ForgerySpotter: Pinpointing Tampered Regions with Multi-scale Evidence and Confidence-Guided Refinement

Xiaolong Cheng[1] and Juanjuan Luo[1] and MingChen Han[1]

[1] Beijing University of Posts and Telecommunications, Beijing 100876, P.R. China

Abstract. Image manipulation detection (IMD) is crucial for maintaining the integrity of digital media, as forged images can be used to spread false information and erode public trust. IMD faces two persistent challenges: i) limited generalization to diverse real-world post-processing operations, and ii) imprecise localization accuracy yielding coarse or incomplete tampering regions. Moreover, existing methods often lack interpretability due to the absence of reliable confidence estimation. The primary research often prioritizes feature or architectural improvements while neglecting the integration of detection reliability with localization refinement. In this paper, we propose a unified framework that incorporates multi-dimensional feature extraction, multi-scale feature fusion, and a confidence-guided refine mechanism. Our method captures tampering traces across types and scales adaptively, while the confidence-guided mechanism refines localization maps and estimates pixel-wise reliability. Extensive experiments on multiple datasets demonstrate that the proposed approach achieves state-of-the art performance and shows strong generalization, validating its effectiveness and practicality.

Keywords: Image Manipulation Detection. .

1 Introduction

“Seeing is believing.” The longstanding adage, however, is being fundamentally challenged by the rapid development of digital image tampering technologies. Commercial image editing tools, such as Photoshop and GIMP, have popularized common manipulations including splicing, removal, and copy-paste. Meanwhile, image generation techniques represented by Generative Adversarial Networks (GANs) and diffusion models have largely overcome the technical barriers to semantic-level content synthesis. These methods can not only create highly realistic faces or scenes, but also blend tampered regions seamlessly with the original image in terms of details like lighting and texture. Furthermore, multimodal large language models (MLLMs) (e.g., SeedEdit and GPT image-1) enable image editing through natural language instructions instead of manual operations, further blurring the boundary between reality and fiction.

The dramatically reduced technical barriers and unprecedented realism of tampered images have profoundly affected critical domains, including journalism, judicial forensics, and social media. As a primary medium for multimedia communication, digital images are increasingly met with skepticism regarding their authenticity. Malicious use

of image manipulation can lead to severe consequences, such as widespread misinformation, privacy violations, infringement of portrait rights, manipulation of public opinion, and even threats to social stability.

In response, *Image Manipulation Detection* (IMD) has become a vital line of defense within the multimedia forensics field, which is also an active research topic in recent years. Despite significant progress, several critical challenges persist (see Fig. 1(a)).

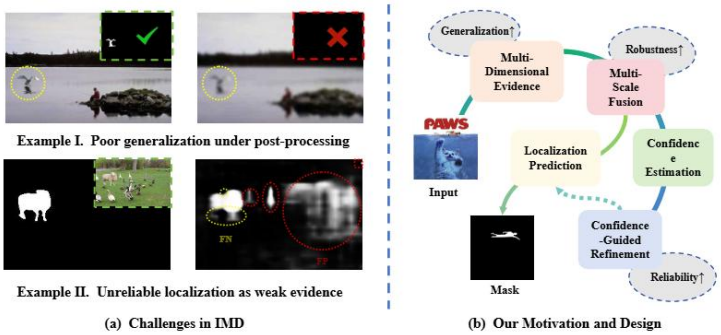

Fig. 1. Key challenges and motivation for reliable IMD

1) Limited generalization and robustness of current models hinder their application in complex real-world scenarios. This limitation stems from the inherent diversity of natural images and the prevalence of unknown post-processing operations (e.g., repeated JPEG compression on social media platforms), which degrade image quality and introduce a domain gap between training data and real-world data distribution. Consequently, models relying on static or single-feature representations often struggle to capture the artifacts under such variations.

