Edge-aware Regional Message Passing Controller for Image Forgery Localization

Dong Li*, Jiaying Zhu*, Menglu Wang, Jiawei Liu†, Xueyang Fu, Zheng-Jun Zha
University of Science and Technology of China, China
{dongli6, zhujy53, vault}@mail.ustc.edu.cn, {jwliu6, xyfu, zhazj}@ustc.edu.cn

Abstract

Digital image authenticity has promoted research on image forgery localization. Although deep learning-based methods achieve remarkable progress, most of them usually suffer from severe feature coupling between the forged and authentic regions. In this work, we propose a two-step Edge-aware Regional Message Passing Controlling strategy to address the above issue. Specifically, the first step is to account for fully exploiting the edge information. It consists of two core designs: context-enhanced graph construction and threshold-adaptive differentiable binarization edge algorithm. The former assembles the global semantic information to distinguish the features between the forged and authentic regions, while the latter stands on the output of the former to provide the learnable edges. In the second step, guided by the learnable edges, a region message passing controller is devised to weaken the message passing between the forged and authentic regions. In this way, our ERMPC is capable of explicitly modeling the inconsistency between the forged and authentic regions and enabling it to perform well on refined forged images. Extensive experiments on several challenging benchmarks show that our method is superior to state-of-the-art image forgery localization methods qualitatively and quantitatively.

1. Introduction

The forged or manipulated images pose risks in various fields, such as removing copyright watermarks, generating fake news, and even falsifying evidence in court [32]. The growing technology of forgery will cause a crisis of trust and affect social equity. Therefore, the detection of image forgery is of great significance. The crucial aspect of the detection is to model the inconsistency between the forged and authentic regions and to locate the forged regions on the suspicious image, i.e., **image forgery localization** (IFL). However, as post-processing techniques such as GAN [16, 26, 63], VAE [27, 44] and homogeneous manipulation [7, 33] are wildly utilized, images can be easily tampered in a visually imperceptible way. These techniques constantly aim to couple the forged and authentic regions' features, making image forgery localization challenging. Therefore, in order to accurately locate the image forgery region, it is particularly essential to decouple the features between forged and authentic regions.

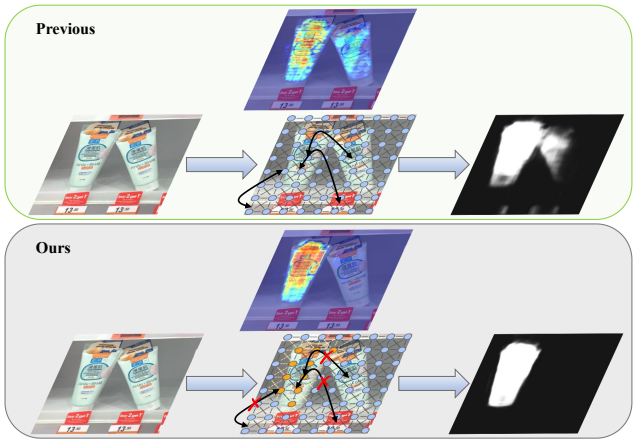

Figure 1. The difference between previous methods and ours. Our method controls the message passing between the forged and authentic regions, ignored by the previous method. As shown by ✗.

In recent years, deep-learning techniques have attracted more attention [23, 57, 58, 65, 66]. Due to the development of deep learning, image forgery localization has achieved remarkable results. For example, ManTra-Net [55] treats the forgery localization problem as a local anomaly detection problem and proposes a novel long short-term memory solution to assess local anomalies. To discriminate between heterologous regions, SPAN employs CNNs to extract anomalous local noise features from noise maps. MVSS-Net [5] learns a multiview feature with multi-scale

* Co-first authors contributed equally, † corresponding author. This work was supported by the National Key R&D Program of China under Grant 2020AAA0105702, the National Natural Science Foundation of China (NSFC) under Grants 62225207, 62276243, 62106245 and U19B2038, the University Synergy Innovation Program of Anhui Province under Grant GXXT-2019-025.

supervised networks to jointly exploit the noise view and the boundary artifacts. However, these methods do not decouple the features between the forged and authentic regions, making it difficult to locate the tampered regions for elaborately forged images accurately. As shown in Figure 1, in the previous method, the features of the forged region are coupled with some of the features of the authentic region, resulting in wrong localization.

In this work, we propose a novel method to avoid feature coupling of the two regions (forged and authentic) for image forgery localization. One of the keys of this method is to construct a dynamic graph, where the edges between the forged and authentic regions participate in its construction. We control the message passing of regions inside and outside the edges (i.e., forged and authentic) by reconstructing the adjacency matrix of nodes inside and outside the edges, thus achieving effective disentanglement of the features between the forged and authentic regions. Based on its functionality, the edge-aware dynamic graph convolution is named the regional message passing controller (RMPC).

In order to use the proposed RMPC for image forgery localization, it is necessary to obtain the edge information between the forged and authentic regions, which is the other key of this method. Therefore, an edge reconstruction (ER) module was developed, including a context-enhanced graph (CEG) and a threshold-adaptive differentiable binarization module. We specially design an adjacency matrix learner in the CEG, which encodes global information along the nodes to assemble the global semantic information. Inspired by the Sigmoid function, we develop the threshold-adaptive differentiable binarization edge algorithm, which stands on the output of the CEG to provide the learnable edges.

In summary, in this work we propose a new two-step framework named Edge-aware Regional Message Passing Controller (ERMPC) for Image Forgery Localization, including RMPC and ER. The ERMPC could effectively control the message passing between the forged and authentic regions to achieve effective disentanglement of the two regions, thus boosting the performance of image forgery localization. We take edge information as the main task and use it as a basis to explicitly model the inconsistency between two regions. To the best of our knowledge, this work is the first attempt to explicitly weaken the message passing between the forged and authentic regions. Our contributions are as follows:

- We propose ERMPC, a novel two-step coarse-to-fine framework for image forgery localization, explicitly modeling the inconsistency between the forged and authentic regions with edge information.
- We propose an edge-aware dynamic graph, also known as RMPC, to control the message passing between two regions (forged and authentic) in the feature map.
- We develop an edge reconstruction module containing a context-enhanced graph and a threshold-adaptive differentiable binarization module to obtain the desired edge information.
- We conduct extensive experiments on multiple benchmarks and demonstrate that our method is qualitatively and quantitatively superior to state-of-the-art image forgery localization methods.

