CatmullRom Splines-Based Regression for Image Forgery Localization

Li Zhang[1,2], **Mingliang Xu**[2], **Dong Li**[2], **Jianming Du**[1,†], **Rujing Wang**[1,2†] *

[1] Hefei Institute of Physical Science, Chinese Academy of Sciences, China
[2] University of Science and Technology of China, China

Abstract

IFL (*Image Forgery Location*) helps secure digital media forensics. However, many methods suffer from false detections (i.e., FPs) and inaccurate boundaries. In this paper, we proposed the **C**atmullRom **S**plines-based **R**egression Network (**CSR-Net**), which first rethinks the IFL task from the perspective of regression to deal with this problem. Specifically speaking, we propose an adaptive CutmullRom splines fitting scheme for coarse localization of the tampered regions. Then, for false positive cases, we first develop a novel rescoring mechanism, which aims to filter out samples that cannot have responses on both the classification branch and the instance branch. Later on, to further restrict the boundaries, we design a learnable texture extraction module, which refines and enhances the contour representation by decoupling the horizontal and vertical forgery features to extract a more robust contour representation, thus suppressing FPs. Compared to segmentation-based methods, our method is simple but effective due to the unnecessity of post-processing. Extensive experiments show the superiority of CSR-Net to existing state-of-the-art methods, not only on standard natural image datasets but also on social media datasets.

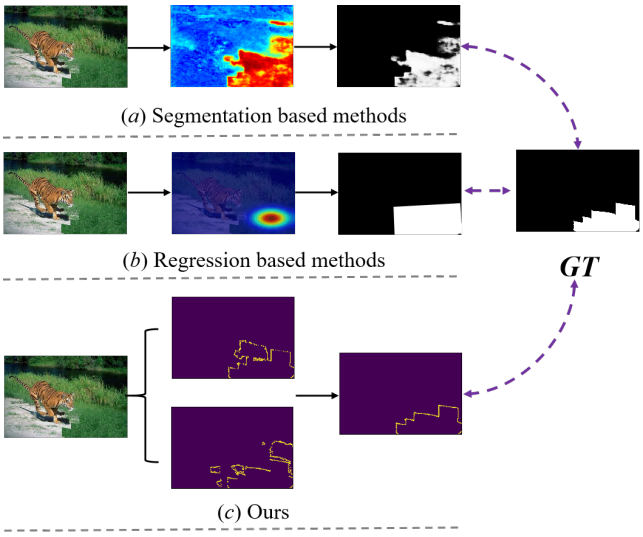

Figure 1: The categorization of existing methods applied into IFL. Please zoom in for better visualization.

Introduction

Image **F**orgery **L**ocation (**IFL**), also known as image tampering detection, refers to the task of detecting the location of forged regions in a suspicious image. With the increasing availability of digital image editing software, image forgery has become a prevalent issue in various fields such as journalism, forensics, and biometrics.

However, IFL has not been adequately studied due to the various techniques used to manipulate images, including removal, splicing, and cloning. Furthermore, the forgers may use carefully crafted tools to conceal the tampered regions and make the detection more difficult. Therefore, the development of accurate and efficient forgery localization algorithms is crucial to address this issue, ensuring the integrity and credibility of digital images in today's world. Thanks to the rapid development of deep learning in recent years, many excellent methods have been introduced and continue to drive the progress of the field (Wang et al. 2022; Hu et al. 2020). However, due to the specific properties of forgery regions such as large variance in color, texture, shape, etc., there are still mainly two challenging issues that haven't been addressed satisfactorily in Image Forgery Localization.

The first issue is **false positives (FPs)**. False positives refer to test results that indicate the presence of a satisfactory target region when in reality it is not convincing. Traditional segmentation-based methods often suffer from this situation (as shown in Fig. 2). Binarization, an indispensable and decisive strategy used in these methods, is a threshold-sensitive task that directly determines the number of foreground region blocks. An unreasonable threshold value often leads to the appearance of unexpected regions (i.e., false positive cases) in traditional segmentation methods. However, many methods, when focusing on potentially tampered regions, usually ignore the false alarm rate. This has negative implications on the propagation of digital content, impacting the profitability of relevant journalistic sources, which constrains the development of assay results in a more convincing direction.

The second issue is **inaccurate boundaries**. As shown in

*† corresponding author.

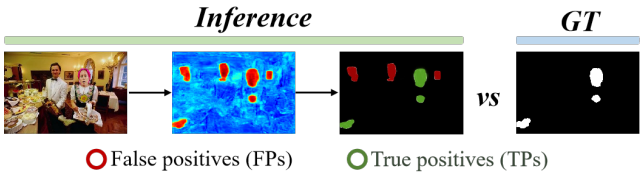

Figure 2: The illustration of FPs in traditional segmentation-based methods.

the results displayed in Fig. 1 (a), traditional segmentation-based methods suffer from inconsistent mask predictions between consecutive decoder layers, which leads to inconsistent optimization goals and weak coupling of feature spaces. On the other hand, the localization effect is also unsatisfactory when the general regression method is directly introduced to handle the task because the bounding boxes used can only localize the target region in a quadrilateral fashion, and often the target region appears mostly in irregular curves (Fig. 1 (b) shows the localization effect of rotation detection (Li et al. 2022), where we use the minimum bounding quadrilateral of the masked region as the Ground Truth). The increasingly elaborate tampered image poses a greater challenge, as most methods do not constrain or explicitly model forged region boundaries well, which can easily lead to the blending of other targets or incompatible backgrounds in the detection results.

