Multi-domain Probability Estimation Network for Forgery Detection over Online Social Network Shared Images

Jiaxin Chen[1,2], Xin Liao[2,*], Zhenxing Qian[3], Zheng Qin[2]
[1]School of Computer and Communication Engineering, Changsha University of Science and Technology
[2]College of Computer Science and Electronic Engineering, Hunan University
[3]School of Computer Science, Fudan University
jxchen@csust.edu.cn, xinliao@hnu.edu.cn, zxqian@fudan.edu.cn, zqin@hnu.edu.cn

2024 IEEE International Conference on Multimedia and Expo (ICME) | 979-8-3503-9015-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICME57554.2024.10687645

***Abstract*—Fake images spread on online social networks can lead to public misconceptions, causing potential harm to politics, economics and social culture. Researching the authenticity of digital images is of great significance. Almost all online social networks perform multiple lossy operations on uploaded images, which will change the forgery traces and degrade the forgery detection performance of existing forensic methods. Therefore, forgery detection for images transmitted on online social networks has become a severe challenge. To solve this problem, we propose a multi-domain probability estimation network (PRest-Net). Specifically, considering that the wavelet coefficient of useful information is larger than that of lossy noise, and the image residual processing can highlight the edge texture, we design a multi-domain probability estimation method based on feature coupling and selective search to mine forgery traces in the spatial, residual, and wavelet domains. As a result, the lossy noise introduced by transmission on online social networks can be suppressed and the most differentiated regional information can be captured. Furthermore, a forgery detector based on low-level and high-level feature learning is proposed to adaptively capture rich operation clues and local edge differences between tampered and real areas, achieving fine-grained tampered region prediction. Extensive experiments show that the proposed PRest-Net outperforms existing state-of-the-art forgery detectors and is robust to online social network transmission.**

***Index Terms*—Image forensics, probability estimation, online social network transmission**

I. Introduction

The Internet has brought enormous benefits to the industry and community, especially in promoting and applying various online social networks (OSNs), such as Facebook, Whatsapp, Wechat and Weibo, making information transmission more convenient. Meanwhile, digital images have become an essential carrier for acquiring and transmitting information. However, the popularity of image processing technology makes it easy to forge images, and it is difficult for the public to judge whether images have been modified from the visual senses [1], [2]. There is a growing concern about ensuring the authenticity of images shared over various OSNs.

This work was supported in part by the National Natural Science Foundation of China under Grants U22A2030, U20A20174, U20B2051, 61972142, and 61972395, Hunan Provincial Funds for Distinguished Young Scholars under Grant 2024JJ2025. (*Corresponding author: Xin Liao.)

Many forensic efforts utilize clues such as color inconsistency [3] and affine transformation matrix [4] to determine if a specific block in an image has been forged. Most of these methods are designed for a single tampering manipulation detection. Some recently proposed methods [5]–[9] show robustness to general-purpose localization. A fully convolutional network based on local anomaly detection [5] was proposed to perform localization on many manipulations such as splicing, copy-move and removal. Cozzolino *et al.* [6] extracted a camera fingerprint to detect the tampered regions. Mayer *et al.* [7] proposed a two-part network to determine whether two patches contain different forensic traces. An encoder-decoder architecture based on dense connections [8] can obtain excellent localization performance.

However, digital multimedia forgery usually involves multiple operations [10]–[12]. Besides, images are often shared on online social networks and almost all OSNs will modify the uploaded images with several lossy operations, such as resizing and compression. These lossy operations would change the image content forgery traces, making the forgery detection performance decrease [13]. Therefore, forgery localization for images transmitted over online social networks remains an important issue.

