

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-18

论文引用格式: Xue Jun-Lin, Tang Yun-Qi, Wu Si-Kang, Yu Bo-Ya. Direction-aware Cross-modal Transformer for Image Tampering Localization [J]. Journal of Image and Graphics, XXXX:1-18. DOI: 10.11834/jig.260002. (薛峻凜, 唐云祁, 吴思慷, 余博雅. 融合方向先验的跨模态图像篡改定位方法[J/OL]. 中国图象图形学报, XXXX:1-18. DOI: 10.11834/jig.260002.) [DOI: 10.11834/jig.260002]

融合方向先验的跨模态图像篡改定位方法

薛峻凜, 唐云祁*, 吴思慷, 余博雅

中国人民公安大学 侦查学院, 北京 100038

摘要: 随着图像编辑技术和生成式模型的快速发展, 数字图像在社交媒体、新闻传播和司法鉴定等场景中更易受到多类型篡改与多轮压缩、重采样等处理, 导致篡改区域仅保留稀疏且弱激活的残差伪迹, 并呈现明显的方向依赖性和模态不一致性。这类弱残差跨越语义区域, 易被深层语义特征抹除, 使现有方法在真实退化条件下难以稳定定位篡改区域。同时, 现有研究普遍受到方向建模缺失、跨模态融合策略静态、空间与通道增强割裂等限制, 难以充分挖掘在红绿蓝 (red, green, and blue, RGB) 外观与噪声残差中隐含的取证线索。针对上述问题, 本文提出一种方向感知跨模态图像篡改定位框架 DA-CMTL (Direction-aware Cross-modal Multi-level Tampering Localization)。该框架以 RGB 图像域与噪声特征域为互补信息源, 构建跨模态特征表示框架, 引入方向感知机制以显式建模篡改区域在不同方向上的结构一致性。同时, 通过跨模态推理模块对不同模态特征进行自适应交互与联合推断, 从而增强对篡改区域边界和细粒度结构特征的表达能力。在多个公开图像篡改定位数据集上进行实验验证, 结果表明, 所提出的方法在 F1 值 (F1-score) 和 交并比 (intersection over union, IoU) 等评价指标上均优于多种主流篡改定位方法, 尤其在复杂篡改类型及多重后处理条件下表现出更高的定位精度与稳定性。实验结果表明, 方向感知跨模态推理机制能够有效提升模型对篡改区域结构信息的建模能力, 为复杂场景下的图像篡改定位提供了一种具有较强鲁棒性和泛化能力的解决方案。代码已开源于 <https://www.scidb.cn/anonymous/cnFFZll2>

关键词: 图像篡改定位; 跨模态特征融合; 方向感知; 深度学习; 图像取证

Direction-aware Cross-modal Transformer for Image Tampering Localization

Xue Jun-Lin, Tang Yun-Qi*, Wu Si-Kang, Yu Bo-Ya

School of Criminal Investigation, People's Public Security University of China, Beijing 100038, China

Abstract: With the rapid development of image editing tools, mobile image processing applications, and generative artificial intelligence techniques, visually realistic manipulated images can now be produced with increasing ease, which poses serious challenges to the authenticity and credibility of digital visual content. In practical scenarios, tampered images are often subjected not only to splicing, copy-move, or content removal operations, but also to multiple post-processing procedures, such as compression, resampling, filtering, smoothing, and format conversion. These operations tend to weaken the forensic traces left by tampering and make the manipulated regions exhibit weak residual artifacts, ambiguous boundaries, and inconsistent structural cues. In addition, the discriminative clues useful for tampering localization are usually distributed across different feature domains. RGB images provide rich semantic appearance information, whereas noise-

收稿日期: 2026-01-02; 修回日期: 2026-03-28

* 通信作者: 唐云祁, E-mail: tangyunqi@ppsuc.edu.cn

基金项目: 中国人民公安大学刑事科学技术双一流创新研究项目 (项目编号: 2023SYL06)

Supported by: Double First-Class Innovation Research Project of Criminal Science and Technology, People's Public Security University of China (Project No.: 2023SYL06)

related representations contain subtle forensic evidence that is less dependent on visual semantics. However, existing methods often fail to fully exploit the complementarity between these heterogeneous modalities, and they remain vulnerable to semantic interference, weak trace attenuation, and complex degradation conditions. To address these problems, a direction-aware cross-modal multi-level tampering localization framework, termed DA-CMTL, was proposed for robust image tampering localization. The proposed method aimed to improve the representation and reasoning of manipulation traces by jointly modeling appearance information, forensic residual information, and directional structural priors in a unified framework. Specifically, a dual-branch architecture was constructed to extract complementary features from the RGB domain and the noise domain, respectively. The RGB branch was used to preserve contextual and structural appearance cues, while the noise branch was designed to highlight subtle statistical anomalies and residual inconsistencies caused by tampering. To enhance the discriminative ability of the extracted features, a hierarchical spatial-channel attention mechanism was introduced. This mechanism enabled the network to focus on suspicious local regions in the spatial dimension while adaptively emphasizing manipulation-sensitive responses in the channel dimension, thereby improving the perception of weak tampering traces and reducing the influence of irrelevant background content. Furthermore, considering that manipulated boundaries and residual artifacts often exhibit directional dependency, a multi-directional cross-modal fusion module was designed to incorporate horizontal, vertical, and diagonal directional priors into the feature interaction process. Through explicit direction-aware modeling, the proposed framework was able to better capture structural continuity, boundary variation, and anisotropic trace distribution in manipulated regions. In addition, an adaptive cross-modal gating strategy was employed to achieve collaborative fusion between RGB features and noise features. This strategy allowed the network to suppress redundant or conflicting information across modalities and selectively strengthen integrity-critical cues that were more informative for tampering localization. Extensive experiments were conducted on multiple public image tampering localization datasets containing diverse manipulation types and challenging post-processing conditions. The effectiveness of the proposed method was evaluated by using quantitative metrics such as F1-score and intersection over union, together with qualitative comparisons, ablation studies, and robustness analyses. Experimental results showed that the proposed DA-CMTL framework achieved superior or highly competitive performance compared with several mainstream tampering localization methods on different benchmarks. In particular, the proposed method demonstrated stronger robustness and more stable localization performance in scenarios involving complex manipulations, weak residual traces, and multiple post-processing disturbances. The predicted tampering masks were more accurate and structurally complete, especially for irregular boundaries, fine-grained manipulated regions, and semantically confusing areas. Comparative analysis further indicated that some existing methods were prone to semantic-related false detections or incomplete region localization when forensic clues became weak or fragmented, whereas the proposed framework was more effective in suppressing semantic interference and preserving detailed manipulation structures. The ablation results verified that each designed component contributed positively to the final performance, and the combination of hierarchical attention, direction-aware modeling, and adaptive cross-modal fusion yielded the most significant improvement. These findings demonstrated that explicitly integrating directional priors and cross-modal complementary reasoning was beneficial for enhancing image tampering localization under realistic conditions. Overall, the proposed method provided an effective solution for modeling weak and direction-sensitive manipulation traces, and it showed promising application value in multimedia forensics, digital evidence analysis, misinformation detection, and trustworthy image authentication.