Recently, some regression-based strategies have made significant advances in false-positive determination in the field of object detection (You et al. 2022; Li and Košecká 2022; Chen et al. 2023). Differing from object detection tasks, IFL is a pixel-level task, which means that the directness of approach migration will bring performance degradation. To this end, some customized methods and processing need to be introduced which could bridge these two tasks effectively. Specifically, firstly, for the mask labeled GT, we introduce CatmullRom splines to transform it into polygonal frames, thus enabling regression strategies to be applied to pixel-level tasks such as IFL. Meanwhile, during the training and inference process, in order to make the polygonal labeling closer to the real label, adaptive parametric CatmullRom splines method is proposed, which can minimize the similarity gap between the predicted region and the Ground Truth and for the curvature of the target region. secondly, to go further and explicitly suppress false positives in the localization results, we propose an effective re-scoring mechanism: we directly reject false positives that do not receive a response in both branches through two independent prediction branches, each with a regional classification score and an instance score. In addition, to get more accurate boundaries, we further refine the contours of the predicted regions by decoupling horizontal texture features and vertical texture features for modeling the forged region boundaries and reducing the overlap between them and other masks.

Our contributions can be summarized as three folds:

- We tailor a **C**atmullRom **S**plines-based **R**egression Network (**CSR-Net**) to make the first attempt to introduce regression methods into the pixel-level task (referring IFL in this paper).
- To explicitly suppress the false positive samples and to avoid unclear edges, we design two mutually complementary and reinforcing components, i.e., Comprehensive Re-scoring Algorithm (CRA) to synthetically evaluate the confidence score of each region as a tampered region, while Vertical Texture-interactive Perception (VTP) is developed to control the generation of more accurate region edges.
- Extensive experiments on multiple public datasets (including natural image datasets and social media datasets) demonstrate the superiority of our method compared to state-of-the-art methods in IFL.

Related Work

Classic Methods in IFL

The prior art in IFL mainly relies on feature extraction and matching techniques, e.g. color filter array (Ferrara et al. 2012), photo-response non-uniformity noise (Chierchia et al. 2014), illumination (Carvalho et al. 2015), JPEG artifacts (Iakovidou et al. 2018) and so on. Despite the achievements, these methods often struggle with complex forgery techniques or when the forged region is well-blended into the background of the image. In recent years, deep learning-based methods have shown great potential in IFL. Many methods have been proposed to promote the progress and development of this field. For instance, In (Liu et al. 2022), Liu et al. proposed PSCC-Net, which uses a two-path (top-down and bottom-up route) methodology to analyze the image. A self-adversarial training strategy and a reliable coarse-to-fine network (Zhuo et al. 2022) is designed which utilizes a self-attention mechanism to localize forged regions in forgery images. However, these methods are all conducted from the point of segmentation, an unlearnable hyperparameter needs to be predefined for the binarization of different regions, limiting the method's further development.

Regression Based Methods

Regression-based methods have been widely used (Savran, Sankur, and Bilge 2012; Xia et al. 2021) in computer vision, particularly in tasks such as object detection (Carion et al. 2020) and localization (Choe et al. 2020). Over the years, various algorithms were proposed such as Fast R-CNN (Girshick 2015), YOLO (Redmon and Farhadi 2018) and Diffusion-Det (Chen et al. 2022). Some improved algorithms can adapt quickly to more complex scenes, such as solving false-positive samples (You et al. 2022) caused by uneven sample segmentation in 3D scenes through detection frames. In certain scenarios where quadrilateral regions cannot be detected, some parametric curves have been introduced to solve the regression problem. This mechanism is achieved through an interpolation spline or an approximation spline function. For example, gesture recognition uses a customized way to fit points to *Bezier curves* with constant memory usage, while *B-splines* are used to detect lane markings and regress their 3D location (Pittner, Condurache, and Janai 2023). In this paper, we show how to

tailor a customized CatmullRom detection for IFL and find that reasonable parameter values can significantly improve the model fit. Our results demonstrate that a proper balance of the tension factor (τ) can help improve the characterization performance of the CatmullRom splines, emphasizing the importance of flexible parameter adjustment in practical applications.

Method

Overview

Fig. 3 is an overview of our framework. The input image is represented as $X \in \mathbb{R}^{H\times W\times 3}$. First, we use FPN embedded with ResNet-50 as the backbone network to conduct Catmull spline detection. More specifically, our approach utilizes a parameterization method based on CatmullRom splines to fit the target segmentation area adaptively (orange part). Following (Chen et al. 2017), we adopt atrous spatial pyramid pooling (ASPP) together with ResNet-50 to capture the long-range contextual information and multi-scale features. This anchor-free convolutional neural network significantly simplifies the detection for our task and also allows us to obtain a coarse feature map. Later on, we use a re-scoring mechanism (CRA) to filter out the false positive samples for suspicious regions highlighted on the coarse feature maps (blue part). Finally, we perform texture extraction (by VTP) on the regions from both horizontal and vertical directions simultaneously in anticipation of obtaining more accurate boundaries (green part). Note that each tampered region reserved will be processed by VTP independently.

CatmullRom Splines Detection

Most traditional methods use the idea of segmentation-based in IFL (Liu et al. 2022; Wang et al. 2023; Wu, AbdAlmageed, and Natarajan 2019; Li et al. 2023). However, regression-based methods tend to be a more efficient approach when dealing with such mask or polygon-based datasets, e.g. (Liu et al. 2019; Pranav, Zhenggang et al. 2020; Zhang et al. 2022). In general, mainstream regression-based methods require complex processing to fit the instance boundaries, which leads to unreliability and instability in practice. In recent years, spline curves have been used in computer graphics applications to generate curves of various shapes. For example, automatic driving lane lines (Ma et al. 2019; Yu and Chen 2017), text detection (Liu et al. 2020; Tang et al. 2022; Nguyen et al. 2021), Fault detection (Park et al. 2011; Guo and Wang 2005), etc. Among them, CatmullRom spline function is a classic interpolating spline, which is suitable for parameterization of tampered regions due to its fitting effect and inference cost (Chandra 2020; Li 2022).