In this work, we propose a multi-domain probability estimation network PRest-Net to address the above issue. From the perspective of reducing the influence of lossy noise on image forgery detection, we propose a tampering probability estimation method to analyze the coupling relationship between multiple forensic features and capture the most differentiated regional information. By estimating tampering probability maps from the spatial domain, residual domain, and wavelet domain, and integrating them with the given image based on multiplicative fusion, coarse-grained tampered region localization is realized. For accurately detecting the tampered regions, we design a forgery detector that limits the influence of complex image content on forensics via low-level feature extraction and adaptively learns rich tampering clues through high-level feature extraction and regional edge difference learning. Experimental results demonstrate that the proposed PRest-Net achieves better detection performance in

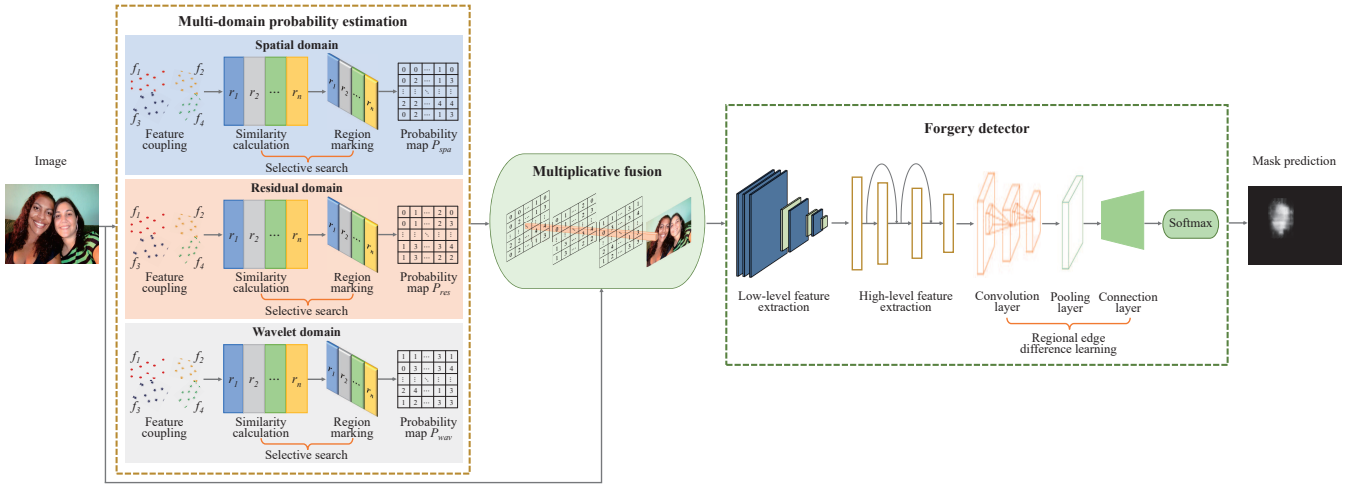

Fig. 1. Overview of the proposed PRest-Net architecture for online social network shared image forgery localization.

comparison with several state-of-the-art methods.

II. The Proposed Network

Fig. 1 shows the proposed PRest-Net, which mainly consists of multi-domain probability estimation and forgery detector. In the multi-domain probability estimation module, since 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 extract multiple forensic features from the perspective of the spatial domain, residual domain, and wavelet domain, and analyze the coupling relationship among these features. By similarity calculation and marking the suspicious tampered areas, the tampering probability maps can be obtained. Then, applying multiplicative fusion to integrate probability maps calculated from multi-domain with the given image, the negative impact of lossy noise generated by OSN transmission can be reduced and coarse-grained image forgery detection can be realized. Moreover, the forgery detector consists of low-level feature extraction, high-level feature extraction, and regional edge difference learning can capture more detailed semantic tampering traces, thus achieving fine-grained forgery localization.

A. Multi-domain Probability Estimation

The multi-domain probability estimation method is designed based on feature coupling and selective search. It is well known that image forgery is usually performed in texture regions, and local semantic tampering can easily lead to inconsistent tones in each region. Besides, the peak signal to noise ratio can measure the difference between the tampered region and the real region, and the size feature can fit the boundary information of each region. Therefore, we first extract the texture pixels, color histograms, peak signal to noise ratio and size features of different regions from the perspective of the spatial domain, residual domain, and wavelet domain. To capture the most differentiated regional information of the investigated image, we analyze the coupling relationships of these features and construct new coupled features. Then, combined with selective search [14], the suspicious tampered regions of the image are marked by calculating the similarity of coupled features among different image blocks, so as to obtain the tampering probability maps. The detailed steps are demonstrated as follows.