Key words: image tampering localization; cross-modal feature fusion; direction-aware modeling; deep learning; image forensics

论文引用格式: Xue Jun, Lin, Tang Yun, et al. Direction-aware Cross-modal Transformer for Image Tampering Localization [J]. ****, ****, **(**):00-00 (薛峻凜, 唐云祁, 吴思慷, 等. 融合方向先验的跨模态图像篡改定位方法 [J]. ****, ****, **(**):00-

00) [DOI:]

0 引言

随着图像编辑工具与生成式模型的持续演进,
© 中国图象图形学报版权所有

数字图像篡改在社交媒体、新闻传播、司法鉴定和公共安全等领域愈加普遍(Omer等,2023)。现代篡改操作(如拼接、复制-移动、修补/重采样)通常伴随平滑、压缩与重采样等后处理步骤,使得篡改区域的可见伪迹被显著弱化,仅残留微弱、模态依赖且具有方向敏感性的残差结构(Alex等,2022)。这些弱残差包括边界结构不连续、噪声水平变化,以及沿水平、

垂直或对角方向出现的轴向纹理扰动(如联合图像专家组(Joint Photographic Experts Group, JPEG)块效应、重采样纹理偏移等(Cristian等,2020)。如图1所示,这类证据跨越不同语义层次,具有稀疏、弱激活和易被压缩破坏等特征,使得依赖单模态或各向同性算子的传统方法难以稳健捕获(Michele等,2023)。

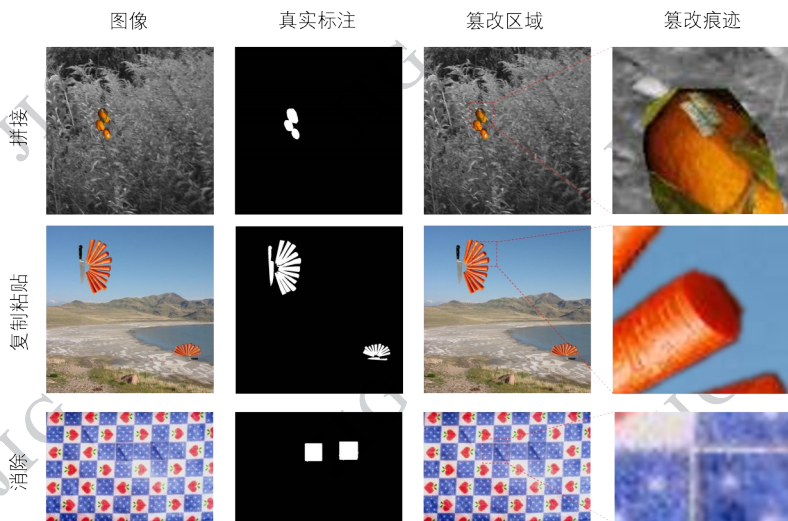


图1 展示了三种不同类型篡改及其对应伪迹的示例。

Fig. 1 Illustrative examples of different manipulation types (splicing, copy - move, and inpainting/resampling) and their heterogeneous forensic evidence.

现有篡改定位方法大致可分为手工取证特征、基于卷积神经网络(convolutional neural networks, CNN)的纹理/噪声学习、频域或噪声域的统计建模,以及基于Transformer的全局推理与跨模态融合等方向(Oran等,2020)。尽管这些方法在不同维度上取得一定进展,但在真实退化条件下普遍存在以下三个相互关联、并会形成负向循环的核心问题:

1)跨模态交互缺乏自适应性

RGB外观与空间丰富模型(spatial rich model, SRM)残差(Jessica等,2012)在不同压缩强度、噪声条件和纹理复杂度下的可靠性差异显著,但许多方法仍采用静态加权或对称融合策略。当局部纹理复杂度较高或压缩强度加大时,固定融合规则容易导致模态主导性失衡,使弱残差被高层语义淹没,进一步削弱整体定位稳定性。

2)缺乏显式方向建模机制

卷积与自注意力均为各向同性算子,而篡改操作产生的JPEG水平块效应、重采样方向纹理偏移与插值导致的对角平滑结构具有明确方向性

(Berkay,2023),是典型的统计不一致性。这类方向性残差往往跨越语义区域,与SRM噪声残差存在天然相关性。如果缺乏显式方向先验,模型将难以区分真实纹理方向与篡改扰动,使跨模态融合缺乏可靠的低层对比基础(蒋小玉等,2021)。

3)空间与通道增强相互割裂

空间位置异常(如伪边界)与通道级噪声偏移(如高频扰动)共同构成篡改证据(Wei等,2011);两者具有天然耦合关系。然而现有注意力机制通常分别对空间或通道建模(蔺聪等,2024),导致方向扰动在深层特征中被语义平滑抹除,使方向信息与噪声残差难以在后续跨模态阶段有效对齐,从而进一步削弱模态自适应能力。

需要强调的是,这三个问题呈现出清晰的因果链条关系:

空间-通道割裂(问题3)使方向扰动在深层被平滑,减弱方向建模(问题2);方向先验缺失(问题2)导致模型无法可靠区分噪声残差与纹理细节,造成模态权重错配(问题1);模态不平衡(问题1)又反