2) Insufficient localization precision despite high detection scores. Although current methods perform well at the image level, their pixel-wise localization remains imprecise. Specifically, the predicted masks often contain both incomplete artifacts (FN) and false-positive pixels (FP), leading to high false alarm rates. This limitation undermines the reliability of detection results for downstream applications such as image authentication and content restoration, placing the research in a "trust but verify" dilemma.

To address the above challenges in IMD, we propose a novel framework that incorporates multi-dimensional feature extraction, multi-scale feature fusion, and a confidence-guided mechanism (see Fig.1(b)). The multi-dimensional extractor captures complementary cues from both low-level artifacts and high-level semantic discrepancies, thereby reducing the limitations of single-feature representations. The dynamic fusion module aggregates fine-grained local cues and global contextual information, enhancing generalization, robustness, and localization of small tampered regions. Confidence-guided refinement suppresses false positives and recovers missing artifacts through adaptive strategies, improving localization precision and reliability.

The main contributions of this paper are summarized as:

• We propose a unified framework for manipulation artifact representation that integrates multi-dimensional feature extraction with hierarchical dynamic fusion. This design captures both low-level forensic artifacts and high-level semantic inconsistencies, enabling more robust detection and improved generalization.

• We propose a confidence-guided mechanism that refines localization maps and improves detection accuracy dynamically. By suppressing false alarms and recovering in complete regions, it provides more trustworthy evidence for decision-making.

• We achieve state-of-the-art performance on the IMDL-Benco benchmark [1] for image manipulation detection. Extensive experiments on multiple datasets further demonstrate the effectiveness and generalization of the proposed approach.

2 Related Work

2.1 Tampering Artifacts

The theoretical basis of IMD assumes that tampering introduces pixel-level artifacts such as JPEG blockiness, boundary inconsistencies, and noise anomalies [2]. Early IMD approaches relied on handcrafted features like DCT-based JPEG Ghost detection [3] and CFA irregularities [4], but their fixed or single-feature representations were easily disrupted by post-processing and multi-step editing operations. With the development of deep learning, data-driven representations have become dominant. SRM-based filters, Noiseprint, and other forensic trace modeling modules have significantly improved the extraction of tampering artifacts [5],[6]. In addition, most methods adopt dual-branch architectures to fuse RGB and frequency-domain cues. However, they often employ static fusion mechanisms and only cover limited artifact types, restricting the adaptability to diverse manipulation pipelines [7].

2.2 Manipulation Detection

Early tampering detectors produced only binary decisions or image-level scores, which limited their applicability in real-world scenarios. Subsequent works integrated pixellevel localization with authenticity assessment, forming a bimodal paradigm that outputs both detection scores and localization heatmaps. However, localization maps often suffer from instability, including isolated false positives, blurred boundaries, and missing regions in low-contrast areas [8]. These issues can propagate errors to the detection stage, since many pipelines derive global scores from localization statistics. Consequently, the results are less useful for high-precision applications. To improve reliability, TruFor [9] introduced pixel-wise confidence estimation to calibrate detection scores, yet its confidence maps primarily refine image-level predictions and do not effectively enhance localization quality.

2.3 Confidence and Interpretability

Confidence modeling and uncertainty estimation are widely used in safety-critical domains such as medical imaging [10] and autonomous driving [11] to quantify the prediction reliability and prevent over-confident errors. In contrast, IMD methods based on deep learning still lack interpretability, as CNNs typically generate localization maps

through a black-box process. The resulting heatmaps offer limited insight into the model's decision rationale. Such instability, including inconsistent outputs across samples and the coexistence of false positives with correctly localized tampered regions, further undermines user trust. These limitations highlight the need for more reliable and interpretable IMD outputs.