Step 1: To explore forensic clues from multiple domains, we convert the spatial domain image I_{spa} into the wavelet domain image I_{wav} by using Haar wavelet transform. Meanwhile, converting I_{spa} into the residual domain image I_{res}. That is, 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: Generating the initial region $R = \{r_1, r_2, ..., r_n\}$ for I_i (I_i is contained in $\mathbb{I}$=$\{I_{spa}$,I_{res},$I_{wav}\}$) based on the segmentation method [15]. Extracting the texture pixels feature f_1 from neighbouring region pair (r_i, r_j).

$$f_1 = \sum_{k=1}^{N} min(t_i^k, t_j^k), \tag{2}$$

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

Step 3: Extracting the color histograms feature f_2 from neighbouring region pair (r_i, r_j).

$$f_2 = \sum_{k=1}^{N} min(c_i^k, c_j^k), \tag{3}$$

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

Step 4: Extracting the peak signal to noise ratio feature f_3 from neighbouring region pair (r_i, r_j).

$$f_3 = 10 \times (\log_{10} \frac{255}{MSE(t_i, t_j)} + \log_{10} \frac{255}{MSE(c_i, c_j)}), \quad (4)$$

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 5: Extracting the size feature f_4 from neighbouring region pair (r_i, r_j).

$$f_4 = 1 - \frac{size(r_i) + size(r_j)}{size(image)}, \quad (5)$$

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

Step 6: To capture the intrinsic linear and non-linear coupling relationships among the texture pixels, color histograms, peak signal to noise ratio, and size features, we first expand these features with their powers.

$$\mathbf{f}_i = [f_i, f_i^2, f_i^3]. \quad (6)$$

Then, we measure the coupling relationship between the features with the Euclidean distance. Given the features $\mathbf{f}_i$ and $\mathbf{f}_j$, the Euclidean distance can be calculated as

$$dis(\mathbf{f}_i, \mathbf{f}_j) = \sqrt{\sum_{k=1}^{m} (\mathbf{f}_{i_k} - \mathbf{f}_{j_k})^2}, \quad (7)$$

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),

$$FCM = \begin{bmatrix} dis(\mathbf{f}_1, \mathbf{f}_1) & dis(\mathbf{f}_1, \mathbf{f}_2) & dis(\mathbf{f}_1, \mathbf{f}_3) & dis(\mathbf{f}_1, \mathbf{f}_4) \\ dis(\mathbf{f}_2, \mathbf{f}_1) & dis(\mathbf{f}_2, \mathbf{f}_2) & dis(\mathbf{f}_2, \mathbf{f}_3) & dis(\mathbf{f}_2, \mathbf{f}_4) \\ dis(\mathbf{f}_3, \mathbf{f}_1) & dis(\mathbf{f}_3, \mathbf{f}_2) & dis(\mathbf{f}_3, \mathbf{f}_3) & dis(\mathbf{f}_3, \mathbf{f}_4) \\ dis(\mathbf{f}_4, \mathbf{f}_1) & dis(\mathbf{f}_4, \mathbf{f}_2) & dis(\mathbf{f}_4, \mathbf{f}_3) & dis(\mathbf{f}_4, \mathbf{f}_4) \end{bmatrix}. \quad (8)$$

The smaller the distance, the higher the coupling degree, which reflects the higher support for the judgment of whether the two regions r_i and r_j are similar. Therefore, we assign more weight to features with a high coupling degree. Based on this analysis, the new coupled features can be expressed as

$$\overrightarrow{f_i} = f_i \times min(dis(\mathbf{f}_i, \mathbf{f}_j)), \quad (9)$$

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

Step 7: Coupled features $\{\overrightarrow{f_1}, \overrightarrow{f_2}, \overrightarrow{f_3}, \overrightarrow{f_4}\}$ are used to calculate the similarity of adjacent regions, and similar regions are merged based on selective search. The final similarity is calculated as

$$Sim(r_i, r_j) = \sum_{i=1}^{4} \overrightarrow{f_i}. \quad (10)$$

The suspicious tampered regions can be obtained by using selective search. We then mark the pixel values in the suspicious regions as 1. Pixel values 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 to obtain the tampering probability map P_i.