过来增强空间/通道误抑制现象,使弱残差进一步消失。

这一负向循环正是现有方法在真实压缩/重采样场景下性能不稳定的根本原因。

为解决以上系统性挑战,本文提出 DA-CMTL (Direction-aware Cross-modal Multi-level Tampering Localization) 框架,将方向先验、模态自适应与跨维度耦合推理统一于同一 Transformer 中。需要特别指出的是,弱残差往往呈间断分布,其相关性可能跨越语义边界;跨模态交互与方向先验也依赖远程统计关系。卷积依赖局部邻域,需堆叠大量层才能覆盖足够感受野,并易在堆叠过程中引入平滑效应,使噪声判据进一步弱化。相比之下,Transformer 的全局依赖建模可直接跨越弱残差的不连续区域,对齐跨模态与方向域的稀疏证据,使其成为构建三类能力协同推理机制的最适合甚至唯一可行的基础架构。

基于上述动机,本文设计两个关键模块:

HSA (Hierarchical Spatial - Channel Attention): 对应解决问题3。通过联合建模空间显著性与通道噪声统计,显式捕获二者的耦合结构,使结构异常与噪声偏移的共现模式得到强化,避免弱残差在深层语义中被抹除。

MCFM (Multi-directional Cross-modal Fusion Module): 对应解决问题2。利用水平、垂直与对角方向的深度可分离滤波构建方向先验,并通过跨模态门控自适应调节 RGB 与 SRM 特征权重。方向扰动与噪声残差本质上是同类统计不一致性,因此在方向滤波后执行自适应模态交互可显著增强噪声证据,为跨模态融合提供稳定物理依据。

二者协同作用解决问题1,在 Transformer 框架中形成方向建模、模态自适应与跨维度耦合推理的闭环,使模型在不同篡改方式和真实压缩/重采样条件下仍保持稳定与高敏感性的弱残差捕获能力。

本文的主要贡献如下:

1) 提出 DA-CMTL 框架,首次从方向先验、模态自适应和跨维度耦合三个维度构建可协同作用的统一范式,为弱残差增强与鲁棒篡改定位提供系统性设计思路;

2) 提出 HSA 模块,通过空间 - 通道联合注意力显式建模结构异常与噪声偏移的耦合关系;

3) 提出 MCFM 模块,通过多方向滤波与跨模态

门控构建方向敏感且模态自适应的融合策略,显著提升弱残差的可辨识性;

4) 在五个公开基准与多种社交网络退化条件下取得领先性能,并通过系统性消融验证各模块的必要性、互补性与设计合理性。

全文结构如下:第2节综述相关工作;第3节描述 DA-CMTL 的总体框架;第4节呈现实验结果和分析;第5节总结全文并讨论未来方向。

1 相关工作

图像篡改定位技术从依赖人工设计的取证线索逐步演化为能够联合建模跨模态残差与全局结构的深度神经网络。为确保综述结构的可解释性与一致性,本文根据方法所依赖的核心证据来源(语义外观、噪声残差、频域先验)与建模机制(卷积特征学习、全局注意力、显式取证先验)而非网络类型进行分类,将现有方法划分为四类:(1)手工取证特征;(2)CNN 操纵特征学习;(3)频域与噪声域建模;(4)Transformer 跨模态全局推理。此分类标准可避免噪声域 CNN 与外观 CNN 的混合干扰,也有助于对比不同取证线索与推理方式的能力差异。

在综述这些方法时,本文重点关注三类与本文动机直接相关的“知识维度”:

- 1) 是否显式引入取证先验;
- 2) 是否包含方向敏感性;
- 3) 是否具备自适应跨模态交互能力。

1.1 手工取证线索

传统方法主要依靠人工设计的操作痕迹,包括 JPEG 压缩不一致性分析、彩色滤光阵列(color filter array, CFA)去马赛克模式检测、光响应非均匀性噪声(photo-response non-uniformity, PRNU)估计以及基于统计矩的复制 - 移动检测等。这类方法具备较强可解释性,但依赖于成像管线与压缩模式的严格假设。在真实社交平台压缩、重采样和多操作序列下,这些手工特征往往遭到严重破坏,导致鲁棒性不足。因此,后续研究逐渐转向数据驱动的深度特征学习,以获得更具泛化性的取证能力(Y. Tomioka 等, 2011; Alin 等, 2004; Tiberio 等, 2012; M. B. İmamoğlu 等, 2013; Micah 等, 2007; N. Kanagavalli 等, 2017)。

这些局限性促使研究转向数据驱动方法,以捕

表 1 代表性的图像篡改定位方法概览

Table 1 Overview of state-of-the-art image tampering localization methods.

Method	Backbone		Resolution	Manipulation supervision	dataset	
	CNN	Tran			Type	Amount
ManTra-Net	√	-	512x512	Noise	Private	102k
SPAN	√	-	224x224	Noise	Private	102k
GSR-Net	√	-	300x300	Edge Prediction	Public	5k
MVSS-Net	√	-	512x512	Noise+Edge	Public	5k
TruFor	√	√	512x512	Noise	Public	35k
CAT-Net	√	-	512x512	DCT	Public	910k
EITLNet	-	√	512x512	Noise	Public	60k
DA-CMTL	-	√	512x512	Noise	Public	62k

捉更具普适性的操纵残差。

1.2 基于 CNN 的操作特征提取

深度卷积网络能够从数据中自动学习低层噪声不一致性与高层结构异常。ManTra-Net(Yuezun 等, 2019)、SPAN(Xiaoyu 等, 2020)等方法展示了 CNN 在捕获伪边界与局部操纵残差方面的能力; GSR-Net(Peng 等, 2020)、MVSS-Net(Chongyang 等, 2023)通过引入边缘分支、多尺度监督等机制进一步提升了细粒度定位性能。

尽管 CNN 灵活且易训练,但其局部卷积结构存在两方面天然限制:感受野有限,难以对弱残差进行跨区域对齐建模;卷积核各向同性,使其难以捕捉 JPEG 块效应、重采样纹理偏移等方向性伪迹。