3 Proposed Method

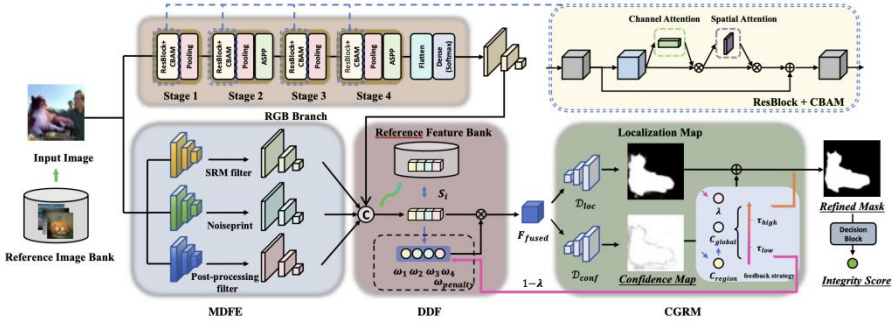

Fig. 2. Overview of ForgerySpotter for image manipulation detection.

To address the challenges of limited generalization, imprecise localization, and insufficient interpretability, we propose a unified framework named ForgerySpotter (see Fig. 2 for the network architecture). The framework consists of three core components: (*i*) multi-dimensional feature extraction (MDFE), (*ii*) dynamic feature fusion (DFF), and (*iii*) confidence-guided refine mechanism (CGRM). Specifically, the multi-dimensional extractor first aggregates RGB cues and frequency-domain artifacts. Then, the dynamic fusion module integrates these features effectively through dynamic weighting. Finally, the confidence-guided refine mechanism refines the localization maps based on the fused representations. These three components are tightly interconnected. Enriched multi-domain features strengthen the evidence for accurate localization and improve generalization, while reliable localization in turn supports more trustworthy confidence estimation. The estimated confidence information further feeds back to refine the localization map, thereby enhancing both interpretability and robustness. The proposed framework is detailed as follows.

3.1 Multi-Dimensional Feature Extraction

To comprehensively capture diverse tampering artifacts, we propose a multi-dimensional feature extractor based on a dual-branch architecture. The RGB branch focuses on spatial texture features, while the frequency one integrates specialized noise submodules. This design enables effective modeling of diverse manifestations of tampering artifacts.

The RGB branch is built upon the ResNet-50 backbone and is lightly modified to enhance its sensitivity to manipulation-related cues. Specifically, Convolutional Block

Attention Modules (CBAM) are introduced to emphasize informative channels and spatial regions associated with abnormal textures, while suppressing background responses. And an Atrous Spatial Pyramid Pooling (ASPP) module is employed to aggregate contextual information at multiple receptive fields. Together, these modifications enable the RGB branch to produce multi-scale feature representations.

Complementary to the RGB branch, we design a frequency-domain noise branch to explicitly capture tampering artifacts that are difficult to observe in the spatial domain. This branch consists of three specialized noise extractors with distinct sensitivities to different types of artifacts. First, an SRM-based filter is adopted to extract residual noise patterns that reflect local statistical inconsistencies introduced by manipulation. Such high-frequency residuals are known to be effective in revealing subtle tampering traces, especially under compression or smoothing operations. Second, Noiseprint is employed to model camera-dependent noise characteristics. Since manipulation often disrupts the inherent imaging pipeline, inconsistencies in camera noise provide valuable cues for detecting splicing or copy-move operations. Third, a post-processing noise extractor is designed to capture artifacts introduced by common image processing operations, such as resampling, compression, and filtering, which frequently appear in real-world manipulation pipelines.

These three noise extractors are complementary rather than redundant, as they respond to different manipulation operations and processing characteristics. Their outputs are jointly aggregated to form a robust frequency-domain representation, which is subsequently aligned with the RGB features at corresponding scales for downstream fusion.

3.2 Dynamic Feature Fusion

To address the limited adaptability of traditional fixed-weight fusion under diverse tampering scenarios, we propose a dynamic fusion mechanism based on feature contribution scores. In this mechanism, multi-scale feature maps from different extractors are adaptively integrated according to their relevance to tampering cues.