Multiplicative fusion module: In this module, to get more diversified forensic information, we integrate the tampering probability maps P_{spa}, P_{res}, and P_{wav} obtained from the spatial domain, residual domain, and wavelet domain with the given image. The multiplicative fusion can be expressed as follows.

$$I_{fuse} = P_{spa} \times P_{res} \times P_{wav} + I_{spa}. \quad (11)$$

Through the proposed multi-domain probability estimation, the lossy noise can be suppressed and coarse-grained image forgery localization can be realized.

B. Forgery Detector

The forgery detector consists of low-level feature extraction layer, high-level feature extraction layer, and regional edge difference learning layer. In the forgery detector, we mine richer forensic information to achieve accurate tampered region localization. Precisely, to capture tampering traces of various operations and learn universal features, the constrained convolution [16] is exploited in the low-level feature extraction layer, which can suppress the complex image content. We feed the output of the multi-domain probability estimation module to the low-level feature extraction layer. As a result, the prediction error of the multi-domain probability estimation module and the changes in local pixel relationships caused by different tampering operations would be learned adaptively.

Since these low-level features may be fragile, we capture richer semantic tampering information via high-level feature extraction. The residual block [17] is adopted in the high-level feature extraction layer. By stacking residual blocks, feature details and global tampering information can be better captured. The combined use of low-level and high-level features enables PRest-Net to obtain forensic clues from both local edge and macroscopic perspectives.

To enhance the local edge difference between tampered region and real region, in the regional edge difference learning layer, we input the deep feature extracted earlier into the convolution layer followed by the RoI pooling layer and the connection layer, where the convolution layer utilizes the convolution block in region proposal network [18]. After the softmax layer, PRest-Net will judge whether the corresponding region is the tampered region according to the edge difference learned, thus completing the fine-grained tampered region prediction.

III. Experimental Evaluations

A. Experimental Settings

Implement Details. The proposed network is implemented with the TensorFlow deep learning framework. Cross entropy loss and smooth L_1 loss are employed as the training loss function. 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. We optimize our model by utilizing the `SGD` optimizer with default hyperparameters.

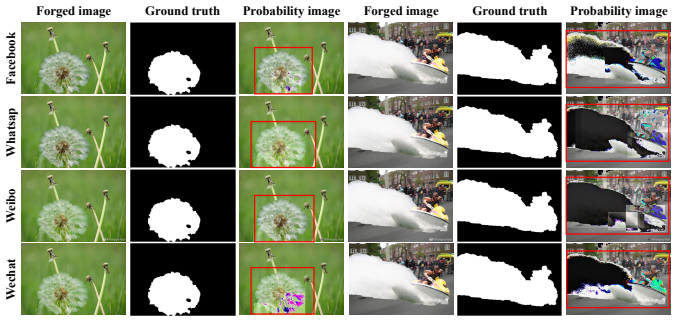

Fig. 2. Visualization results of multi-domain probability estimation. Images are selected from the NIST dataset. The rectangular box marks the estimated tampered region, whose average coincidence error with the actually tampered region in the ground truth is 0.091.

ALL experiments are carried out with a single NVIDIA 2080 Ti GPU server.

Datasets. We adopt the training dataset as [13] to train our model, which contains 9984 forged images. For evaluations, we compare PRest-Net with current state-of-the-art methods on the NIST [19] dataset (containing 564 forged images) and the DSO [20] dataset (containing 100 forged images). To evaluate the robustness of PRest-Net against the online social network transmission, the NIST and DSO dataset are transmitted through the four most popular online social networks: Facebook, Whatsapp, Weibo, and Wechat.