Specifically, CatmullRom splines are a family of cubic interpolating splines formulated such that the tangent at each point $\mathbf{P}_i$ is calculated using the previous and next point on the spline. Under the given control points, there are variations of CatmullRom spline functions that can be adapted to any shape desired (Li and Chen 2016; Li, Liu, and Liu 2022). In addition, only integer coefficients are involved in constructing a cubic CatmullRom spline function which reduces the implementation cost when compared to other spline functions. All these properties mentioned above contribute to the faster inference speed and lower calculation consumption (Flops).

Mathematically, CatmullRom spline is defined as Eq. 1:

$$\boldsymbol{c}_i(t) = \sum_{j=0}^{3} b_j(t)\boldsymbol{p}_{i+j}, \quad i = 0, 1, \ldots, n-3 \tag{1}$$

where $0 \le t \le 1, \quad \boldsymbol{p}_i (i = 0, 1, \ldots, n-3; n \ge 3)$ are control points, $b_j(t)$ is the basis. For example, it can be expressed by Eq. 2 when the highest power of t in the function $b_j(t)$ is 3:

$$\begin{aligned}\boldsymbol{c}_i(t) = \tfrac{1}{2}\cdot\begin{bmatrix}1 & t & t^2 & t^3\end{bmatrix}\cdot \\ \begin{bmatrix} 0 & 2 & 0 & 0 \\ -\tau & 0 & \tau & 0 \\ 2\tau & \tau-6 & -2(\tau-3) & -\tau \\ -\tau & 4-\tau & \tau-4 & \tau \end{bmatrix}\cdot\begin{bmatrix}\boldsymbol{p}_i \\ \boldsymbol{p}_{i+1} \\ \boldsymbol{p}_{i+2} \\ \boldsymbol{p}_{i+3}\end{bmatrix}\end{aligned} \tag{2}$$

In order to reconcile arbitrary shapes of the tampered regions with CatmullRom splines, we thoroughly studied oriented or curved tampered from existing datasets and the authentic images. In CatmullRom splines, τ (tension factor) is an important parameter that is used to control the tightness of the splines. A higher value of the tension factor will cause the curve to bend more tightly between the control points, thus fitting closer to the given data points during the fitting process. Conversely, lower values of the tension factor will cause the curve to be smoother between the control points. Intuitively, the conventional CatmullRom spline (parameter τ=1) is a poor fit for the IFL task directly, so we sought to find the right balance between fitting accuracy and curve smoothness by adjusting τ. Ablation experiments (In the ablation analysis part) show that CatmullRom splines can be reliable for this task when τ is set to 16. It also allows the learned control points to be closer to the foreground (tampered) area.

CatmullRom Ground Truth Generation

In IFL, many benchmarks use Mask or polygon-based datasets as public datasets (Dong, Wang, and Tan 2013; Hsu and Chang 2006; Alibaba 2021/2022). Given the annotated points $\{p_i\}_{i=1}^{n}$ from the curved boundary where p_i represents the $i-th$ annotating point, the main goal is to obtain the optimal parameters for CatmullRom splines $c(t)$ in Eq. 1. To achieve this, we can simply apply the standard least square method, as shown in Eq. 3:

$$\begin{bmatrix} \boldsymbol{p}_{03t_0} & \cdots & \boldsymbol{p}_{33t_0} \\ \boldsymbol{p}_{03t_1} & \cdots & \boldsymbol{p}_{33t_1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{p}_{03t_m} & \cdots & \boldsymbol{p}_{33t_m} \end{bmatrix}\begin{bmatrix} c_{x_0} & c_{y_0} \\ c_{x_1} & c_{y_1} \\ c_{x_2} & c_{y_2} \\ c_{x_3} & c_{y_3} \end{bmatrix} = \begin{bmatrix} \mathscr{P}_{x_0} & \mathscr{P}_{y_0} \\ \mathscr{P}_{x_1} & \mathscr{P}_{y_1} \\ \vdots & \vdots \\ \mathscr{P}_{x_m} & \mathscr{P}_{y_m} \end{bmatrix} \tag{3}$$

where m represents the number of annotated points for a curved boundary. while t is calculated by using the ratio of the cumulative length to the perimeter of the polyline. $\boldsymbol{p}_{ij}$ can be refered from Eq. 1, and we use $\mathscr{P}_i$ represents the new coordinate points after the transformation. According to Eq. 1 and Eq. 3, we convert the original masked annotation to a parameterized CatmullRom spline. Illustration can be referred from Fig. 4.