Specifically, the contribution score S_i of each feature map F_i is defined as

$$S_i = 1 - \frac{1}{N}\sum_{k=1}^{N} cos(F_i, F_{0k}) \tag{1}$$

where F_i represents the feature map extracted from the current sample, and F_{0k} are the feature maps from untampered reference images in the database. The function $\cos(\cdot,\cdot)$ denotes the cosine similarity measure computed between scale aligned feature representations. The reference feature bank is constructed from a set of generic untampered images during training to model the distribution of non manipulated features. Importantly, the reference feature bank is fixed after training and does not require access to any external image database at inference period. The comparison is performed against a fixed feature distribution rather than matched reference images. Therefore, the proposed method remains directly comparable to other blind detectors.

The contribution score S_i quantifies the importance of each scale's feature map and serves as the basis for dynamic weight assignment. A larger S_i indicates that the feature map deviates more from untampered references and thus contributes more to tampering

detection. For features captured at different scales, the tampering contribution score is inversely correlated with similarity to untampered reference features.

The normalized fusion weights ω_i are computed as

$$\omega_i = \frac{S_i}{\sum_{j=1}^{M} S_j} \tag{2}$$

where M denotes the total number of feature maps across all scales.

The fused multi-scale feature map F_{fused} is obtained via weighted summation

$$F_{fused} = \sum_i \omega_i F_i \tag{3}$$

To prevent any single feature from dominating the fusion, the weight ω_i is constrained by an upper bound of 0.3, empirically determined to promote balanced and complementary integration across multiple scales.

3.3 Confidence-Guided Refine Mechanism

To improve localization accuracy and reduce false positives in complex tampering scenarios, we propose a confidence-guided refine mechanism. This mechanism incorporates confidence estimation and confidence-guided localization refinement to enhance the stability and reliability of manipulation detection. Unlike TruFor, which primarily uses confidence maps to adjust final detection scores, our approach enables refinement of localization. Furthermore, since Trufor freezes its core parameters during the second training phase, its confidence generation process remains static, limiting adaptability and robustness in complex scenarios like weak noise and post-processing operations. The confidence estimation module takes the fused feature map F_{fused} as input and employs a lightweight CNN classifier, producing a tamper region probability map P_{pred} with pixel values in [0,1]. Two principal confidence metrics are then defined.

Regional Confidence C_{region} computes the mean prediction probability within each connected tamper region R_{tamper}. Here, N denotes the number of pixels in R_{tamper}. $C_{region}, k \in [0,1]$, and a higher value of C_{region}, k indicates greater reliability.

$$C_{region} = \frac{1}{N} \sum_{x,y \in R_{tamper}} P_{pred}(x, y) \tag{4}$$

Global Confidence C_{global} quantifies the overall detection reliability by weighting the confidence of each region C_{region}, k according to its relative area.

$$C_{global} = \sum_{k=1}^{K} (C_{region}, k \times \frac{N_k}{N_{total}}) \tag{5}$$

where K is the total number of predicted connected tampering regions, C_{region}, k and N_k denote the confidence and the number of pixels in the k-th region, respectively, and N_{total} represents the total number of pixels in the image.

To guide accurate manipulation localization, confidence thresholds are τ_1 and τ_2 ($\tau_1 > \tau_2$) are established. Refine strategies are selected according to C_{global}.

- For low-confidence cases($C_{global} \leq \tau_2$) , the localization is considered unreliable, and an enhanced correction strategy is activated. This strategy prioritizes features sensitive to weak noise and post-processing traces. Specifically, the response weights used for localization reconstruction are reweighted across domains, and a

boundary-clarity penalty term λ is introduced to suppress responses with blurred edges. The penalty adjustment is defined as

$$\omega_{penalty} = \omega_{old} \times (1 - \lambda) \quad (6)$$

- For medium-confidence cases ($\tau_2 < C_{global} < \tau_1$), local confidence values are evaluated. For regions where local confidence C_{region}, k is below the regional average $\overline{C}_{region}$, conservative corrections are applied to reduce false positives. For regions with higher confidence, localization results are reinforced to better capture tampering features.
- For high-confidence cases ($C_{global} \geq \tau_1$), the localization is regarded as reliable and directly used as the final output without adjustment.