部分研究尝试利用 Sobel、Laplacian 或 Gabor 等方向滤波器强调梯度特性,但这些算子通常只提供单方向或固定方向的局部增强,且无法与深层跨模态推理机制协同工作。

1.3 频域与噪声域线索

噪声域与频域特征聚焦于语义无关弱残差,被证明在压缩与纹理复杂场景下具有较高稳定性。CAT-Net(Kwon 等, 2021)基于 DCT 块统计建模 JPEG 不一致性; TruFor(Fabrizio 等, 2023)结合 Noiseprint 与多尺度噪声统计捕捉通用操纵指纹。这类方法在鲁棒性方面具有优势,因为它们显式引入了取证先验而非纯语义特征。

然而,大部分频域/噪声域方法仍存在两方面限制:通常将 RGB 外观与噪声残差视为独立分支,采用静态拼接或固定注意力策略,难以在真实退化条

件下依据上下文动态调整模态权重;虽然噪声残差中包含某些方向性扰动,但缺乏与外观结构的深层对齐机制,使方向线索难以在后续推理中发挥作用。

1.4 基于 Transformer 的全局推理与跨模态融合

Transformer 通过全局自注意力建模长程依赖,天然适合对稀疏弱残差进行跨区域关联推理。EITLNet(Kun 等, 2024)、IML-ViT(Ma 等, 2023)、SparseViT(Su 等, 2025)等方法有效缓解了 CNN 在长距离结构推理上的局限,提高了边界保持性和跨区域一致性。

然而现有 Transformer 方法在跨模态取证场景中仍面临以下核心瓶颈:

1.4.1 模态交互机制大多为静态或弱交互

例如简单拼接、对称协同注意力等,无法适配 RGB 与噪声残差在不同压缩与纹理条件下的可靠性变化。

1.4.2 自注意力各向同性,缺乏显式方向建模能力
标准注意力机制难以捕捉 JPEG 块效应、重采样方向纹理偏移等关键方向扰动,导致弱残差在深层出现“被语义淹没”的现象。

1.4.3 空间-通道增强多以解耦方式建模

导致结构异常与通道噪声偏移无法形成有效耦合,限制了弱残差在深层特征中的表达能力。

这些观察与本文提出的三大缺口一一对应,是 DA-CMTL 构建方向感知、模态自适应与空间-通道联合推理框架的直接动机来源。

1.5 小结与动机

综上四类方法,可总结出有研究的三项根本

性不足:

1) 跨模态交互多为静态策略,缺乏对局部内容、压缩强度及噪声水平的自适应调节能力;

2) 缺少显式方向建模机制,难以捕捉 JPEG 重采样与插值造成的典型方向扰动;

3) 空间与通道增强常被割裂处理,使结构异常与噪声偏移的耦合关系在深层语义中被弱化。

因此,需要一种能够同时整合方向先验、模态自适应性与空间-通道耦合推理能力的统一框架,以充分挖掘弱残差线索并提升在真实退化场景下的鲁棒性。该需求直接引出了本文第 3 节提出的 DA-CMTL 模型设计。

2 本文方法

2.1 概述

如图 2 所示,DA-CMTL 采用双分支 Transformer:一支编码 RGB 外观特征,另一支编码 SRM 噪声残差。两条分支均使用四阶段的 MiT-B2 骨干网络,输出 1/4、1/8、1/16、1/32 多尺度特征图。借鉴 SegFormer,本文通过重叠式补丁嵌入提供隐式位置编码,以引入局部结构先验而无需绝对位置偏置;由于 MCFM 显式建模轴向(水平/垂直/对角)的方向先验,骨干的隐式位置信息与之构成互补而非冗余。

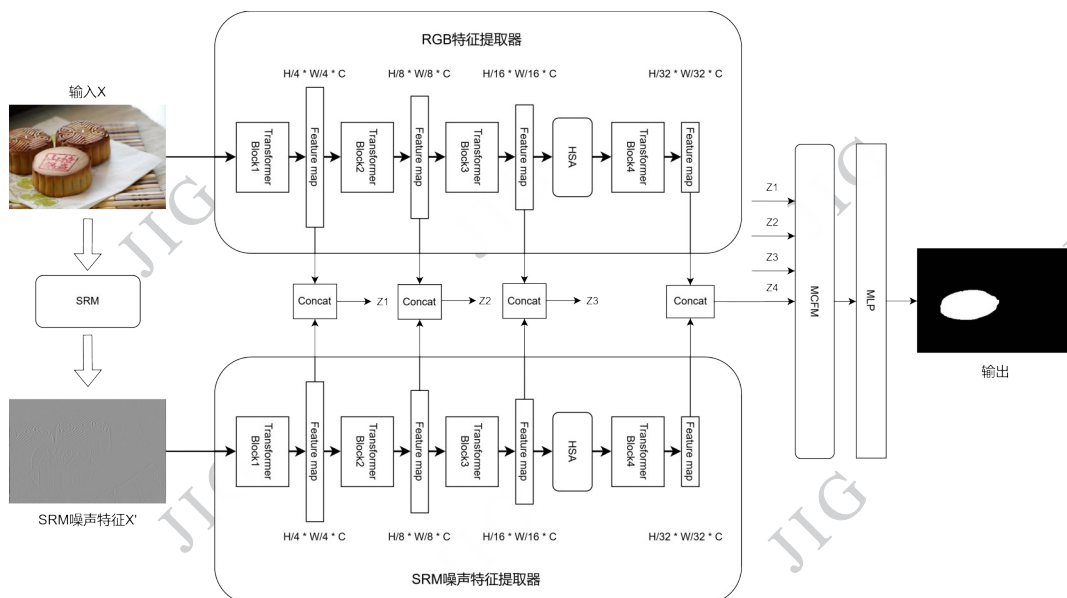


图2 所提出的 DA-CMTL 框架的总体架构

Fig. 2 Overall architecture of the proposed DA-CMTL framework

本文在第 3 阶段(1/16)之后插入 HSA,以在语义丰富度与取证敏感性之间取得最佳平衡:过早阶段难以得到稳定的操纵证据,而第 4 阶段的高层语义又易抑制微弱的残差线索;第 4.3 节的消融结果也印证了该选择。不同于 SegFormer 风格的解码器,DA-CMTL 保留四阶段输出,并在每一分辨率上施加 MCFM 以进行方向聚合与交互式跨模态门控,最后以轻量级 MLP 进行掩码预测。