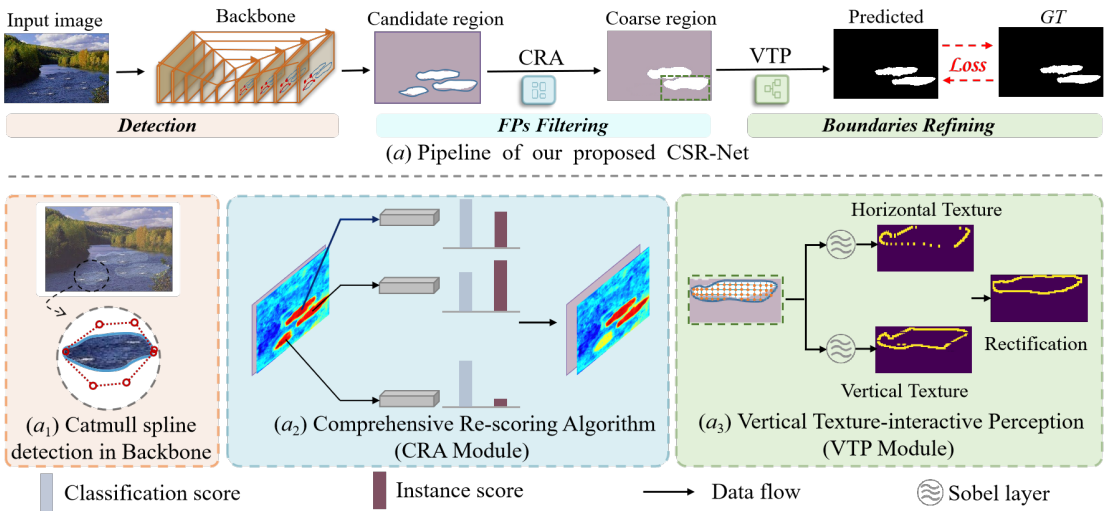

Figure 3: Overall of our proposed CSR-Net. The top part is our pipeline, which takes a suspicious image (H×W×3) as input, and the output is the predicted mask (H×W×1, the tampered regions). Formally, an uncertain number of potential regions will be obtained after CRA processing, and VTP will refine each region independently. The bottom parts are details of each module.

Comprehensive Re-scoring Algorithm

The fundamental mechanism behind Mask R-CNN is to treat the classification confidence of the resulting bounding boxes as scores, and then a predetermined threshold is used to filter out the background boxes. However, despite the advances, when bounding boxes contain an obviously incompatible region instance, it is accompanied by a large amount of background information, and Mask R-CNN often filters out such low-score true positives, while it retains some FPs with relatively high confidence in contrast. Therefore, we re-assign scores for each region instance. Specifically speaking, the comprehensive score of region instance is composed of two parts: classification score (CLS) and instance score (INS). Mathematically, the comprehensive score for the $i-$th proposal, given the predicted n-class scores CLS = $\left\{s_{ij}^{cls} \mid j \in [0, \cdots, n-1]\right\}$ and INS = $\left\{s_{ij}^{ins} \mid j \in [0, \cdots, n-1]\right\}$ is computed via the customized softmax function: Eq. 4 .

$$s_{ij} = \frac{e^{s_{ij}^{cls} + s_{ij}^{ins}}}{\sum_{l=0}^{n-1} e^{s_{il}^{cls} + s_{il}^{ins}}} \tag{4}$$

In our work, we adopt $n = 2$, where the two classes represent tampered (foreground) and authentic (background) regions. Therefore, we only need to calculate the score for the foreground class. CLS is directly obtained by a classification branch similar to Mask R-CNN, and INS is the activation value of the region instance on the global region segmentation map. In detail, it is projected onto a tampered region segmentation map for each region instance, containing $P_i = \left\{p_i^1, p_i^2 \dots p_i^n\right\}$, and the mean of P_i in the region instance area can be formulated as:

$$s_{i1}^{ins} = \frac{\sum_j p_i^j}{N} \tag{5}$$

where P_i is the set of the pixels' value of $i-$th region instance on region segmentation map. The classification score is organically integrated with the instance score to get the comprehensive score, which can reduce the FP confidence in practice. This is because FPs tend to have weaker responses than regions on the segmentation map.

Experiments results in the following show that our design is more friendly for splicing cases because the splicing cases usually enjoy a stronger response on the segmentation map, a high instance score will compensate for a low classification score.

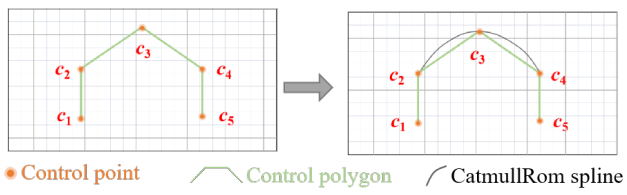

Figure 4: An example of Cubic CatmullRom splines. Note that with only two end-points c_1 and c_5 the CatmullRom spline degenerates to a straight line.

Vertical Texture-interactive Perception

Traditional edge detection operators (e.g. Sobel, Roberts, Prewitt, etc.) help to extract handcrafted features on natural image processing tasks, while the biggest drawback is that they cannot learn dynamically according to the specificity of the task. Inspired by (Holla and Lee 2022), we adopt an edge detection operator into a learnable module coined Sobel layer, see Fig. 5. Furthermore, for better modeling of tampered area boundaries, we introduce Vertical Texture-interactive Perception (VTP) into our network. In VTP, tampered region is represented with a set of contour points,

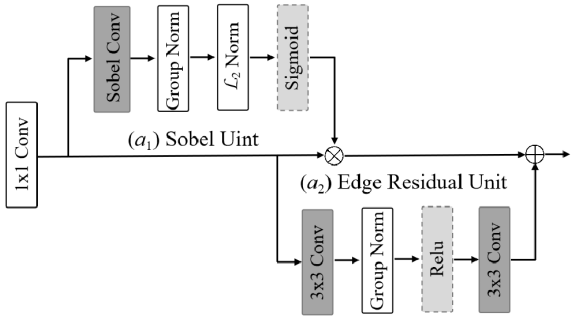

Figure 5: Diagrams of Sobel layer, used in VTP for enhancing edge-related patterns and manipulation edge detection. Features from the i-th block first go through the Sobel Unit (SU) followed by an Edge Residual Unit (ERU). For training and optimization reasons, a residual learning strategy is introduced.

these points containing strong texture characteristics can accurately localize tampered regions with arbitrary shapes.