2. Related Works

2.1. Image Forgery Localization

Most early works propose to localize a specific type of forgery, including splicing [2, 3, 9, 10, 24, 28, 37, 51, 60], copy-move [8, 14, 25, 48, 52, 54], and removal [1, 50, 56, 64]. While these methods demonstrate satisfactory performance in detecting specific types of forgery, they exhibit limitations in practical applications due to the prevalence of unknown and diverse forgery types. Consequently, recent studies have emphasized the need for an approach to tackle multiple forgery types with one model. RGB-N [62] proposes a dual-stream Faster R-CNN network. The first stream is designed to extract RGB features and identify tampering artifacts, while the second stream utilizes noise features to model the noise inconsistency between tampered and authentic regions, thereby localizing image forgery with enhanced accuracy. ManTra-net [55] is an end-to-end network that performs both detection and localization, which treats the problem as anomaly detection and introduces a long short-term memory solution to assess local anomalies. SPAN [22] attempts to model the spatial correlation by constructing a pyramid of local self-attention blocks. MVSS-Net [5] has designed an edge-supervised branch that uses edge residual blocks to capture fine-grained boundary detail in a shallow to deep manner. It is worth noting that edge information is important for image forgery detection and localization, as the tampered regions are commonly surrounded by unnatural artifacts. Nevertheless, most studies only utilize edges in a supervised strategy, such as MVSS-Net [5], MFCN [41], GSR-Net [61], and CAIFL [47]. PSCCNet [35] uses a progressive spatial-channel correlation module that uses features at different scales and dense cross-connections to generate operational masks in a coarse-to-fine fashion. ObjectFormer [46] captures forgery traces by extracting high-frequency parts of the image and combining them with RGB features. In this work, we explicitly employ learnable edges as guiding information to impede the message passing between the forged and untouched regions and employ refined edge reconstruction to achieve inconsistent modeling of the two regions to localize manipulation artifacts.

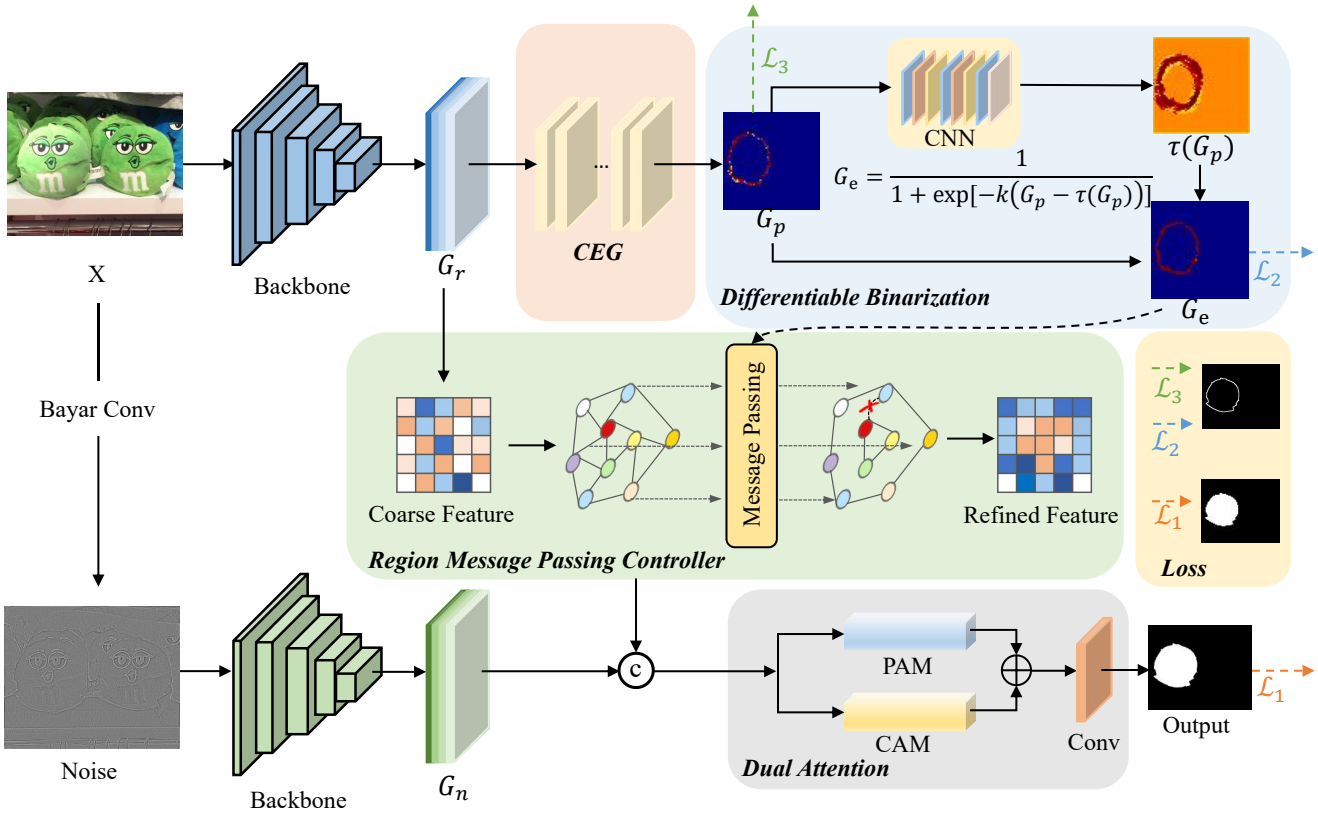

Figure 2. An overview of the proposed framework ERMPC. The input is a suspicious image ($H \times W \times 3$), and the output is a predicted mask (H×W×1), which localizes the forged regions.

2.2. Graph Reasoning

Graph-based approaches have attracted growing attention from the computer vision community these days and have shown to be an effective relational reasoning method with a robust capacity for non-local feature aggregation. Specifically, CDGCNet [21] utilizes a class-wise learning strategy and designs a class-wise graph convolution, which avoids heavy graph concatenation and facilitates feature learning. To overcome the problem that most graph-based approaches assign a fixed number of neighbors for each query item, DAGL [39] proposes a dynamic attentive graph learning model to explore the dynamic non-local property for image restoration. CTL [34] learns a context-reinforced topology to construct multi-scale graphs which have a robust representational capacity by considering both global contextual information and physical connections of the human body. On other tasks in computer vision, e.g., object detection [43], multi-label image recognition [6], and skeleton-based action recognition [36, 59], graph convolutional neural networks have also achieved impressive performance. In contrast, we propose an improved edge-guided graph attention model to obstruct the message passing between forged and untouched regions with dynamic construction for forgery region localization. As graph convolution is an effective relational reasoning method that is very suitable for detecting forgery traces, this paper applies it to the task of image forgery localization for the first time. An improved edge-guided graph attention model is proposed for forgery region localization, which obstructs the message passing between forged and untouched regions with dynamic construction.