In summary, the proposed confidence feedback mechanism effectively addresses the limitations of static feature fusion and confidence-based correction methods such as TruFor. By leveraging confidence to guide localization refinement, it substantially improves detection accuracy and robustness in complex tampering environments.

4 Experiments

4.1 Experimental Settings

Table I. Summary of datasets used in experiments.

Dataset	Images	Characteristics
CASIA v2.0	7,491/5,123	splicing and copy-move forgeries
CASIA v1.0	800/921	official pixel-level masks
NIST16	0/564	splicing, copy-move, removal
Columbia	180/180	splicing forgeries
COVERAGE	100/100	copy-move forgeries
IMD2020	414/2,010	realistic manipulations with post-processing

Datasets. As shown in Table I, we evaluate our method on five standard datasets: *CASIA*, *NIST16*, *Columbia*, *COVERAGE*, and *IMD2020*. We follow the standard protocol by training on CASIA v2.0 and testing on other datasets. All experimental settings follow the protocols in IMDL-Benco.

Evaluation metrics. The model is evaluated at pixel level with the benchmark metrics: F1 score and Intersection over Union (IoU). The F1 score reflects detection robustness by balancing precision and recall, while IoU emphasizes the spatial accuracy of localized manipulation regions.

Implementation details. The input images are resized to 512×512. In this work, the backbone is enhanced ResNet-50 [18], pre-trained on ImageNet [11]. Implemented in PyTorch and IMDL-Benco, our model is trained on two NVIDIA GeForce RTX 3090 GPUs. We adopt Adam optimizer with a learning rate decayed from 10−4 to 10−7. Training is conducted for 150 epochs with a batch size of 8

4.2 Comparison with state-of-the-art approaches

We compare our method against existing state-of-the-art approaches on five standard datasets. Qualitative results are shown in Fig. 3. Table II reports pixel-level F1 performance under the MVSS protocol, while Table III presents IoU results. Across both metrics, our method achieves the highest average performance or competitive results across most datasets, indicating consistent improvements in both localization accuracy and cross-dataset generalization. In particular, our approach shows substantial gains on the more challenging datasets such as COVERAGE and IMD2020, where traditional methods relying on limited forensic artifacts often struggle. On IMD2020, our approach improves pixel-level F1 by 3.4% and IoU by 3.1% over the strongest baseline, highlighting its robustness under realistic and diverse manipulation pipelines. Similarly, on the COVERAGE dataset, our method achieves a 2.9% improvement in IoU. These results highlight the effectiveness of our method in capturing complementary cues and producing of our method in capturing complementary cues and producing more reliable localization predictions.

Table II. Pixel-level F1 Performance.Authentic images are not used . The best and second-best performances for each dataset are highlighted in **bold** and underlined, respectively.

Protocol	Method	COVERAGE	Columbia	NIST16	CASIAv1	IMD2020	Average
Protocol-MVSS	Mantra-Net [8]	0.090	0.243	0.104	0.125	0.055	0.123
	MVSS-Net [12]	0.259	0.386	0.246	0.534	0.279	0.341
	CAT-Net [13]	0.296	0.584	0.269	0.581	0.273	0.401
	ObjectFormer [14]	0.294	0.336	0.173	0.429	0.173	0.281
	PSCC-Net [15]	0.231	0.604	0.214	0.378	0.235	0.333
	IML-ViT [16]	0.225	0.446	0.260	0.502	0.237	0.334
	TruFor [9]	**0.419**	0.865	0.324	**0.721**	0.322	0.530
	ours	0.414	**0.876**	**0.347**	0.718	**0.356**	**0.542**

Table III. Pixel-level IoU Performance. The best and second-best results are highlighted in **bold** and underlined, respectively.