SOTA Models. We compare our method with five state-of-the-art methods: **MT-Net** [5], **NoiPri** [6], **ForSim** [7], **DFCN** [8], **RIFD** [13].

Evaluation Metric. The pixel-level F_1 score and Area Under the receiver operating characteristic Curve (AUC) are utilized to measure the image forgery detection performance.

B. Simulation of Multi-domain Probability Estimation

In order to intuitively illustrate the help of our multi-domain probability estimation method for OSN shared image forgery detection, we perform a visual simulation of multi-domain probability estimation. As demonstrated in Fig. 2, the forged image is transmitted over Facebook, Whatsapp, Weibo, and Wechat. The tampering probability of different image regions can be calculated using multi-domain probability estimation. We fuse the probability maps with the given image at the pixel level based on multiplicative fusion to generate the visual tampering probability image. The tampered region is marked with rectangular boxes, which is shown in the third and sixth columns of Fig. 2. By comparing the ground truth and probability image, it can be found that the tampered region predicted by probability estimation has a high coincidence degree with ground truth. Furthermore, to quantify the coincidence degree of the predicted tampered region and the actually tampered region, we calculate the mean absolute percentage error $MAPE$ between the tampered region obtained by probability estimation and ground truth as follows,

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{coor_{real} - coor_{pred}}{coor_{real}}\right|, \tag{12}$$

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 $MAPE$ between the tampered regions predicted by our probability estimation and the actually tampered regions on Facebook, Whatsapp, Weibo, and Wechat is 0.072, 0.082, 0.126, and 0.082, respectively. These results indicate that the coarse-grained forgery detection for images transmitted over Facebook, Whatsapp, Weibo, and Wechat can be realized by utilizing the proposed multi-domain probability estimation, although different online social networks will use diverse lossy operations to manipulate uploaded images. The reason is that we comprehensively consider the multi-domain information fusion and the coupling relationships between multiple forensics features so that 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.

TABLE I

COMPARISON OF ABLATION STUDY: DETECTION EVALUATION WITH AND WITHOUT (W/OUT) PROBABILITY ESTIMATION (PREST). THE TESTING IMAGES ARE FROM THE NIST DATASET. "WITH SPATIAL", "WITH RESIDUAL", AND "WITH WAVELET" INDICATE THAT ONLY THE CORRESPONDING SINGLE IMAGE DOMAIN IS USED TO PERFORM PROBABILITY ESTIMATION. "-" MEANS THAT THE TESTING FORGED IMAGES HAVE NOT BEEN TRANSMITTED OVER OSNS.

Metric	Method	-	Facebook	Whatsapp	Weibo	Wechat
F_1	w/out PRest	0.41	0.40	0.36	0.40	0.39
	with spatial	0.41	0.40	0.40	0.41	0.40
	with residual	0.40	0.41	0.39	0.41	0.40
	with wavelet	0.40	0.40	0.40	0.41	0.40
	with PRest	**0.42**	**0.42**	**0.41**	**0.42**	**0.42**
AUC	w/out PRest	0.70	0.69	0.66	0.69	0.68
	with spatial	0.71	0.71	0.70	0.71	0.70
	with residual	0.71	0.72	0.69	0.72	0.70
	with wavelet	0.71	0.71	0.71	0.71	0.70
	with PRest	**0.73**	**0.73**	**0.72**	**0.73**	**0.72**

C. Ablation Study

We perform an ablation study to validate the effectiveness of the proposed multi-domain probability estimation method, which reduces the impact of lossy noise generated by online social network transmission on image forgery detection. The comparison results are shown in Table I. We can find that

TABLE II
COMPARISON AGAINST SOTA METHODS BY USING F_1 AND AUC AS CRITERIA. THE FORGED IMAGES ARE TRANSMITTED OVER FACEBOOK, WHATSAPP, WEIBO, AND WECHAT.