3. Methodology

This section details the scheme of image forgery localization based on edge-aware message passing control. Sec. 3.1 describes an overview of the framework. One of the keys of the scheme is to control the message passing with edge and thus model the inconsistency between the forged and authentic regions (Sec. 3.2). Another key lies in reconstructing accurate edges from coarse features (Sec. 3.3). In addition, following [5, 22, 55, 62], we use the noise branch and fuse it with RGB branch at the end of the network (Sec. 3.4). The optimization is introduced in Sec. 3.5.

3.1. Overview

Figure 2 is an overview of the framework. The input image is represented as $X \in \mathbb{R}^{H\times W\times 3}$, where H and W represent the height and width of the image. First, we use two branches to handle RGB and noise separately, obtaining $G_r \in \mathbb{R}^{H_s\times W_s\times C_s}$ and $G_n \in \mathbb{R}^{H_s\times W_s\times C_s}$ respectively. Then, we use ResNet-50 pretrained on ImageNet [11] as the backbone network. Following [4], we adopt atrous spatial pyramid pooling(ASPP) together with ResNet-50 to capture the long-range contextual information. The coarse features extracted from RGB branches are converted into edges through the edge reconstruction block. Meanwhile, the coarse features are constructed as graph structures, guided by the reconstructed edge information. Finally, the RGB features after the graph convolutional network are fused with noise information through dual attention [5] to output the predicted forgery localization map.

3.2. Region Message Passing Controller

Most forged images are carefully processed to hide tampering artifacts, making it challenging to model inconsistencies in the RGB branch. To overcome this problem, the edges between the forged and authentic regions are utilized to control the message passing explicitly.

The edge feature $G_e \in \mathbb{R}^{H_e\times W_e\times 1}$ comes from the edge reconstruction block (described in detail in the next subsection), where $H_e = H_s$ and $W_e = W_s$. First, we calculate the relationship between two node features P_i, P_j of the G_E using an algorithm similar to XNOR gate:

$$\mathrm{XN}\left(P_i, P_j\right) = \begin{cases} 0 & \text{Inside and outside the edges} \\ 1 & \text{On the same side of edges.} \end{cases} \tag{1}$$

If two nodes are inside and outside the edge respectively, then their XN is set to 0. For each of the N ($N = H_e \times W_e$) nodes features, we calculate its XN, thus generating a matrix $A_e \in \mathbb{R}^{N\times N}$.

Next, we apply graph learning to process the $G_r \in \mathbb{R}^{H_s\times W_s\times C_s}$. Following GAT [45], we calculate the similarity between two nodes as attention coefficients:

$$\alpha_{i,j} = \psi\left(x_i\right)^T \psi'\left(x_j\right), \tag{2}$$

where ψ, ψ' denotes two learnable linear transformations. Specifically, we utilize $\psi = Wx$ and $\psi' = W'x$, where $W \in \mathbb{R}^{C_s\times C_s}$ and $W' \in \mathbb{R}^{C_s\times C_s}$ are both weight matrixes. To make coefficients easier to compare across different nodes, we normalize them using the softmax function:

$$A_{r_{i,j}} = \frac{exp\left(\alpha_{i,j}\right)}{\sum_{j=1}^{N} exp\left(\alpha_{i,j}\right)}, \tag{3}$$

where $A_r \in \mathbb{R}^{N\times N}$ is the preliminary adjacency matrix. It reflects the relationship between any two nodes in the feature map. Larger values represent a greater flow of information between two nodes. In order to realize message passing control better, a method of dynamically adjusting the adjacency matrix is adopted. Specifically, if two nodes are on and off the boundary respectively, their adjacency is broken due to the dynamic adjustment of the adjacency matrix. In practice, the adjacency matrix is recalculated as follows:

$$A'_r = A_r \odot A_e, \tag{4}$$

where $\odot$ is the Hadamard product. $A'_r \in \mathbb{R}^{N\times N}$ re-models the weights between nodes in the feature map and severs the connections between the forged and real regions.

Once the adjacency matrix is obtained, it is weighted by the learnable attention weights. Then the original nodes are updated with the following:

$$Z_r = \mathrm{ReLU}\left(A'_r G'_r W_z\right), \tag{5}$$

where $Z_r \in \mathbb{R}^{N\times C_s}$ is RGB features after graph reasoning, $W_z \in \mathbb{R}^{C_s\times C_s}$ is the learnable parameter and $G'_r \in \mathbb{R}^{N\times C_s}$ is the graph representation transformed by G_r.

It is worth noting that this work not only introduces the idea of controlling the message passing in image forgery localization for the first time, but also the approach implemented is different from previous research [12, 20]. BFP [12] uses directed acyclic graphs (DAG) for feature propagation and introduces boundary information into the propagation to control the message passing between different segments. However, it has to scan the image pixel by pixel and requires a lot of loops, causing it difficult to implement in real-world applications. BGC [20] emphasizes reducing the weights of the edges and does not really focus on the message passing between regions.

3.3. Edge Reconstruction

As described in the previous section, the Edge-aware Message Passing Control Graph requires accurate edge information. Designing such an edge-access network is nontrivial. The main challenge is how to learn edge information from the coarse feature. To this end, a novel approach to edge reconstruction is proposed. Specifically, we started by using the Sobel layer [5] to enhance the edge-related patterns in $G_r \in \mathbb{R}^{H_s\times W_s\times C_s}$,

$$G_c = G_r \odot \sigma\left(\mathrm{Norm}\left(\mathrm{Sobel}(G_r)\right)\right), \tag{6}$$

where Norm is L2 Normalization, σ is Sigmoid and Sobel is the SobelConv [5].