See them all: there are two core parallel branches in VTP, in the top branch, we introduce a convolutional kernel with size $1 \times k$ sliding over the feature maps to model the local texture information in the horizontal direction, which only focuses on the texture characteristics in a k-range region. This flexible trick is proved to be simple but works a lot through our pre-experiments. Moreover, It is nearly cost-free while maintaining competitive efficiency at the same time. Through a similar paradigm, the bottom branch is conducted to model the texture characteristics in the vertical direction through a convolutional kernel with size $k \times 1$. k is a hyper-parameter to control the size of the receptive field of texture characteristics. In the actual experiment, we take $k = 3$. Finally, two independent sigmoid layers are involved to normalize the heatmaps to [0, 1] in both directions. In this way, tampered regions can be detected in two orthogonal directions and represented with contour points in two different heatmaps, either of which only responds to texture characteristics in a certain direction.

As false positive predictions can be effectively suppressed by considering the response value in both orthogonal directions, two heatmaps from VTP are further processed through Point Re-scoring Algorithm. Concretely, points in different heatmaps are first processed through NMS to achieve a tight representation. Then, to suppress the predictions with strong unidirectional or weakly orthogonal responses, we only select the points with distinct responses in both heatmaps as candidates. Finally, the tampered region can be represented with a polygon made up of these high-quality contour points.

Optimization

As described above, our network includes multi-task. Therefore, we calculate the loss function for the following components:

$$L = L_{rpn} + \lambda_1 \cdot L_{cls} + \lambda_2 \cdot L_{mask} + \lambda_3 \cdot L_{gts} + \lambda_4 \cdot L_{CR} \tag{6}$$

where L_{rpn} , L_{cls} and L_{mask} are the standard loss derived from Mask R-CNN. The L_{gts} is used to optimize tampered region detection, defined as:

$$L_{gts} = \frac{1}{N} \sum_i - \log \left(\frac{e^{p_i}}{\sum_j e^{p_j}} \right) \tag{7}$$

The L_{gts} is Softmax loss, where p is the output prediction of the network.

The L_{CR} is used to optimize the fit of CatmullRom spline detection, defined as:

$$L_{CR} = L_{ctr} + L_{bias} \tag{8}$$

The L_{ctr} and L_{bias} are all FCOS loss (Tian et al. 2019). The former is used to optimize distance loss from the center of CatmullRom control points, while the offset distance of these control points from the center is constrained by the latter.

Experiment

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, and 3) removal.

Testing Datasets Following (Wang et al. 2022; Hu et al. 2020), we evaluate our model on CASIA (Dong, Wang, and Tan 2013), Columbia (Hsu and Chang 2006), NIST16 (Guan et al. 2019), COVER (Wen et al. 2016).

Evaluation Metrics To quantify the localization performance, following previous works (Hu et al. 2020), we use pixel-level Area Under Curve (AUC) and F1 score on manipulation masks. Since binary masks are required to compute F1 scores, we adopt the Equal Error Rate (EER) threshold to binarize them.

Implementation Details The input images are resized to 512 × 512. In this work, the backbone network is ResNet-50, pre-trained on ImageNet. Implemented by PyTorch, our model is trained with GeForce GTX 3090, using Adam as the optimizer.

Comparison with the SOTA Methods

Following classic methods (Hu et al. 2020; Wang et al. 2022), our model is compared with other state-of-the-art tampering localization methods under two settings: 1) training on the synthetic dataset and evaluating 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 Tab. 1 shows the localization performance of pre-trained models for different SOTA methods on five datasets under pixel-level AUC. Our CSR-Net achieves the best localization performance on Coverage, CASIA, NIST16 and IMD20, ranking second on Columbia. Especially, It achieves 94.4% on the copy-move dataset

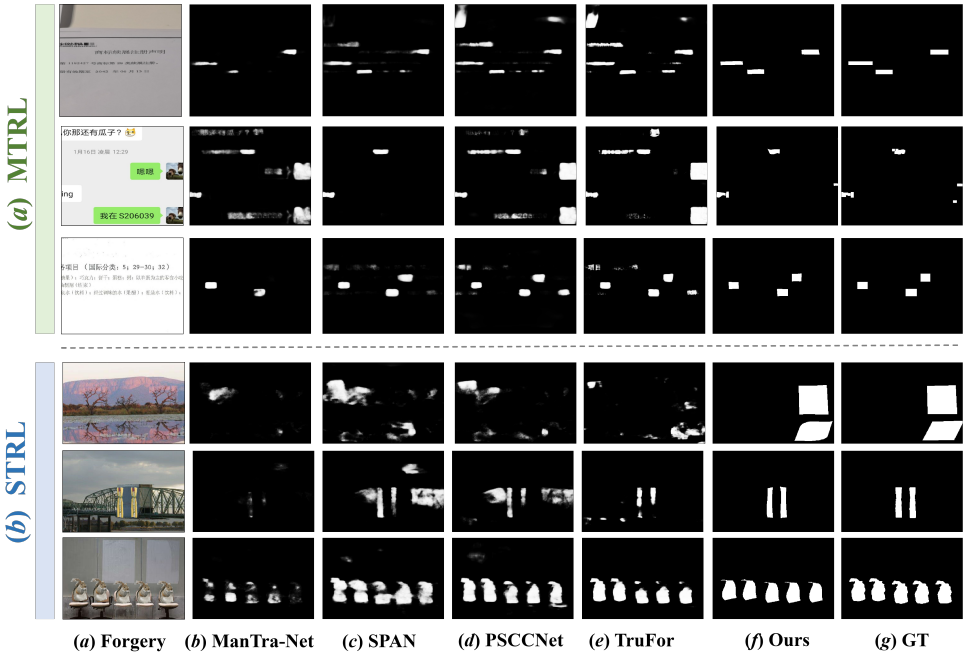

Figure 6: Visualization of the predicted manipulation mask by different methods. From left to right, we show forged images, predictions of ManTra-Net, SPAN, PSCCNet, TruFor, Ours and GT masks.