OSNs	Method	NIST		DSO	
		F_1	AUC	F_1	AUC
Facebook	**MT-Net** [5]	0.13	0.55	0.07	0.53
	NoiPri [6]	0.06	0.58	0.24	0.50
	ForSim [7]	0.14	0.58	0.37	0.69
	DFCN [8]	0.21	0.63	0.22	0.55
	RIFD [13]	0.33	0.72	**0.46**	**0.72**
	PRest-Net	**0.42**	**0.73**	0.44	**0.72**
Whatsapp	**MT-Net** [5]	0.11	0.53	0.07	0.53
	NoiPri [6]	0.07	0.58	0.24	0.49
	ForSim [7]	0.14	0.59	0.23	0.54
	DFCN [8]	0.23	0.65	0.31	0.65
	RIFD [13]	0.31	0.71	0.34	**0.67**
	PRest-Net	**0.41**	**0.72**	**0.42**	**0.67**
Weibo	**MT-Net** [5]	0.13	0.55	0.07	0.52
	NoiPri [6]	0.05	0.58	0.23	0.50
	ForSim [7]	0.15	0.58	0.26	0.57
	DFCN [8]	0.22	0.64	0.30	0.64
	RIFD [13]	0.29	0.70	0.38	0.69
	PRest-Net	**0.42**	**0.73**	**0.43**	**0.71**
Wechat	**MT-Net** [5]	0.11	0.54	0.07	0.53
	NoiPri [6]	0.04	0.58	0.24	0.50
	ForSim [7]	0.14	0.58	0.25	0.56
	DFCN [8]	0.23	0.64	0.30	0.65
	RIFD [13]	0.28	0.69	0.37	0.68
	PRest-Net	**0.42**	**0.72**	**0.44**	**0.72**

TABLE III
COMPARISON OF DETECTION TIME EFFICIENCY (UNIT: SECONDS) BETWEEN PREST-NET AND **RIFD** ON DIFFERENT DATASETS.

Dataset	Method	Facebook	Whatsapp	Weibo	Wechat
NIST	**RIFD** [13]	2566.47	555.19	1357.53	1349.10
	PRest-Net	**1320.44**	**300.32**	**540.18**	**570.27**
DSO	**RIFD** [13]	539.42	113.54	257.91	260.38
	PRest-Net	**165.30**	**43.60**	**79.82**	**80.43**

the forgery localization performance of the proposed network with probability estimation (with PRest) outperforms that without probability estimation (w/out PRest). Furthermore, using the spatial domain, residual domain, and wavelet domain to estimate the tampering probability map can improve the localization performance. This is because the spatial domain-based probability estimation can effectively reflect the pixel discontinuity caused by image forgery, the residual domain-based probability estimation can strengthen the edge difference between the tampered region and the real region, and the wavelet domain-based probability estimation can extract the difference between useful information and lossy noise information.

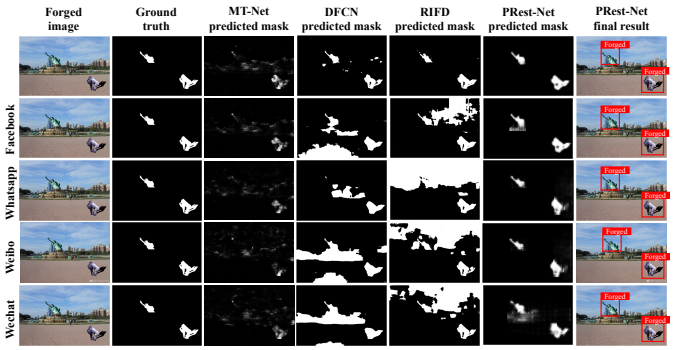

Fig. 3. Qualitative comparison results of forgery localization using different methods.

The expression of forensic knowledge can be strengthened through the multiplicative fusion of multi-domain tampering probability maps, hence, the lossy noise is suppressed and the detection performance is further improved by utilizing the multi-domain probability estimation method (with PRest).