Then, we explore a Context-Enhanced Graph (CEG), extracting the local and global features separately for $G_c \in \mathbb{R}^{H_s\times W_s\times C_s}$. Specifically, the local information is extracted by the convolutional layers [5]. For global information, the contextual information of the feature map is encoded as the adjacency matrix $A_c \in \mathbb{R}^{N\times N}$ in a simple and

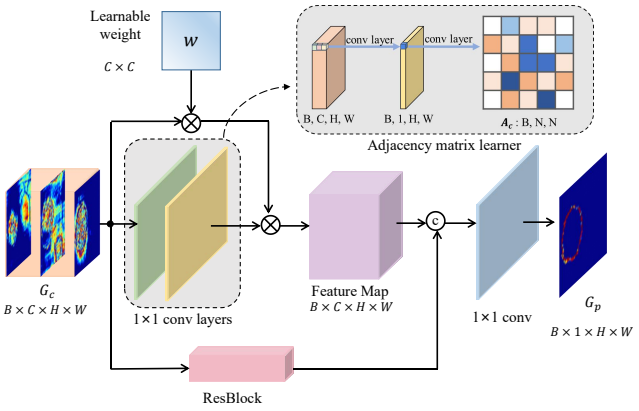

Figure 3. Context-Enhanced Graph. Global features are extracted by the specific graph, while local features by the ResBlock.

efficient way. As shown in Figure 3, A_c is generated by a specially designed adjacency matrix learner which consists of two 1×1 convolution layers. The process is written as

$$A_c = \mathrm{Norm}\left(\mathrm{Conv}\left(\mathrm{Conv}\left(\mathrm{G}_c'\right)\right)\right), \tag{7}$$

where $G_c' \in \mathbb{R}^{N \times C_s}$ is reshaped by G_c. Given the node features of G_c', we first squeeze the feature dimensions by a 1×1 convolution layer. Then, another convolution layer with 1×1 kernel is used to transfer the N-dimension feature vector into the $N \times N$ adjacency matrix. In addition, L2 Normalization is applied to each row of A_c to facilitate stable optimization. Next, we use this adjacency matrix to complete the graph reasoning. In this way, we can obtain a global feature map. The process can be written as

$$\mathrm{Global}\left(G_c'\right) = A_c G_c' W_c, \tag{8}$$

where $W_c \in \mathbb{R}^{C_s \times C_s}$ is the learnable parameter. To match with the local information dimension, we reshape $\mathrm{Global}\left(G_c'\right) \in \mathbb{R}^{N \times C_s}$ back to $\mathrm{Global}'\left(G_c'\right) \in \mathbb{R}^{H_s \times W_s \times C_s}$. Then combining the local and global information, we can obtain the edge probability map $G_p \in \mathbb{R}^{H_s \times W_s \times 1}$:

$$G_p = \sigma\left(\mathcal{C}\left(\mathrm{Cat}\left(\mathrm{Local}\left(G_c\right), \mathrm{Global}'\left(G_c'\right)\right)\right)\right), \tag{9}$$

where σ denotes the Sigmoid, $\mathcal{C}$ is a 1×1 convolution layer, and *Local* contains ReLU and two convolutional layers.

Furthermore, in order to determine the edges, a threshold needs to be determined to binarize the probability map. Most previous studies have used fixed thresholds, and the process is non-differentiable. Inspired by [30], we utilize a variable threshold map which is adaptive for each point on the probability map G_p. Besides, we propose a variant of the Sigmoid function to complete the binarization, which is capable of participating in the backpropagation involved. Binarize threshold adjustment is achieved by translating the Sigmoid function along the x-axis. Therefore, we explore the threshold-adaptive differentiable binarization (TDB) for edge reconstruction. It is computed as follows:

$$G_e = \frac{1}{1 + e^{-k(G_p - \tau(G_p))}}, \tag{10}$$

where τ denotes the learnable transformation, which in practice is 3×3 convolution operators, and k is the amplifying factor. In particular, k is set to 500 empirically.

3.4. Branch Fusion

Following most studies [5,22,55,62], we also employ the noise branch. However, this is not the focus of this work, so we have used some common methods. As in Figure 2, the noise is extracted with BayarConv [55]. For the fusion of the two branches, we follow [5] and adopt the Dual Attention (DA) [15]. DA includes channel attention module (CAM) and position attention module (PAM). It can effectively fuse two branches. The process can be written as

$$G_o = \mathrm{DA}(G_z, G_n), \tag{11}$$

where $G_z \in \mathbb{R}^{H_s \times W_s \times C_s}$ is reshaped by the Z_τ. Finally, we transform $G_o \in \mathbb{R}^{H_s \times W_s \times 1}$ with bilinear upsampling into the final predicted mask $G_{out} \in \mathbb{R}^{H \times W \times 1}$.

3.5. Optimization

As shown in Figure 2, we calculate the loss function for three components: the final prediction $G_{out} \in \mathbb{R}^{H \times W \times 1}$, the binary edge prediction $G_e \in \mathbb{R}^{H_e \times W_e \times 1}$ and the edge probability map $G_p \in \mathbb{R}^{H_e \times W_e \times 1}$. For edge loss, the ground-truth edge $E \in \mathbb{R}^{H \times W \times 1}$ is downsampled to a smaller size $E' \in \mathbb{R}^{H_e \times W_e \times 1}$to match G_e, G_p. This strategy outperforms upsampling G_e, G_p in terms of computational cost and performance. The overall loss function can be written as:

$$\mathbf{L} = \lambda_1 \mathcal{L}\left(\mathrm{Y}, \mathrm{G_{out}}\right) + \lambda_2 \mathcal{L}\left(\mathrm{E}', \mathrm{G_e}\right) + \lambda_3 \mathcal{L}\left(\mathrm{E}', \mathrm{G_p}\right), \tag{12}$$

where $\mathcal{L}$ denotes the Dice loss [5], $Y \in \mathbb{R}^{H \times W \times 1}$ is the ground-truth mask, and $\lambda_1, \lambda_2, \lambda_3$ are the parameters to balance the three terms in loss function ($\lambda_1 + \lambda_2 + \lambda_3 = 1$). In our setting, they are set as 0.50, 0.25 and 0.25 respectively.

4. Experiments

4.1. Experimental Setup

Pre-training Data We create a sizable image tampering dataset and use it to pre-train our model. This dataset includes three categories: 1) splicing, 2) copy-move, 3) removal. For splicing, we use the MS COCO [31] to generate spliced images, where one annotated region is randomly selected per image and pasted into a different image after several transformations. We use the same transformation as [5],

Method	Data	Columbia	Coverage	CASIA	NIST16	IMD20
ManTraNet	64K	82.4	81.9	81.7	79.5	74.8
SPAN	96k	93.6	92.2	79.7	84.0	75.0
PSCCNet	100k	**98.2**	84.7	82.9	85.5	80.6
ObjectFormer	62K	95.5	92.8	84.3	87.2	82.1
Ours	60K	96.8	**94.4**	**87.6**	**89.5**	**85.6**

Table 1. Comparisons of manipulation localization AUC (%) scores of different pre-trained models.