与现有篡改定位体系相比,DA-CMTL 的三点核心差异为:

(1) 自适应跨模态推理: MCFM 采用交互式门控,随空间上下文动态调节 RGB 与残差线索的贡献;

(2) 显式方向先验:多方向滤波编码水平/垂直/对角线索,提升轴向伪迹的可探测性;

(3) 跨维注意耦合: HSA 联合建模空间-通道依赖,支撑分层强化。

该框架以显式取证先验为中心:多方向滤波实现方向先验,交互式门控落实跨模态“相互佐证”,而 SRM 残差提供与语义弱相关的噪声域证据(见第 3.2/3.3 节)。

2.2 HSA (Hierarchical Spatial - Channel Attention)

提出 HSA 的动机在于克服取证场景中常见注意力机制的基本局限:多数方法执行“单维度增强”,并隐含假设空间与通道可独立建模。经典模块

(SE、ECA、CBAM)可被理解为学习通道加权 $w_c = \phi_c(\mathbf{F})$ 或空间掩码 $w_s = \phi_s(\mathbf{F})$, 等价于将空间/通道的联合依赖近似为

$$p(\mathbf{S}, \mathbf{C} | \mathbf{F}) \approx p(\mathbf{S} | \mathbf{F})p(\mathbf{C} | \mathbf{F}), \quad (1)$$

这对通用识别任务尚可, 但对图像取证往往不足。

在操纵定位中, 伪迹很少仅表现于单一维度: 边界不一致同时改变局部梯度(空间域)与通道层面的残差统计分布; 重采样或压缩会产生轴向结构并伴随细微的通道偏移。因此更合理的描述是联合项 $p(\mathbf{S}, \mathbf{C} | \mathbf{F})$: 异常往往表现为空间不规则与通道统计偏移的共现。忽略这种耦合, 注意力可能(i)高亮无取证证据的语义显著区, 或(ii)放大与真实边界不对齐的弱通道响应。

HSA 明确近似这一联合依赖: 先分别提取反映噪声—纹理不一致性的通道描述子与捕捉边界/方向异常的空间描述子, 再融合为统一的调制权重, 作用于原特征。概念上, HSA 近似

$$\hat{\mathbf{F}} = \mathbf{F} \odot \sigma(\phi_s(\mathbf{F}) // \phi_c(\mathbf{F})), \quad (2)$$

设输入特征图为 $\mathbf{F} \in \mathbf{R}^{C \times H \times W}$, 其中 \mathbf{C} 表

示通道数, \mathbf{H} 和 \mathbf{W} 分别表示特征图的高度和宽度。为同时建模篡改线索在空间维和通道维上的分布特性, 本文构建层次化空间—通道注意力模块(HSA)对输入特征 \mathbf{F} 进行联合增强。式(1) - (2)中, $\mathbf{S}(\mathbf{F})$ 表示由输入特征 \mathbf{F} 提取的空间描述子, 用于表征篡改痕迹在不同空间位置上的异常响应; $\mathbf{C}(\mathbf{F})$ 表示由输入特征 \mathbf{F} 提取的通道描述子, 用于刻画各通道对取证线索的敏感程度; $\phi(\cdot)$ 表示跨维融合函数, 用于对空间描述信息与通道描述信息进行联合建模; $\hat{\mathbf{F}}$ 表示经 HSA 模块增强后的输出特征。这里的联合建模表示同时考虑空间响应与通道响应之间的相关性, 以提升模型对弱篡改痕迹的表征能力。

这一设计在 DA-CMTL 的双分支框架中尤为重要: 篡改痕迹往往弱且稀疏, 且易被高层语义抑制。通过“以通道统计约束空间定位, 反之亦然”, HSA 使网络聚焦于“结构异常与残差统计异常一致”的位置, 抑制语义相关的误报并增强微弱但可靠的取证证据。同时, HSA 采用与伪迹多尺度特性一致的分层设计: 浅层强调细粒度纹理异常, 深层关注更广泛的结构不一致, 实现取证激活的逐级细化。

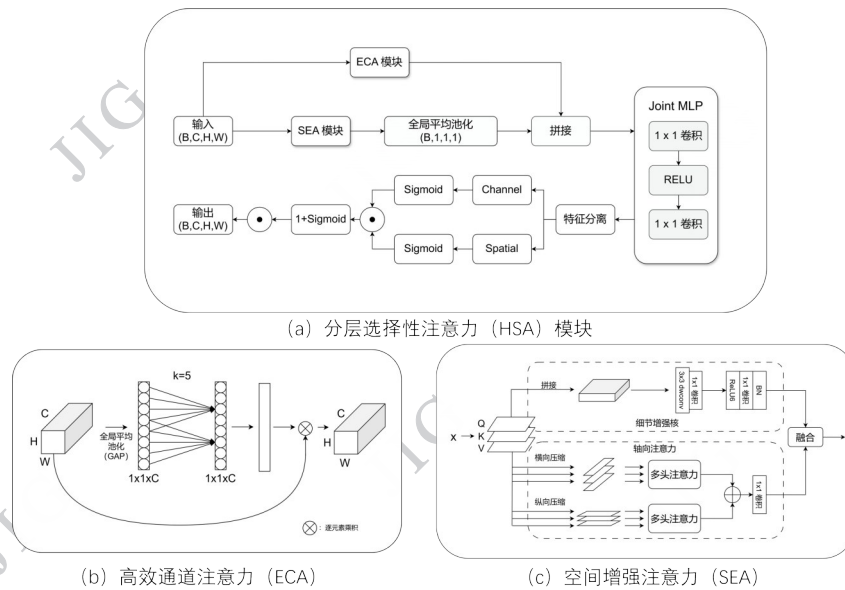


图3 所提出的分层注意力设计示意图

Fig. 3 Illustration of the proposed hierarchical attention design: (a) HSA as the main module, integrating (b) ECA and (c) SEA submodules for channel and spatial enhancement.