D. Quantitative Comparisons With SOTA Methods

In this experiment, we test the robustness of PRest-Net and compare it with **MT-Net** [5], **NoiPri** [6], **ForSim** [7], **DFCN** [8], **RIFD** [13]. Table II shows the comparison results on different datasets. As seen from this table, the proposed PRest-Net can achieve better forgery detection performance compared with these SOTA methods. The reason why PRest-Net can achieve better performance may be that PRest-Net mines tampering clues from the spatial domain, residual domain, and wavelet domain, respectively, and uses 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 features learning and macroscopic features learning, which enhances the representation ability of PRest-Net and thus improves the localization performance.

In addition, considering the detection time efficiency, Table III shows the comparison results between **RIFD** and the proposed PRest-Net. We can find that 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.

E. Qualitative Comparisons With SOTA Methods

Fig. 3 provides the qualitative results for comparison of the proposed PRest-Net, **MT-Net** [5], **DFCN** [8], and **RIFD** [13]. We can find that the SOTA methods perform well in cases where forged images are not uploaded to online social networks. However, in cases of different OSN transmitted images, their detection performance decreased because the lossy operations weaken the semantic tampering traces. In contrast, the proposed PRest-Net can capture more robust forensic clues and thereby obtain more precise localization results over

various OSN transmission due to the multi-domain probability estimation.

IV. Conclusion

In this paper, we propose PRest-Net for the forgery localization of online social network shared images. The localization issue is solved from the perspective of multi-domain information acquisition, which introduces a probability estimation idea to suppress the lossy influence caused by online social network transmission on the forged images and learn the most differentiated regional features. The designed network architecture is capable of combining low-level features with higher-level features and focuses on the overall and local detail changes caused by various tampering operations, which significantly improves the forgery detection performance. Experimental results show that PRest-Net has competitive and robust localization performance in different online social network transmission cases.

References

[1] H. Xie, J. Ni, and Y.-Q. Shi, "Dual-domain generative adversarial network for digital image operation anti-forensics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1701–1706, 2022.

[2] X. Liu, W. Lu, Q. Zhang, J. Huang, and Y.-Q. Shi, "Downscaling factor estimation on pre-jpeg compressed images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 618–631, 2020.

[3] W. Wang, J. Dong, and T. Tan, "Effective image splicing detection based on image chroma," in *IEEE International Conference on Image Processing (ICIP)*, 2009.

[4] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 507–518, 2015.

[5] Y. Wu, W. AbdAlmageed, and P. Natarajan, "Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[6] D. Cozzolino and L. Verdoliva, "Noiseprint: A cnn-based camera model fingerprint," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 144–159, 2020.

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

[8] P. Zhuang, H. Li, S. Tan, B. Li, and J. Huang, "Image tampering localization using a dense fully convolutional network," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2986–2999, 2021.

[9] F. Guillaro, D. Cozzolino, A. Sud, N. Dufour, and L. Verdoliva, "Trufor: Leveraging all-round clues for trustworthy image forgery detection and localization," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

[10] J. Chen, X. Liao, W. Wang, Z. Qian, Z. Qin, and Y. Wang, "Snis: A signal noise separation-based network for post-processed image forgery detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 2, pp. 935–951, 2023.

[11] C. Chen, B. Li, R. Cai, J. Zeng, and J. Huang, "Distortion model-based spectral augmentation for generalized recaptured document detection," *IEEE Transactions on Information Forensics and Security*, vol. 19, pp. 1283–1298, 2024.

[12] X. Kang, P. Su, Z. Huang, Y. Chen, and J. Wang, "Double compression detection based on the de-blocking filtering of hevc videos," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

[13] H. Wu, J. Zhou, J. Tian, and J. Liu, "Robust image forgery detection over online social network shared images," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[14] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, pp. 154–171, 2013.

[15] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, vol. 59, pp. 167–181, 2004.

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

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

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

[20] T. J. De Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. de Rezende Rocha, "Exposing digital image forgeries by illumination color classification," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 7, pp. 1182–1194, 2013.