Methods	Coverage		CASIA		NIST16	
	AUC	F1	AUC	F1	AUC	F1
J-LSTM	61.4	-	-	-	76.4	-
H-LSTM	71.2	-	-	-	79.4	-
RGB-N	81.7	43.7	79.5	40.8	93.7	72.2
SPAN	93.7	55.8	83.8	38.2	96.1	58.2
PSCCNet	94.1	72.3	87.5	55.4	99.1	74.2
ObjectFormer	95.7	75.8	88.2	57.9	99.6	82.4
Ours	**98.4**	**77.3**	**90.4**	**58.6**	**99.7**	**83.6**

Table 2. Comparison of manipulation localization results using fine-tuned models.

Distortion	SPAN	ObjectFormer	Ours	
no distortion	83.95	87.18	**89.49**	
Resize(0.78×)	83.24	87.17	**89.33**	**0.16**↓
Resize(0.25×)	80.32	86.33	**87.72**	**1.77**↓
Blur($k = 3$)	83.10	85.97	**89.22**	**0.27**↓
Blur($k = 15$)	79.15	80.26	**87.13**	**2.36**↓
Noise($\sigma = 3$)	75.17	79.58	**88.25**	**1.24**↓
Noise($\sigma = 15$)	67.28	78.15	**83.40**	**6.09**↓
Compress($q = 100$)	83.59	86.37	**89.42**	**0.07**↓
Compress($q = 50$)	80.68	86.24	**88.82**	**0.67**↓

Table 3. Localization performance on NIST16 dataset under various distortions. AUC scores are reported (in %), (Blur: GaussianBlur, Noise: GaussianNoise, Compress: JPEGCompress.)

including the scale, rotation, shift, and luminance changes. Since the spliced region is not always an object, we create random outlines using the Bezier curve [38] and fill them to create splicing masks. For copy-move, the datasets from MS COCO and [53] are adopted. For removal, we adopt the SOTA inpainting method [29] to fill one annotated region that is randomly removed from each chosen MS COCO image. We randomly add Gaussian noise or apply the JPEG compression algorithm to the generated data to resemble the visual quality of images in real situations.

Testing Datasets Following [46], we evaluate our model on CASIA [13] dataset, Coverage [49] dataset, Columbia [19] dataset, NIST16 [17] dataset and IMD20 [40] dataset. Specifically, CASIA [13], which contains two types of tampered images (splicing and copy-move), is widely used in the image forgery field. The database V1.0 has fixed size tampered images, which are generated only by using crop-and-paste operation under Adobe Photoshop. The database v2.0 is more comprehensive and challenging, and most tampered examples are generated with post-processing. COVER [49] provides 100 images, and all of them are generated by copy-move tampering technique. Columbia [19] consists of 180 splicing images, whose size ranges from 757×568 to 1152×568. In detail, the tampered images in Columbia are all uncompressed and without any post-processing. NIST16 [17] is a high-quality dataset. This dataset contains three types of tampering, and some regions of manipulation are difficult for humans to identify. IMD20 [40] collects real-life manipulated images from the Internet, and involves all three manipulations as well. We apply the same training/testing splits as [22, 46, 62] to fine-tune our model for fair comparisons.

Evaluation Metrics To quantify the localization performance, following previous works [22, 46], we use pixel-level Area Under Curve (AUC) and F1 score on manipulation masks. And we adopt the Equal Error Rate (EER) threshold to binarize masks because binary masks are necessary to calculate F1 scores.

Implementation Details The input images are resized to 512×512. In this work, the backbone is ResNet-50 [18], pre-trained on ImageNet [11]. Implemented by PyTorch, our model is trained with GeForce GTX 3090. We use Adam as the optimizer, and the learning rate decays from 10^{-4} to 10^{-7}. We train 100 epochs with a batch size of 8, and the learning rate decays by 10 times every 30 epochs.

4.2. Comparison with the State-of-the-Art Methods

Following SPAN [22] and ObjectFormer [46], our model is compared with other state-of-the-art tampering localization methods under two settings: 1) training on the synthetic dataset and evaluating on the full test datasets, and 2) fine-tuning the pre-trained model on the training split of test datasets and evaluating on their test split. The pre-trained model will demonstrate each method's generalizability, and the fine-tuned model will demonstrate how well each method performs locally once the domain discrepancy has been significantly reduced.

Pre-trained Model Table 1 shows the localization performance of pre-trained models for different methods on four standard datasets under pixel-level AUC. We compare our model ERMPC with MantraNet [55], SPAN [22], PSCCNet [35], and ObjectFormer [46] when evaluating pre-trained models. The pre-trained ERMPC achieves the best localization performance on Coverage, CASIA, NIST16 and IMD20, and ranks second on Columbia. Especially, ERMPC achieves 94.4% on the copy-move dataset

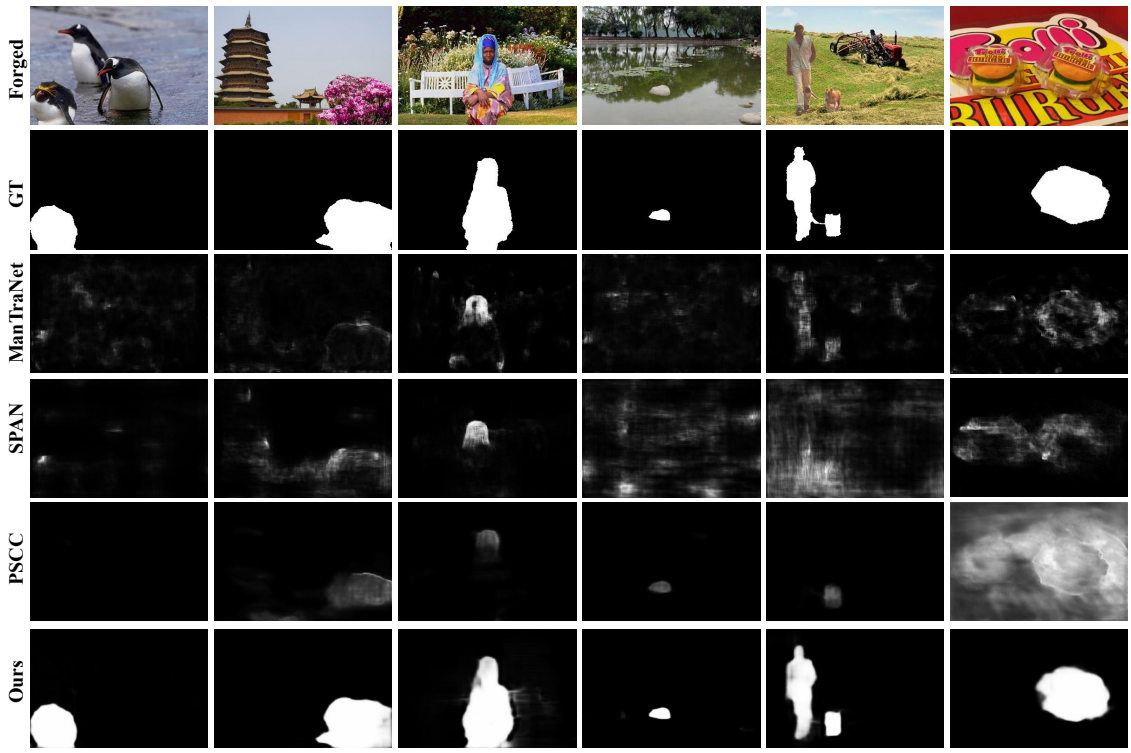

Figure 4. Visualization of the predicted manipulation mask by different methods. From top to bottom, we show forged images, GT masks, predictions of ManTraNet, SPAN, PSCC-Net, and ours.