Protocol	Method	COVERAGE	Columbia	NIST16	CASIAv1	IMD2020	Average
Protocol-MVSS	Mantra-Net	0.066	0.348	0.135	0.092	0.062	0.141
	MVSS-Net	0.192	0.290	0.176	0.440	0.201	0.260
	CAT-Net	0.204	0.469	0.196	0.509	0.198	0.315
	ObjectFormer	0.181	0.224	0.111	0.320	0.106	0.188
	PSCC-Net	0.146	0.477	0.150	0.310	0.157	0.248
	IML-ViT	0.372	0.685	0.250	0.648	0.264	0.444
	TruFor	0.352	0.832	0.271	**0.666**	0.267	0.478
	ours	**0.381**	**0.851**	**0.286**	0.654	**0.298**	**0.494**

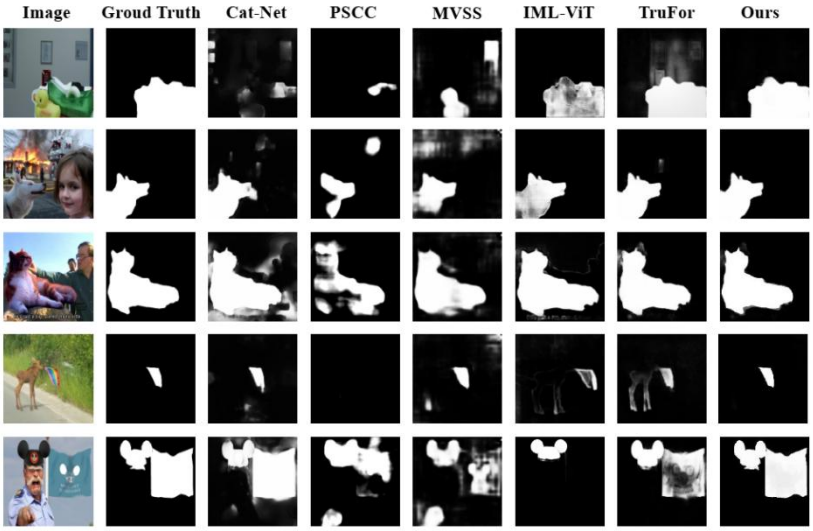

Fig. 3. Qualitative comparison of forgery detection on some testing images.

4.3 Ablation study

Table IV. Summary of datasets used in experiments.

Method	CASIA (F1/IoU)	NIST16 (F1/IoU)
Baseline (ResNet-50)	0.632 / 0.581	0.268 / 0.203
w/o MDFE & DFF	0.681 / 0.617	0.305 / 0.236
w/o CGRM	0.704 / 0.632	0.326 / 0.251
Ours (ForgerySpotter)	0.718 / 0.654	0.347 / 0.286

To validate the effectiveness of key components in our framework, we conduct ablation studies on the CASIA and NIST16 datasets. The quantitative results are summarized in Table IV. The baseline model employs a standard ResNet-50 backbone without any proposed module. Removing CGRM results in a noticeable performance drop (CASIA F1: 0.718→0.704; NIST16 IoU: 0.286→0.251), indicating its contribution to enhancing boundary consistency and fine-grained localization. When MDFE & DFF are removed, the degradation becomes more pronounced, especially on NIST16 (F1: 0.347→0.305; IoU: 0.286→0.236), highlighting the critical role of complementary forensic cues under cross-dataset conditions.

4.4 Sensitivity Analysis

To analyze the impact of key parameters in the confidence refine mechanism on model performance, we conduct sensitivity experiments on global confidence thresholds , and penalty coefficient on the CASIA v1 dataset, with results shown in Table V.

Table V. Summary of datasets used in experiments.

Parameters Setting	τ_1	τ_2	λ	F1/IoU
A	0.65	0.35	0.20	0.714/0.652
B	0.70	0.30	0.20	**0.715/0.656**
C	0.80	0.50	0.20	0.713/0.653
D	0.75	0.30	0.10	0.716/0.651
E	0.70	0.30	0.30	0.712/0.649

The experimental results show that the model achieves balanced pixel-level F1-score and IoU when thresholds are set to Setting B. In contrast, performance decreases slightly when thresholds are too low (Setting A) or too high (Setting C), as over-relaxed or over-strict thresholds reduce the effectiveness of feedback refinement.

