transactions on Information Forensics and Security* 13, 2 (2018), 377–392.

[13] Yingjie He, Yuanman Li, Changsheng Chen, and Xia Li. 2023. Image Copy-Move Forgery Detection via Deep Cross-Scale PatchMatch. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*. 2327–2332.

[14] Peiyu Zhuang, Haodong Li, Shunquan Tan, Bin Li, and Jiwu Huang. 2021. Image Tampering Localization Using a Dense Fully Convolutional Network. *IEEE Transactions on Information Forensics and Security* 16 (2021), 2986–2999.

[15] Xianjin Liu, Wei Lu, Qin Zhang, Jiwu Huang, and Yun-Qing Shi. 2020. Downscaling Factor Estimation on Pre-JPEG Compressed Images. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 3 (2020), 618–631.

[16] Jiaxin Chen, Xin Liao, Wei Wang, and Zheng Qin. 2023. Identification of Image Global Processing Operator Chain based on Feature Decoupling. *Information Sciences* 637 (2023), 1–18.

[17] Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan. 2019. ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries With Anomalous Features. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 9535–9544.

[18] Rongyu Zhang and Jiangqun Ni. 2020. A Dense U-Net with Cross-Layer Intersection for Detection and Localization of Image Forgery. In *2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2982–2986.

[19] Yuan Rao and Jiangqun Ni. 2021. Self-supervised Domain Adaptation for Forgery Localization of JPEG Compressed Images. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. 15014–15023.

[20] Jiaxin Chen, Xin Liao, Wei Wang, Zhenxing Qian, Zheng Qin, and Yaonan Wang. 2023. SNIS: A Signal Noise Separation-Based Network for Post-Processed Image Forgery Detection. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 2 (2023), 935–951.

[21] Weiwei Sun, Jiantao Zhou, Ran Lyu, and Shuyuan Zhu. 2016. Processing-Aware Privacy-Preserving Photo Sharing over Online Social Networks. In *2016 ACM international conference on Multimedia (ACM MM)*. 581–585.

[22] Shuren Qi, Yushu Zhang, Chao Wang, Jiantao Zhou, and Xiaochun Cao. 2023. A Principled Design of Image Representation: Towards Forensic Tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 5 (2023), 5337–5354.

[23] Edoardo Ardizzone, Alessandro Bruno, and Giuseppe Mazzola. 2015. Copy-Move Forgery Detection by Matching Triangles of Keypoints. *IEEE Transactions on Information Forensics and Security* 10, 10 (2015), 2084–2094.

[24] Yuanman Li and Jiantao Zhou. 2019. Fast and Effective Image Copy-Move Forgery Detection via Hierarchical Feature Point Matching. *IEEE Transactions on Information Forensics and Security* 14, 5 (2019), 1307–1322.

[25] Yuan Rao and Jiangqun Ni. 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)*. 1–6.

[26] Davide Cozzolino and Luisa Verdoliva. 2020. Noiseprint: A CNN-Based Camera Model Fingerprint. *IEEE Transactions on Information Forensics and Security* 15 (2020), 144–159.

[27] Tong Chen, Bin Li, and Jinhua Zeng. 2023. Learning Traces by Yourself: Blind Image Forgery Localization via Anomaly Detection With ViT-VAE. *IEEE Signal Processing Letters* 30 (2023), 150–154.

[28] Long Zhuo, Shunquan Tan, Bin Li, and Jiwu Huang. 2022. Self-Adversarial Training Incorporating Forgery Attention for Image Forgery Localization. *IEEE Transactions on Information Forensics and Security* 17 (2022), 819–834.

[29] Owen Mayer and Matthew C. Stamm. 2020. Forensic Similarity for Digital Images. *IEEE Transactions on Information Forensics and Security* 15 (2020), 1331–1346.

[30] Weiwei Sun, Jiantao Zhou, Yuanman Li, Ming Cheung, and James She. 2021. Robust High-Capacity Watermarking Over Online Social Network Shared Images. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 3 (2021), 1208–1221.

[31] Haiwei Wu, Jiantao Zhou, Jinyu Tian, Jun Liu, and Yu Qiao. 2022. Robust Image Forgery Detection Against Transmission Over Online Social Networks. *IEEE Transactions on Information Forensics and Security* 17 (2022), 443–456.

[32] Haiwei Wu, Jiantao Zhou, Jinyu Tian, and Jun Liu. 2022. Robust Image Forgery Detection over Online Social Network Shared Images. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13430–13439.

[33] P. F. Felzenszwalb and D. P. Huttenlocher. 2004. Efficient Graph-based Image Segmentation. *International Journal of Computer Vision* 59 (2004), 167–181.

[34] J.R.R. Uijlings, K.E.A. van de Sande, T. Gevers, and A.W.M. Smeulders. 2013. Selective Search for Object Recognition. *International Journal of Computer Vision* 104 (2013), 154–171.

[35] Belhassen Bayar and Matthew C. Stamm. 2018. Constrained Convolutional Neural Networks: A New Approach Towards General Purpose Image Manipulation Detection. *IEEE Transactions on Information Forensics and Security* 13, 11 (2018), 2691–2706.

[36] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 770–778.

[37] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 6 (2017), 1137–1149.

[38] T. Gloe and R. BoHme. 2010. The Dresden Image Database for Benchmarking Digital Image Forensics. *Journal of Digital Forensic Practice* 3, 2-4 (2010), 150–159.

[39] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *European Conference on Computer Vision (ECCV)*. 1–16.

[40] Nist. 2016. NIST manipulation evaluation dataset. https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation.

[41] Tiago Jose De Carvalho, Christian Riess, Elli Angelopoulou, Helio Pedrini, and Anderson de Rezende Rocha. 2013. Exposing Digital Image Forgeries by Illumination Color Classification. *IEEE Transactions on Information Forensics and Security* 8, 7 (2013), 1182–1194.

[42] Columbia. 2004. Image splicing detection evaluation dataset. http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm.

[43] Kun Guo, Haochen Zhu, and Gang Cao. 2024. Effective Image Tampering Localization Via Enhanced Transformer and Co-Attention Fusion. In *2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 4895–4899.