Variants	CASIA		NIST16	
	AUC	F1	AUC	F1
Baseline	71.6	38.3	77.1	52.6
w/o RMPC	76.9	45.6	86.4	60.7
w/o CEG	85.1	51.5	93.4	75.3
w/o TDB	88.6	57.3	98.2	81.9
Ours	**90.4**	**58.6**	**99.7**	**83.6**

Table 4. Ablation results on CASIA and NIST16 dataset using different variants of ERMPC. AUC and F1 scores (%) are reported.

COVER, whose image forgery regions are indistinguishable from the background. This validates our model owns the superior ability to control message passing between two regions(forged and authentic). We fail to achieve the best performance on Columbia, falling behind PSCCNet 1.4% under AUC. We contend that the explanation may be that the distribution of their synthesized training data closely resembles that of the Columbia dataset. This is further supported by the results in Table 2, which show that ERMPC performs better than PSCCNet in terms of both AUC and F1 scores. Furthermore, it is important to note that ERMPC attains decent results with less pre-training data.

Fine-tuned Model The fine-tuned models, which are initiated with the network weights of the pretrained model, will be trained on the training split of Coverage, CASIA, and

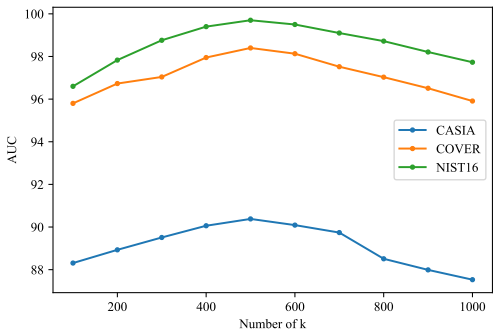

Figure 5. The effect of parameter k in TDB

NIST16 datasets, respectively. We evaluate the fine-tuned models of different methods in Table 2. As for AUC and F1, our model achieves significant performance gains. This validates that ERMPC could capture subtle tampering artifacts by controlling the message passing between two regions (forged and authentic) in the feature map.

4.3. Robustness Evaluation

We degrade the raw manipulated images from NIST16 using the distortion settings in [46] to analyze the robustness of ERMPC for forgery localization. These distortions types include scaling images to various scales (Re-

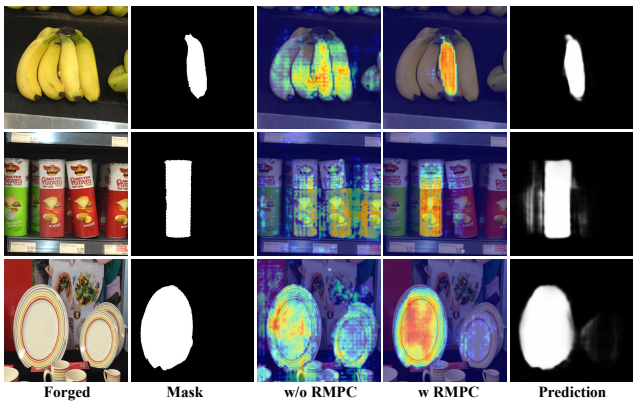

Figure 6. Visualization of message passing controller. From left to right, we display the forged images, masks, GradCAM [42] of the feature map without (w/o) and with (w) RMPC, and predictions.

size), applying Gaussian blur with a kernel size k (GaussianBlur), adding Gaussian noise with a standard deviation σ (GaussianNoise), and performing JPEG compression with a quality factor q (JPEGCompress). We compare the forgery localization performance (AUC scores) of our pre-trained models with SPAN [22] and ObjectFormer on these corrupted data and report the results in Table 3. ERMPC demonstrates more resistance to various distortion techniques. It is worth noting that JPEG compression is commonly performed when uploading images to social media. Our model ERMPC performs significantly better on compressed images than other methods.

4.4. Ablation Analysis

The Region Message Passing Controller (RMPC) module of our method ERMPC is designed to weaken the message passing between the forged and authentic regions. The Context-Enhanced Graph (CEG) encodes global information along the node to obtain a better edge probability map, while the Threshold-adaptive Differentiable Binarization (TDB) adaptively performs the binarization process for learnable edges. In order to evaluate the effectiveness of RMPC, CEG and TDB, we remove them separately from ERMPC and evaluate the forgery localization performance on CASIA and NIST16 datasets.

Table 4 presents the quantitative outcomes. The baseline denotes that we just use ResNet-50. It can be seen that without TDB, the AUC scores decrease by 2.0 % on CASIA and 1.5 % on NIST16, while without CEG, the AUC scores decrease by 5.9 % on CASIA and 6.3 % on NIST16. Furthermore, when RMPC is discarded, serious performance degradation in Table 4, i.e., 14.9% in terms of AUC and 22.2% in terms of F1 on CASIA can be observed.

In Figure 5, we show the different values of parameters k in threshold-adaptive differentiable binarization to verify its effect over three datasets. With its increasing, the curve of binarization becomes steeper. Moreover, being smaller is inadequate for weakening message passing, while being larger will destroy the adaptive capacity of the network. It is obvious that the setting of 500 is the optimal solution.

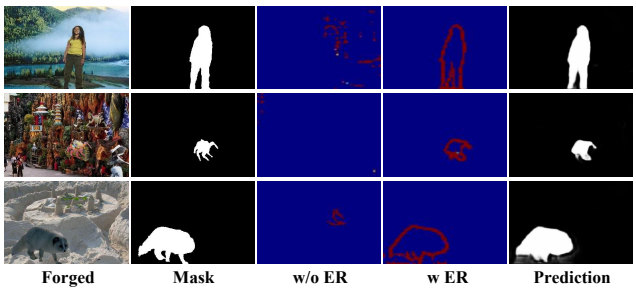

Figure 7. Visualization of edge reconstruction. From left to right, we display the forged images, masks, the features without (w/o) and with (w) the edge reconstruction module, and prediction.