Method	Data	Columbia	Coverage	CASIA	NIST16	IMD20
SPAN	96k	93.6	92.2	79.7	84.0	75.0
TruFor	100k	97.7	85.4	83.3	83.9	81.8
PSCCNet	100k	**98.2**	84.7	82.9	85.5	80.6
ObjectFormer	62K	95.5	92.8	84.3	87.2	82.1
ManTraNet	64K	82.4	81.9	81.7	79.5	74.8
Ours	60K	96.8	**94.3**	**88.1**	**88.3**	**85.4**

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	-
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
RGB-N	81.7	43.7	79.5	40.8	93.7	72.2
Ours	**97.9**	**78.0**	**90.4**	**58.5**	**99.7**	**83.5**

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

(COVER), whose image forgery regions are indistinguishable from the background. This validates our model owns the superior ability to suppress the FPs and generate more accurate edges. Yet, we fail to achieve the best performance on Columbia, with a gap of 1.4 % AUC score lower than that of PSCCNet. We conjecture that the explanation may be the distribution of their synthesized training data closely resembles that of the Columbia dataset. This is further supported by the results in Tab. 2, which shows that CSR-Net performs better than PSCCNet in terms of both AUC and F1 scores. Furthermore, it is worth pointing out that we achieve decent results with less pre-training data.

Fine-tuned Model The network weights of the pre-trained model are used to initiate the fine-tuned models that will be trained on the training split of Coverage, CASIA, and NIST16 datasets, respectively. We evaluate the fine-tuned models of different methods in Tab. 2. As for AUC and F1, our model achieves significant performance gains. This validates that the CRA module effectively suppresses false positive cases and improves the accuracy of predicted region locations and boundaries by VTP.

After synthesizing the data in Tab. 1 and 2, our approach proves that introducing regression methods for pixel-level tasks is effective as expected which is mentioned in the Introduction.

Ablation Analysis

In this section, we conduct experiments to demonstrate the effectiveness of our proposed method CSR-Net. Formally, the CatmullRom Splines-based Regression (CSR) is introduced to better describe the tampered region compared to traditional regression methods. Comprehensive Rescoring Algorithm (CRA) aims to choose expected regions with high classification scores as well as superior instance scores, while Vertical Texture-interactive Perception (VTP) is used to model texture features both horizontally and vertically to refine the target region. To further evaluate the effectiveness of CSR, CRA, and VTP, we remove them separately and verify the forgery localization performance on CASIA and NIST16 datasets. Tab. 3 shows the quantitative results.

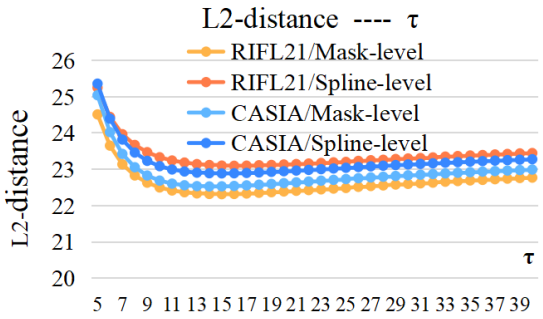

Figure 7: L_2 distance with different value of τ. We show the results through the format of "dataset/label", for example, RIFL21/Mask-level means the average distance of control points in RIFL21 from the Ground Truth with Mask-level.

The baseline (I) denotes that we only use the traditional regression method (Li et al. 2022). In the following ablation experiments, we can infer that the F1 scores decrease by 1.9 % on CASIA and 1.7 % on NIST16 when VTP is not involved. While without CRA, the AUC scores decrease more compared to (IV). Yet, when CRA is not available, significant performance degradation in (II), i.e., 12.3% in terms of AUC and 11.2% in terms of F1 on CASIA can be observed.

Index	Variants	CASIA		NIST16	
		AUC	F1	AUC	F1
I	Baseline	68.5	35.3	75.9	51.2
II	w/o CSR	78.1	47.6	86.1	62.1
III	w/o CRA	86.3	55.5	95.9	78.8
IV	w/o VTP	88.9	56.9	97.8	81.8
V	Ours	**90.4**	**58.8**	**99.7**	**83.5**

Table 3: Ablation results on CASIA and NIST16 dataset using different variants of CSR-Net. AUC and F1 scores (%) are reported.

In Fig. 7, we represent the different values of parameters τ in CatmullRom Ground Truth Generation to validate the respective prediction effects over natural image dataset (i.e., CASIA) and social media dataset (i.e., RIFL21). Intuitively, as the τ gradually increases, the Euclidean distance between the fitted CatmullRom control points and the Ground Truth with mask-level in different datasets gradually decreases, showing a better fit. However, when τ exceeds 16, the Euclidean distance instead shows a tendency to expand, implying that the fitting effect may appear to decrease. Clearly, $\tau = 16$ is an excellent choice to generate the optimal CatmullRom-based Ground Truth.