如图3所示, HSA 的总体流程包含三支: 高效通道注意(ECA)分支、空间增强注意(SEA)分支与特征融合适配器(Adapter), 三者协同提取跨维线索并汇聚为统一的取证表征。本文将 HSA 插入于两条

Transformer 分支的第3阶段: 早期层的操纵证据较弱不稳定, 而第4阶段语义过强易抑制残差; 第3阶段在语义与取证敏感性之间取得最佳折中, 第4.3节的消融结果亦证明其收益最大。

2.2.1 通道分支:全局取证统计的高效建模

给定输入特征 $F \in R^{C \times H \times W}$, 先经全局平均池化 (GAP) 得到通道响应:

$$z_c = \text{GAP}(F). \quad (3)$$

式(3)中, $\text{GAP}(\cdot)$ 表示全局平均池化操作, 输出 $z_c \in R$; 其中, z_c 表示第 c 个通道的全局统计响应, 用于刻画该通道对篡改取证线索的整体敏感程度。

不同于 SE 使用较重的 MLP, 这里采用核长 k 的轻量 1D 卷积以建模跨通道的局部相关性, 刻画由操纵诱发的细微统计偏移:

$$w_c = \sigma(\text{Conv1D}(z_c)). \quad (4)$$

式(4)中, $\text{Conv1D}(\cdot)$ 表示为一维卷积操作, 用于控制局部跨通道交互范围; $w_c \in R$ 表示生成的通道注意权重向量, 其中 w_c 表示第 c 个通道对应的权重系数。若式中包含 $\sigma(\cdot)$, 则其表示 Sigmoid 激活函数, 用于将权重映射到 $[0, 1]$ 区间。

随后将通道权重广播并施加到原特征上:

$$F_c = F \odot w_c. \quad (5)$$

式(5)中, \odot 表示逐元素乘法。广播后的通道权重与输入特征 F 具有相同尺寸, 输出 $F_c \in R^{C \times H \times W}$ 表示经通道重标定后的增强特征图。该过程能够突出与篡改相关的噪声统计异常和纹理分布偏移, 同时抑制冗余通道响应。

该分支强调拼接、修补以及与压缩相关操纵中常见的噪声—纹理不一致与分布异常。

2.2.2 空间分支:方向结构增强与结构异常检测

空间分支旨在捕捉边界破坏与结构不连续等局部伪迹。首先进行线性投影得到 Q, K, V :

$$Q = W_q F, \quad K = W_k F, \quad V = W_v F. \quad (6)$$

式(6)中, $F \in R^{C \times H \times W}$ 表示输入特征图, Q, K 和 V 分别表示由 F 经过线性映射得到的查询、键和值特征, 用于建模不同空间位置之间的相关性。

全局注意易平滑细粒度取证线索, 为保留细节, 引入由深度可分离卷积与逐点卷积构成的细节增强路径, 强化局部响应。为提升对轴向伪迹 (如垂直重采样纹理、水平 JPEG 块边界) 的敏感性, 分别计算行/列注意:

$$\begin{aligned} A_{\text{row}} &= \text{Softmax}(Q_{\text{row}} K_{\text{row}}^T), \\ A_{\text{col}} &= \text{Softmax}(Q_{\text{col}} K_{\text{col}}^T). \end{aligned} \quad (7)$$

式(7)中, A_{row} 和 A_{col} 分别表示沿水平方向和垂直方向计算得到的注意力图, $\text{Softmax}(\cdot)$ 表示归一化

操作。

将二者融合得到空间注意:

$$w_s = W_s ([A_{\text{row}}; A_{\text{col}}]). \quad (8)$$

并基于空间池化特征引入门控以抑制冗余激活:

$$g_s = \sigma(W_g (\text{GAP}(F))). \quad (9)$$

式(9)中, g_s 表示空间门控权重图, $\sigma(\cdot)$ 表示 Sigmoid 激活函数, 最终输出为经门控调制后的空间分支特征。

该分支显式强化对方向依赖结构异常的探测能力。

2.2.3 跨维融合:双向强化

为衔接空间与通道域, 将通道权重 w_c 与门控后的空间注意 $g_s \cdot w_s$ 拼接, 生成统一取证描述子:

$$w = \sigma(\text{MLP}([w_c; g_s \cdot w_s])). \quad (10)$$

并施加于输入特征:

$$F_{\text{HSA}} = F \odot w. \quad (11)$$

该耦合机制实现两种目标:通道引导的空间细化:以全局统计偏移指引可疑位置;空间引导的通道增强:以局部异常修正通道再分配。

相较将空间/通道视作独立顺序变换的 SE/CBAM/ECA, HSA 将二者建模为相互条件的子空间。对应地, 以

$$\hat{F} = F \odot \sigma(\phi_s(F) // \phi_c(F)) \quad (12)$$

近似联合似然 $p(S, C|F)$ 的显式跨维融合, 弥补了 SE/CBAM/ECA 缺失的结构依赖, 更契合取证伪迹中空间—统计相关性的本质。

2.3 MCFM (Multi-directional Cross-modal Fusion Module)

现代图像伪造常引入方向依赖扰动——如水平 JPEG 块边界、垂直重采样伪迹、对角修补痕迹——各向同性卷积或标准自注意难以捕捉。同时, RGB 与 SRM 两种模态在不同空间上下文下的可靠性差异显著:压缩与低纹理区域中 SRM 更稳定, 而在语义轮廓附近 RGB 更占优。因此需要一种既能建模方向先验, 又能动态平衡多模态贡献的融合机制。

为此本文设计 MCFM (如图 4), 将“方向感知深度可分离滤波”与“跨模态门控”结合, 生成统一且可自适应加权的取证表征, 以契合真实操纵中观测到的行为特征。