With fixed thresholds, a too small penalty coefficient (Setting D) leads to insufficient boundary suppression and a slight IoU drop; a too large one (Setting E) causes over suppression, weakening real tampered region responses and degrading overall performance. Thus, Setting B is selected as the default parameter configuration.

4.5 Robustness Evaluation

We further evaluate the robustness of our method under two realistic and challenging scenarios: (***i***) image degradation commonly introduced by *Online Social Networks* (OSNs) and (***ii***) emerging forgeries generated by advanced generative models.

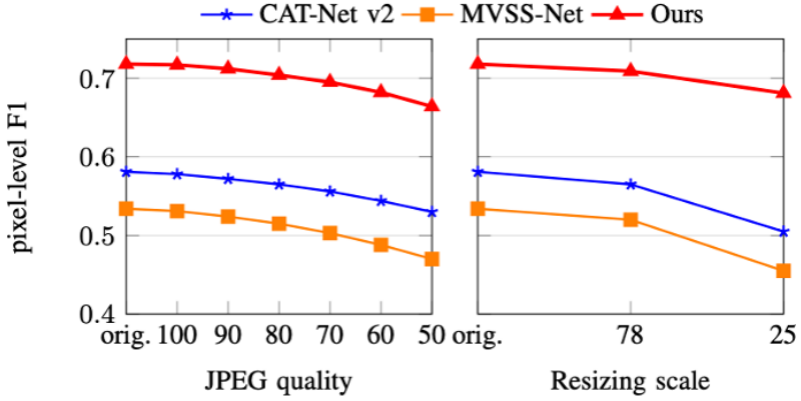

Fig. 4. Robustness Evaluation under JPEG compression and resizing.

Robustness to Social Media Degradation. In OSN-style processing, images typically undergo aggressive JPEG compression and resizing, which may suppress or erase subtle forensic artifacts. We simulate it by applying random JPEG compression (quality factors between 50 and 90) and resizing operations to the CASIAv1 datasets. As shown in Fig. 4, our method consistently outperforms state-of-the-art approaches under these degradations. This robustness can be attributed to our multi-dimensional feature representation.

Generalization to Advanced Generative Manipulations. We further evaluate our method on manipulations produced by a recent GPT-based image generation model. As

shown in Fig. 5, such semantically coherent and artifact-sparse forgeries remain challenging for all detectors. Nevertheless, the pixel-level confidence maps produced by our framework offer valuable interpretability by highlighting regions of high uncertainty, which often correspond to the subtle modifications introduced by Nano Banana. This indicates that our confidence feedback mechanism can capture the inherent ambiguity of these sophisticated forgeries, providing useful diagnostic cues even when binary localization is imperfect.

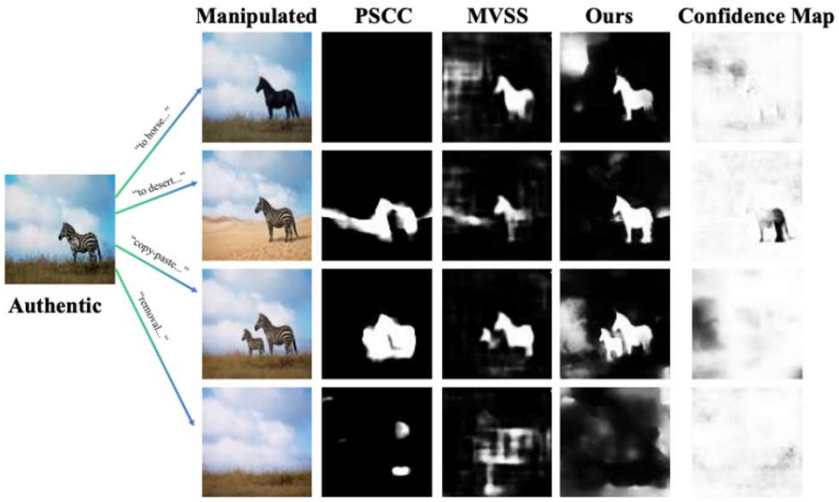

Fig. 5. Qualitative results on advanced generative manipulations.