4.5. Visualization Results

Qualitative results. As shown in Figure 4, We provide predicted forgery masks of various methods. Since the source code of ObjectFormer [46] is not available, their predictions are not available. The results demonstrate that our method could not only locate the tampering regions more accurately but also develop sharp boundaries. It benefits from the explicit modeling of inconsistencies and the full exploitation of edges by our method.

Visualization of message passing controller. To verify the usefulness of the region message passing controller (RMPC), we show the change of features before and after the controller in Figure 6. It is clear that RMPC facilitates the learning of forgery features and prevents false alarms. Specifically, the network without RMPC will make false judgments about objects that are similar to the forgery.

Visualization of edge reconstruction. To verify the effect of the edge reconstruction (ER) module, the change of features before and after EG is shown in Figure 7. The results demonstrate that the EG can effectively acquire accurate edges, thus helping our model to perform well.

5. Conclusion

In this paper, we propose a novel Image Forgery Localization framework with a two-step Edge-aware Regional Message Passing Controlling strategy. In detail, the first step is to account for fully exploiting the edge information. In the second step, guided by the learnable edges, an edge-aware dynamic graph is devised to weaken the message passing between the forged and authentic regions. Our paper provides a new research strategy to solve the misjudgment problem in the field of IFL. Extensive experimental results on several benchmarks demonstrate the effectiveness of the proposed algorithm.

References

[1] Mohammed Aloraini, Mehdi Sharifzadeh, and Dan Schonfeld. Sequential and patch analyses for object removal video forgery detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(3):917–930, 2020. 2

[2] Irene Amerini, Tiberio Uricchio, Lamberto Ballan, and Roberto Caldelli. Localization of jpeg double compression through multi-domain convolutional neural networks. In *2017 IEEE Conference on computer vision and pattern recognition workshops (CVPRW)*, pages 1865–1871. IEEE, 2017. 2

[3] Luca Bondi, Silvia Lameri, David Guera, Paolo Bestagini, Edward J Delp, Stefano Tubaro, et al. Tampering detection and localization through clustering of camera-based cnn features. In *CVPR Workshops*, volume 2, 2017. 2

[4] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 4

[5] Xinru Chen, Chengbo Dong, Jiaqi Ji, Juan Cao, and Xirong Li. Image manipulation detection by multi-view multi-scale supervision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14185–14193, 2021. 1, 2, 3, 4, 5

[6] Zhao-Min Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo. Multi-label image recognition with graph convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5177–5186, 2019. 3

[7] Wenyan Cong, Xinhao Tao, Li Niu, Jing Liang, Xuesong Gao, Qihao Sun, and Liqing Zhang. High-resolution image harmonization via collaborative dual transformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18470–18479, 2022. 1

[8] Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Efficient dense-field copy–move forgery detection. *IEEE Transactions on Information Forensics and Security*, 10(11):2284–2297, 2015. 2

[9] Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Splicebuster: A new blind image splicing detector. In *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–6. IEEE, 2015. 2

[10] Davide Cozzolino and Luisa Verdoliva. Noiseprint: a cnn-based camera model fingerprint. *IEEE Transactions on Information Forensics and Security*, 15:144–159, 2019. 2

[11] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 4, 6

[12] Henghui Ding, Xudong Jiang, Ai Qun Liu, Nadia Magnenat Thalmann, and Gang Wang. Boundary-aware feature propagation for scene segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6819–6829, 2019. 4

[13] Jing Dong, Wei Wang, and Tieniu Tan. Casia image tampering detection evaluation database. In *2013 IEEE China Summit and International Conference on Signal and Information Processing*, pages 422–426. IEEE, 2013. 6

[14] Luca D'Amiano, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. A patchmatch-based dense-field algorithm for video copy–move detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(3):669–682, 2018. 2

[15] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019. 5

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 1

[17] Haiying Guan, Mark Kozak, Eric Robertson, Yooyoung Lee, Amy N Yates, Andrew Delgado, Daniel Zhou, Timothee Kheyrkhah, Jeff Smith, and Jonathan Fiscus. Mfc datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pages 63–72. IEEE, 2019. 6

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6

[19] J Hsu and SF Chang. Columbia uncompressed image splicing detection evaluation dataset. *Columbia DVMM Research Lab*, 2006. 6

[20] Hanzhe Hu, Jinshi Cui, and Hongbin Zha. Boundary-aware graph convolution for semantic segmentation. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 1828–1835. IEEE, 2021. 4

[21] Hanzhe Hu, Deyi Ji, Weihao Gan, Shuai Bai, Wei Wu, and Junjie Yan. Class-wise dynamic graph convolution for semantic segmentation. In *European Conference on Computer Vision*, pages 1–17. Springer, 2020. 3

[22] Xuefeng Hu, Zhihan Zhang, Zhenye Jiang, Syomantak Chaudhuri, Zhenheng Yang, and Ram Nevatia. Span: Spatial pyramid attention network for image manipulation localization. In *European conference on computer vision*, pages 312–328. Springer, 2020. 2, 3, 5, 6, 8

[23] Jie Huang, Man Zhou, Yajing Liu, Mingde Yao, Feng Zhao, and Zhiwei Xiong. Exposure-consistency representation learning for exposure correction. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6309–6317, 2022. 1

[24] Minyoung Huh, Andrew Liu, Andrew Owens, and Alexei A Efros. Fighting fake news: Image splice detection via learned self-consistency. In *Proceedings of the European conference on computer vision (ECCV)*, pages 101–117, 2018. 2

[25] Ashraful Islam, Chengjiang Long, Arslan Basharat, and Anthony Hoogs. Doa-gan: Dual-order attentive generative adversarial network for image copy-move forgery detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4676–4685, 2020. 2

[26] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 1

[27] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 1

[28] Vladimir V Kniaz, Vladimir Knyaz, and Fabio Remondino. The point where reality meets fantasy: Mixed adversarial generators for image splice detection. *Advances in Neural Information Processing Systems*, 32, 2019. 2

[29] Jingyuan Li, Ning Wang, Lefei Zhang, Bo Du, and Dacheng Tao. Recurrent feature reasoning for image inpainting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7760–7768, 2020. 6

[30] Minghui Liao, Zhisheng Zou, Zhaoyi Wan, Cong Yao, and Xiang Bai. Real-time scene text detection with differentiable binarization and adaptive scale fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022. 5