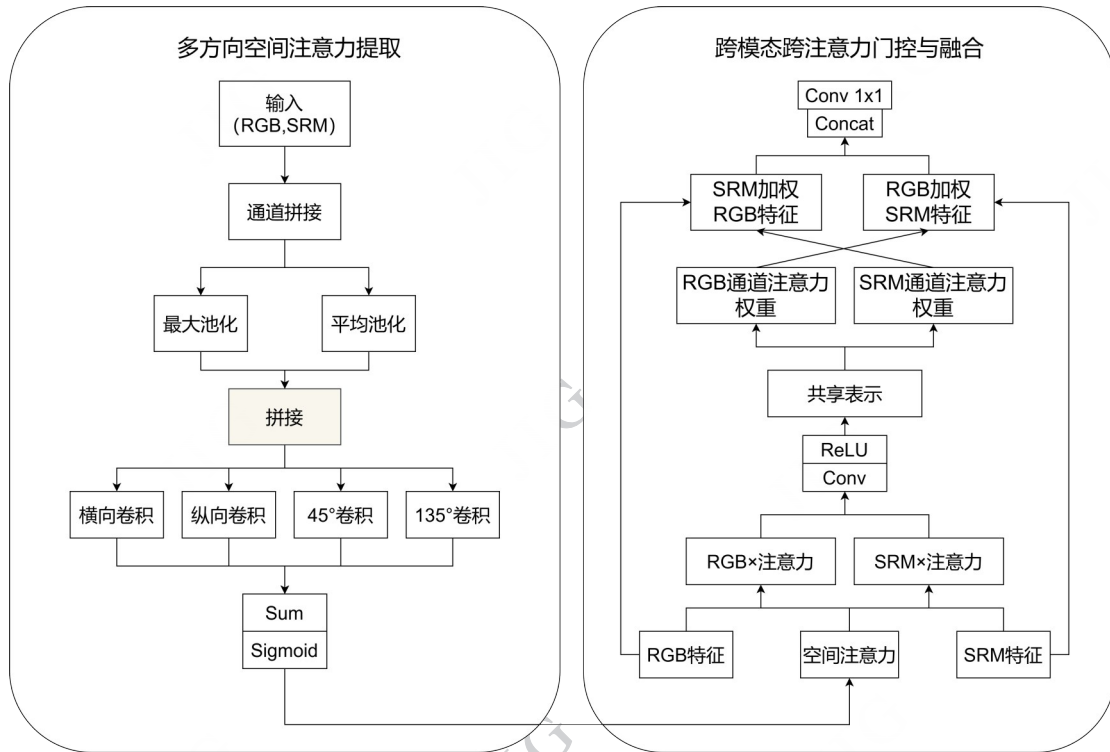


图4 所提出的MCFM模块结构

Fig. 4 Overall architecture of the proposed MCFM framework

2.3.1 多方向空间注意力提取

给定 RGB 与 SRM 特征图

$$\mathbf{F}^r, \mathbf{F}^s \in \mathbb{R}^{C \times H \times W}, \quad (13)$$

式(13)中, \mathbf{F}^r 表示 RGB 模态的输入特征图, 用于提供图像内容的结构与外观信息; \mathbf{F}^s 表示 SRM 模态的输入特征图, 用于提供与噪声残差和取证痕迹相关的统计信息。

先进行通道平均池化得到模态压缩表征并拼接:

$$\mathbf{F}^m = \text{Concat}(\text{AvgPool}(\mathbf{F}^r), \text{AvgPool}(\mathbf{F}^s)) \quad (14)$$

随后施加四组方向特异的深度可分离卷积(水平、垂直、两条对角, 核大小 3×3 或 5×5):

$$\mathbf{A}_d = \sigma(\mathbf{W}_d * \mathbf{F}^m), \quad d \in \{h, v, d1, d2\}. \quad (15)$$

将各向注意求和得到聚合的多方向注意图:

$$\mathbf{A} = \sum_d \mathbf{A}_d, \quad (16)$$

并用于增强两模态:

$$\tilde{\mathbf{F}}^r = \mathbf{F}^r \odot \mathbf{A}, \quad \tilde{\mathbf{F}}^s = \mathbf{F}^s \odot \mathbf{A}. \quad (17)$$

式(17)中, $\tilde{\mathbf{F}}^r$ 和 $\tilde{\mathbf{F}}^s$ 分别表示经多方向空间注意增强后的 RGB 特征和 SRM 特征。

2.3.2 通道降维的共享表征

为统一两模态并降低融合复杂度, 将增强后的

特征拼接后以 1×1 卷积与非线性映射压缩为共享表征:

$$\mathbf{F}^c = \phi(\text{Conv}_{1 \times 1}(\text{Concat}(\tilde{\mathbf{F}}^r, \tilde{\mathbf{F}}^s))). \quad (18)$$

2.3.3 跨模态交互式门控

MCFM 采用跨模态门控而非自门控。取证定位中, 两模态的可靠性互补: RGB 在语义轮廓附近强而在平坦或强压缩区弱; SRM 在压缩/后处理下更稳定, 但在复杂纹理或强语义处响应偏弱。形式化地, 先由共享表征估计两模态的像素级(或通道—像素级)可靠性图: $\alpha_r = \sigma(\phi_r(\mathbf{F}^c))$, $\alpha_s = \sigma(\phi_s(\mathbf{F}^c))$,

默认使用 $[0, 1]^{H \times W}$ 的像素级可靠性并在通道维广播。若采用自门控, 则以 α_r 作用于 RGB、 α_s 作用于 SRM, 这等价于假设“模态可自我诊断可靠性”, 在实践中并不成立: RGB 在语义驱动下常出现高幅但误导的激活, SRM 在纹理未篡改区域也可能产生较强残差。

为此, MCFM 采用“交叉门控”: 以 SRM 推断的可靠性调制 RGB, 反之亦然:

$$\hat{\mathbf{F}}^r = \tilde{\mathbf{F}}^r \odot \alpha_s, \quad \hat{\mathbf{F}}^s = \tilde{\mathbf{F}}^s \odot \alpha_r. \quad (19)$$

该设计等价于施加“相互一致性约束”: 仅当一方证据得到互补模态“背书”时才放大, 从而抑制语义诱发的误报, 并在平坦或强压缩区域(RGB 线索

偏弱而 SRM 仍可靠)稳定检测。可将其视为条件加权的近似:

$$\begin{aligned} w_r &\propto p(\text{RGB reliable} | \text{SRM statistics}), \\ w_s &\propto p(\text{SRM reliable} | \text{RGB structure}). \end{aligned} \quad (20)$$