Visualization Results

Qualitative results. We provide predicted forgery masks of different methods in Fig. 6. Since the source code of ObjectFormer (Wang et al. 2022) is not available, their predictions are not available. Compared with the state-of-the-art methods, our CSR-Net achieves better performance, both in terms of suppressing false positives and in more accurate tampered region boundaries. We have reason to believe that the improvement benefits from the CRA and VTP. CRA is able to consider each possible area more comprehensively and determine the subtle differences between tampered and authentic regions, while VTP models texture boundaries from two orthogonal approaches simultaneously to accurately describe the target regions.

Different splines-based regression. There are many types of interpolation functions, classical ones such as Catmull-Rom splines and Bezier curves, the former is an interpolation spline function, which precisely interpolates a set of known data points by using a series of nodes, while the latter is an approximation spline function, which approximates a set of data points by using nodes. Datasets for IFL are produced from natural images and social media, and the tampered regions share different shapes. Through comparison experiments, we found that CutmullRom splines are more suitable for datasets with diverse curvature (e.g., IFL), while Bezier curve-based methods are sometimes susceptible to interference from other targets. For more details, please follow Fig. 8.

Figure 8: Visualization of the results by different splines-based regression. From left to right, we show forged images, results of different splines-based regression (The left side shows the confidence score, while the right side is the predicted manipulation mask), GT masks. Due to the space limitation, please zoom in for better visualization.

Conclusion

In this paper, we elaborately design a customized Catmull-Rom Splines-based Regression Network (CSR-Net) for IFL, which first attempts to introduce regression methods into the pixel-level (IFL in this paper). In detail, in contrast to traditional detection methods that rely on bounding boxes, we first introduce the CatmullRom fitting technique, which adapts contour modeling for control points in the target region, thereby achieving more accurate and efficient localization of tampered regions. Then, to suppress the FPs, Comprehensive Re-scoring Algorithm (CRA) is designed to filter the exact tampered region with classification score and instance score. Moreover, we proposed a learnable region texture extraction module named Vertical Texture-interactive Perception (VTP) to further refine the edges. Thus the CSR-Net can perceive all tampered regions without nearly FPs and achieve accurate localization. Extensive experiments show the superiority of CSR-Net to existing state-of-the-art approaches, not only on natural image datasets but also on social media datasets.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grant (No.32171888), the Dean's Fund of Hefei Institute of Physical Science, Chinese Academy of Sciences (YZJJ2022QN32), the Natural Science Foundation of Anhui Province (No.2208085MC57), and the National Key Research and Development Program of China (2019YFE0125700).

References

Alibaba. 2021/2022. Real-World Image Forgery Localization dataset. https://tianchi.aliyun.com/competition/entrance/531945/introduction?spm=5176.12281957.0.0.1aaf2448THhlg4,.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-end object detection with transformers. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, 213–229. Springer.

Carvalho, T.; Faria, F. A.; Pedrini, H.; Torres, R. d. S.; and Rocha, A. 2015. Illuminant-based transformed spaces for image forensics. *IEEE transactions on information forensics and security*, 11(4): 720–733.

Chandra, M. 2020. Hardware implementation of hyperbolic tangent function using catmull-rom spline interpolation. *arXiv preprint arXiv:2007.13516*.

Chen, L.-C.; Papandreou, G.; Schroff, F.; and Adam, H. 2017. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*.

Chen, S.; Sun, P.; Song, Y.; and Luo, P. 2022. Diffusiondet: Diffusion model for object detection. *arXiv preprint arXiv:2211.09788*.

Chen, Y.; Yu, Z.; Chen, Y.; Lan, S.; Anandkumar, A.; Jia, J.; and Alvarez, J. 2023. FocalFormer3D: Focusing on Hard Instance for 3D Object Detection. *arXiv preprint arXiv:2308.04556*.

Chierchia, G.; Poggi, G.; Sansone, C.; and Verdoliva, L. 2014. A Bayesian-MRF approach for PRNU-based image forgery detection. *IEEE Transactions on Information Forensics and Security*, 9(4): 554–567.

Choe, J.; Oh, S. J.; Lee, S.; Chun, S.; Akata, Z.; and Shim, H. 2020. Evaluating weakly supervised object localization methods right. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3133–3142.

Dong, J.; Wang, W.; and Tan, T. 2013. Casia image tampering detection evaluation database. In *2013 IEEE China Summit and International Conference on Signal and Information Processing*, 422–426. IEEE.

Ferrara, P.; Bianchi, T.; De Rosa, A.; and Piva, A. 2012. Image forgery localization via fine-grained analysis of CFA artifacts. *IEEE Transactions on Information Forensics and Security*, 7(5): 1566–1577.

Girshick, R. 2015. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448.

Guan, H.; Kozak, M.; Robertson, E.; Lee, Y.; Yates, A. N.; Delgado, A.; Zhou, D.; Kheyrkhah, T.; Smith, J.; and Fiscus, J. 2019. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 63–72. IEEE.

Guo, L.; and Wang, H. 2005. Fault detection and diagnosis for general stochastic systems using B-spline expansions and nonlinear filters. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 52(8): 1644–1652.

Holla, K. S.; and Lee, B. 2022. Convolutional Residual Blocks With Edge Guidance for Image Denoising. In *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, 645–647. IEEE.

Hsu, J.; and Chang, S. 2006. Columbia uncompressed image splicing detection evaluation dataset. *Columbia DVMM Research Lab*.

Hu, X.; Zhang, Z.; Jiang, Z.; Chaudhuri, S.; Yang, Z.; and Nevatia, R. 2020. SPAN: Spatial pyramid attention network for image manipulation localization. In *European conference on computer vision*, 312–328. Springer.