[31] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 5

[32] Xun Lin, Shuai Wang, Jiahao Deng, Ying Fu, Xiao Bai, Xinlei Chen, Xiaolei Qu, and Wenzhong Tang. Image manipulation detection by multiple tampering traces and edge artifact enhancement. *Pattern Recognition*, 133:109026, 2023. 1

[33] Jun Ling, Han Xue, Li Song, Rong Xie, and Xiao Gu. Region-aware adaptive instance normalization for image harmonization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9361–9370, 2021. 1

[34] Jiawei Liu, Zheng-Jun Zha, Wei Wu, Kecheng Zheng, and Qibin Sun. Spatial-temporal correlation and topology learning for person re-identification in videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4370–4379, 2021. 3

[35] Xiaohong Liu, Yaojie Liu, Jun Chen, and Xiaoming Liu. Pscc-net: Progressive spatio-channel correlation network for image manipulation detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022. 2, 6

[36] Ziyu Liu, Hongwen Zhang, Zhenghao Chen, Zhiyong Wang, and Wanli Ouyang. Disentangling and unifying graph convolutions for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 143–152, 2020. 3

[37] Siwei Lyu, Xunyu Pan, and Xing Zhang. Exposing region splicing forgeries with blind local noise estimation. *International journal of computer vision*, 110(2):202–221, 2014. 2

[38] Michael E Mortenson. *Mathematics for computer graphics applications*. Industrial Press Inc., 1999. 6

[39] Chong Mou, Jian Zhang, and Zhuoyuan Wu. Dynamic attentive graph learning for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4328–4337, 2021. 3

[40] Adam Novozamsky, Babak Mahdian, and Stanislav Saic. Imd2020: A large-scale annotated dataset tailored for detecting manipulated images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*, pages 71–80, 2020. 6

[41] Ronald Salloum, Yuzhuo Ren, and C-C Jay Kuo. Image splicing localization using a multi-task fully convolutional network (mfcn). *Journal of Visual Communication and Image Representation*, 51:201–209, 2018. 2

[42] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, pages 618–626, 2017. 8

[43] Weijing Shi and Raj Rajkumar. Point-gnn: Graph neural network for 3d object detection in a point cloud. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1711–1719, 2020. 3

[44] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. 1

[45] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017. 4

[46] Junke Wang, Zuxuan Wu, Jingjing Chen, Xintong Han, Abhinav Shrivastava, Ser-Nam Lim, and Yu-Gang Jiang. Objectformer for image manipulation detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2364–2373, 2022. 2, 6, 7, 8

[47] Menglu Wang, Xueyang Fu, Jiawei Liu, and Zheng-Jun Zha. Jpeg compression-aware image forgery localization. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 5871–5879, 2022. 2

[48] Bihan Wen, Ye Zhu, Ramanathan Subramanian, Tian-Tsong Ng, Xuanjing Shen, and Stefan Winkler. Coverage—a novel database for copy-move forgery detection. In *2016 IEEE international conference on image processing (ICIP)*, pages 161–165. IEEE, 2016. 2

[49] Bihan Wen, Ye Zhu, Ramanathan Subramanian, Tian-Tsong Ng, Xuanjing Shen, and Stefan Winkler. Coverage—a novel database for copy-move forgery detection. In *2016 IEEE international conference on image processing (ICIP)*, pages 161–165. IEEE, 2016. 6

[50] Haiwei Wu and Jiantao Zhou. Iid-net: Image inpainting detection network via neural architecture search and attention. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(3):1172–1185, 2021. 2

[51] Yue Wu, Wael Abd-Almageed, and Prem Natarajan. Deep matching and validation network: An end-to-end solution to constrained image splicing localization and detection. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1480–1502, 2017. 2

[52] Yue Wu, Wael Abd-Almageed, and Prem Natarajan. Busternet: Detecting copy-move image forgery with source/target

localization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 168–184, 2018. 2

[53] Yue Wu, Wael Abd-Almageed, and Prem Natarajan. Busternet: Detecting copy-move image forgery with source/target localization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 168–184, 2018. 6

[54] Yue Wu, Wael Abd-Almageed, and Prem Natarajan. Image copy-move forgery detection via an end-to-end deep neural network. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1907–1915. IEEE, 2018. 2

[55] Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan. Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9543–9552, 2019. 1, 2, 3, 5, 6

[56] Quanxin Yang, Dongjin Yu, Zhuxi Zhang, Ye Yao, and Linqiang Chen. Spatiotemporal trident networks: detection and localization of object removal tampering in video passive forensics. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(10):4131–4144, 2020. 2

[57] Zizheng Yang, Mingde Yao, Jie Huang, Man Zhou, and Feng Zhao. Sir-former: Stereo image restoration using transformer. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6377–6385, 2022. 1

[58] Hu Yu, Jie Huang, Yajing Liu, Qi Zhu, Man Zhou, and Feng Zhao. Source-free domain adaptation for real-world image dehazing. In *Proceedings of the 30th ACM International Conference on Multimedia*, page 6645–6654, 2022. 1

[59] Xikun Zhang, Chang Xu, and Dacheng Tao. Context aware graph convolution for skeleton-based action recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14333–14342, 2020. 3

[60] Yulan Zhang, Guopu Zhu, Ligang Wu, Sam Kwong, Hongli Zhang, and Yicong Zhou. Multi-task se-network for image splicing localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021. 2

[61] Peng Zhou, Bor-Chun Chen, Xintong Han, Mahyar Najibi, Abhinav Shrivastava, Ser-Nam Lim, and Larry Davis. Generate, segment, and refine: Towards generic manipulation segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13058–13065, 2020. 2

[62] Peng Zhou, Xintong Han, Vlad I Morariu, and Larry S Davis. Learning rich features for image manipulation detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1053–1061, 2018. 2, 3, 5, 6

[63] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. 1

[64] Xinshan Zhu, Yongjun Qian, Xianfeng Zhao, Biao Sun, and Ya Sun. A deep learning approach to patch-based image inpainting forensics. *Signal Processing: Image Communication*, 67:90–99, 2018. 2

[65] Yurui Zhu, Xueyang Fu, and Aiping Liu. Learning dual transformation networks for image contrast enhancement. *IEEE Signal Processing Letters*, 27:1999–2003, 2020. 1

[66] Yurui Zhu, Xi Wang, Xueyang Fu, and Xiaowei Hu. Enhanced coarse-to-fine network for image restoration from under-display cameras. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part V*, pages 130–146. Springer, 2023. 1