5 Conclusion and Future work

In this paper, we address three key challenges in image manipulation detection, including limited generalization, inaccurate localization and poor interpretability. To overcome these issues, we propose a unified framework that combines multi-dimensional feature extraction, dynamic feature fusion, and a confidence-guided refine mechanism. Built upon a dual-branch architecture, our approach effectively captures tampering artifacts through multiple extractors with complementary sensitivities. The proposed dynamic fusion strategy aggregates multi-scale contextual information via adaptive weighting, while the confidence-guided mechanism refines localization maps and enhances interpretability. Extensive experiments demonstrate that our approach achieves state-of-the-art performance across multiple benchmarks and exhibits strong generalization. These results highlight its potential for practical and trustworthy manipulation detection applications. However, optimizing the confidence map refinement to improve adaptability under diverse real-world conditions is still an important task, which we will focus on in future work.

References

1. Ma et al., “Imdl-benco: A comprehensive benchmark and codebase for image manipulation detection & localization,” Advances in Neural Information Processing Systems, vol. 37, pp.134591–134613, 2024.
2. Paola Capasso, Giuseppe Cattaneo, and Maria De Marsico, “A comprehensive survey on methods for image integrity,” ACM Transactions on Multimedia Computing, Communications and Applications, vol. 20, no.11, pp. 1–34, 2024.
3. Hany Farid, “Exposing digital forgeries from jpeg ghosts,” IEEE transactions on information forensics and security, vol. 4, no. 1, pp.154–160, 2009.
4. Alin C Popescu and Hany Farid, “Exposing digital forgeries in color filter array interpolated images,” IEEE Transactions on Signal Processing, vol. 53, no. 10, pp. 3948–3959, 2005.
5. HyunJae Lee, Hyo-Eun Kim, and Hyeonseob Nam, “Srm: A style-based recalibration module for convolutional neural networks,” in Proceedings of the IEEE/CVF International conference on computer vision, 2019, pp.1854–1862.
6. Davide Cozzolino and Luisa Verdoliva, “Noiseprint: A cnn-based camera model fingerprint,” IEEE Transactions on Information Forensics and Security, vol. 15, pp. 144-159, 2019.
7. Zhu et al., “Mesoscopic insights: orchestrating multi-scale & hybrid architecture for image manipulation localization,” in Proceedings of the AAAI Conference on Artificial Intelligence, 2025, vol. 39, pp. 11022–11030.
8. Yue Wu, Wael AbdAlmageed, and Premkumar Natarajan, “Mantra-net: Manipulationtracing network for detection and localization of image forgeries with anomalous features,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 9543–9552.
9. Fabrizio Guillaro et al., “Trufor: Leveraging all-round clues for trustworthy image forgery detection and localization,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 20606–20615.
10. Christian Leibig et al., “Leveraging uncertainty information from deep neural networks for disease detection,” Scientific reports, vol. 7, no. 1, pp. 1–14, 2017.
11. Alex Kendall and Yarin Gal, “What uncertainties do we need in Bayesian deep learning for computer vision?,” Advances in neural information processing systems, vol. 30, 2017.
12. Chengbo Dong, Xinru Chen, Ruohan Hu, Juan Cao, and Xirong Li, “Mvss-net: Multi-view multi-scale supervised networks for image manipulation detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 3, pp. 3539–3553, 2022.
13. 13. Myung-Joon Kwon et al., “Cat-net: Compression artifact tracing network for detection and localization of image splicing,” in Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2021, pp. 375–384.
14. Junke Wang et al., “Objectformer for image manipulation detection and localization,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 2364–2373.
15. 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, vol. 32, no. 11, pp. 7505–7517, 2022.
16. Xiaochen Ma et al., “Iml-vit: Benchmarking image manipulation localization by vision transformer,” arXiv preprint arXiv:2307.14863, 2023.