Iakovidou, C.; Zampoglou, M.; Papadopoulos, S.; and Kompatsiaris, Y. 2018. Content-aware detection of JPEG grid inconsistencies for intuitive image forensics. *Journal of Visual Communication and Image Representation*, 54: 155–170.

Li, D.; Zhu, J.; Wang, M.; Liu, J.; Fu, X.; and Zha, Z.-J. 2023. Edge-Aware Regional Message Passing Controller for Image Forgery Localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8222–8232.

Li, H., PhD. 2022. Curve Fitting and Interpolation. In *Numerical Methods Using Kotlin: For Data Science, Analysis, and Engineering*, 169–196. Springer.

Li, J.; and Chen, S. 2016. The cubic α-Catmull-Rom spline. *Mathematical and Computational Applications*, 21(3): 33.

Li, J.; Liu, C.; and Liu, S. 2022. The quartic Catmull–Rom spline with local adjustability and its shape optimization. *Advances in Continuous and Discrete Models*, 2022(1): 1–14.

Li, W.; Chen, Y.; Hu, K.; and Zhu, J. 2022. Oriented reppoints for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1829–1838.

Li, Y.; and Košecká, J. 2022. Uncertainty aware proposal segmentation for unknown object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 241–250.

Liu, X.; Liu, Y.; Chen, J.; and Liu, X. 2022. PSCC-Net: Progressive spatio-channel correlation network for image manipulation detection and localization. *IEEE Transactions on Circuits and Systems for Video Technology*.

Liu, Y.; Chen, H.; Shen, C.; He, T.; Jin, L.; and Wang, L. 2020. Abcnet: Real-time scene text spotting with adaptive bezier-curve network. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9809–9818.

Liu, Y.; Jin, L.; Zhang, S.; Luo, C.; and Zhang, S. 2019. Curved scene text detection via transverse and longitudinal sequence connection. *Pattern Recognition*, 90: 337–345.

Ma, L.; Wu, T.; Li, Y.; Li, J.; Chen, Y.; and Chapman, M. 2019. Automated extraction of driving lines from mobile laser scanning point clouds. In *Proc. Adv. Cartogr. GISci. ICA*, 1–6.

Nguyen, N.; Nguyen, T.; Tran, V.; Tran, M.-T.; Ngo, T. D.; Nguyen, T. H.; and Hoai, M. 2021. Dictionary-guided scene text recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7383–7392.

Park, J.; Kwon, I.-H.; Kim, S.-S.; and Baek, J.-G. 2011. Spline regression based feature extraction for semiconductor process fault detection using support vector machine. *Expert Systems with Applications*, 38(5): 5711–5718.

Pittner, M.; Condurache, A.; and Janai, J. 2023. 3D-SpLineNet: 3D Traffic Line Detection Using Parametric Spline Representations. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 602–611.

Pranav, M.; Zhenggang, L.; et al. 2020. A day on campus-an anomaly detection dataset for events in a single camera. In *Proceedings of the Asian Conference on Computer Vision*.

Redmon, J.; and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Savran, A.; Sankur, B.; and Bilge, M. T. 2012. Regression-based intensity estimation of facial action units. *Image and Vision Computing*, 30(10): 774–784.

Tang, J.; Zhang, W.; Liu, H.; Yang, M.; Jiang, B.; Hu, G.; and Bai, X. 2022. Few could be better than all: Feature sampling and grouping for scene text detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4563–4572.

Tian, Z.; Shen, C.; Chen, H.; and He, T. 2019. Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9627–9636.

Wang, C.; Huang, Z.; Qi, S.; Yu, Y.; Shen, G.; and Zhang, Y. 2023. Shrinking the Semantic Gap: Spatial Pooling of Local Moment Invariants for Copy-Move Forgery Detection. *IEEE Transactions on Information Forensics and Security*.

Wang, J.; Wu, Z.; Chen, J.; Han, X.; Shrivastava, A.; Lim, S.-N.; and Jiang, Y.-G. 2022. Objectformer for image manipulation detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2364–2373.

Wen, B.; Zhu, Y.; Subramanian, R.; Ng, T.-T.; Shen, X.; and Winkler, S. 2016. COVERAGE—A novel database for copy-move forgery detection. In *2016 IEEE international conference on image processing (ICIP)*, 161–165. IEEE.

Wu, Y.; AbdAlmageed, W.; and Natarajan, P. 2019. Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9543–9552.

Xia, W.; Gao, Q.; Wang, Q.; and Gao, X. 2021. Regression-based clustering network via combining prior information. *Neurocomputing*, 448: 324–332.

You, Y.; Ye, Z.; Lou, Y.; Li, C.; Li, Y.-L.; Ma, L.; Wang, W.; and Lu, C. 2022. Canonical voting: Towards robust oriented bounding box detection in 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1193–1202.

Yu, B.; and Chen, Y. 2017. Driving rhythm method for driving comfort analysis on rural highways. *Promet-Traffic&Transportation*, 29(4): 371–379.

Zhang, L.; Du, J.; Dong, S.; Wang, F.; Xie, C.; and Wang, R. 2022. AM-ResNet: Low-energy-consumption addition-multiplication hybrid ResNet for pest recognition. *Computers and Electronics in Agriculture*, 202: 107357.

Zhuo, L.; Tan, S.; Li, B.; and Huang, J. 2022. Self-Adversarial Training incorporating Forgery Attention for Image Forgery Localization. *IEEE Transactions on Information Forensics and Security*, 17: 819–834.