式(20)中的条件加权表示在共享表征约束下对另一模态特征可靠性的近似估计,其本质是特征层面的条件调制关系,而非严格意义上的统计条件概率。

2.3.4 最终跨模态融合

将重加权后的特征拼接并投影得到最终融合表示:

$$\mathbf{F}^{\text{MCFM}} = \mathbf{W}_f \text{Concat}(\hat{\mathbf{F}}^r, \hat{\mathbf{F}}^s), \quad (21)$$

其中 \mathbf{W}_f 为 1×1 卷积。MCFM 通过“多方向深度可分离滤波(方向先验)+跨模态交互门控(相互佐证)”将显式取证先验体现在融合过程中,并以 SRM 残差为统计依据落地实现。

3 实验

3.1 数据集

为了训练 DA-CMTL,本文基于公开数据集构建了约 62,000 张主要由合成拼接操作生成的训练样本(Lin 等,2014)。虽然训练样例以拼接为主,但这种构造方式并不会导致模型对单一操纵类型的过拟合,其原因在于:本文关注的弱残差证据——包括方向扰动、噪声不一致性与局部结构偏移——在不同篡改类型(如 Copy-Move、Inpainting、Resampling、GAN-based Manipulation)下具有高度一致的物理特征,即都表现为语义无关、方向相关且易受压缩退化影响的稀疏扰动。因此,拼接操作能够提供最丰富且最可控的统计扰动形式,使模型在训练阶段充分学习上述通用弱残差特性,而不会依赖特定语义或篡改模式。

此外,本文在构造训练集时引入了随机仿射变换、内容遮挡、区域形状变化以及不同尺度融合等多种扰动,这些变换能够模拟复制、修补、局部重采样等篡改所造成的边界不连续与噪声突变,使得训练数据在统计特征上覆盖多类型篡改的关键伪迹。为了验证泛化能力,本文在测试阶段引入了未在训练中出现的操纵类型(如 IMD 和 Columbia 中的混合型篡改),并进一步在 OSN 传输链路中测试模型在真

实分布下的跨类型鲁棒性,实验结果表明模型在多操纵场景中仍保持稳定性能。

综上,本研究采用拼接为主的训练构造并不会削弱模型的普适性;相反,它使 DA-CMTL 能够更集中地学习篡改过程中最本质的弱残差扰动规律,从而在无需大量操纵特定样例的情况下实现跨类型泛化。

为全面评估方法的泛化性与鲁棒性,本文在五个常用的基准数据集上进行实验:CASIA(Jing 等,2013;Wei 等,2010)、Columbia(Hsu 等,2006)、DSO(Vincent 等,2012)、NIST16(Pawel 等,2015)与 IMD(Markos 等,2017)。

在 SRM 分支中,本文采用固定的高通滤波策略:以深度可分离 5×5 卷积实现,输入通道为 3,输出通道为 9。具体而言,使用三组 SRM 5×5 核(归一化因子为 $\{4, 12, 2\}$),并在 RGB 通道上复制,形成 9 个滤波器;其权重在训练中保持冻结。经滤波得到的 9 通道残差图送入 MiT-B2 骨干(mit_b2_9),与 RGB 分支骨干(mit_b2)并行构成 DA-CMTL 的双分支结构。

优化方面,采用 AdamW,初始学习率 5×10^{-4} ,在 100 个 epoch 内以余弦退火下降至 5×10^{-6} ;batch size 为 8,权重衰减 10^{-2} ,动量参数 $\beta_1=0.9$ 。监督采用 Dice 损失与 Focal 损失的组合,对两类像素等权监督。训练/验证划分由训练根目录下 ImageSets/Segmentation 的 train.txt 与 val.txt 指定;所有图像统一缩放至 512×512 ,并按 4.1 节所述进行数据增强。

3.2 对比实验

本文将 DA-CMTL 与多种图像篡改定位方法进行系统对比,包括 ManTra-Net、SPAN、CR-CNN(Yang 等,2019)、GSR-Net、IF-OSN(Wu 等,2022)、MVSS-Net、CAT-Net、OSN(Wu, Liu 等,2022)、EITLNet、Trufor 与 SparseViT。为保证对比实验的公平性与可重复性,本文对所有对比方法均采用与本文方法 DA-CMTL 完全一致的训练数据与训练设置进行重新训练与测试。具体而言,所有方法均使用本文构建的训练集进行训练,输入分辨率与数据预处理策略保持一致,并在相同的优化器配置、训练轮数与批大小等超参数设置下完成训练与推理,从而避免由于训练数据规模、数据分布差异或训练策略不一致导致的性能偏差。表 2 汇报了五个数据集(CASIA-v1、

Columbia、DSO、NIST16、IMD) 的 F1 与 IoU 对比结果。

DA-CMTL 在所有基准上均取得一致性提升, 尤其在 NIST16 与 IMD 上优势显著——这些数据集中常含强方向性的残差伪迹, 体现了 MCFM 显式建模

水平/垂直/对角线索的价值。即便在强压缩与重采样条件下, DA-CMTL 仍保持稳定; 这表明跨模态门控能够在 RGB 外观与 SRM 残差之间选择更可靠的模态。更高的 IoU 指标进一步说明, 本方法在多样操纵场景下提供了更精确、稳健的边界定位。

表 2 不同篡改定位算法在各数据集上的性能比较(F1 和 IoU)。

Table 2 Performance comparison of different image tampering localization methods across various datasets (F1 and IoU).

Methods	CASIA-v1		Columbia		DSO		NIST		IMD		Average	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
ManTra-Net	0.130	0.086	0.357	0.258	0.332	0.243	0.088	0.054	0.183	0.124	0.218	0.153
SPAN	0.112	0.061	0.235	0.14	0.2	0.11	0.559	0.36	0.17	0.098	0.255	0.154
CR-CNN	0.405	0.290	0.436	0.32	0.31	0.21	0.238	0.14	0.262	0.17	0.330	0.226
GSR-Net	0.387	0.272	0.613	0.43	0.28	0.18	0.283	0.17	0.243	0.16	0.361	0.242
IF-OSN	0.553	0.391	0.753	0.61	0.35	0.24	0.33	0.2	0.41	0.28	0.479	0.344
MVSS-Net	0.509	0.465	0.684	0.596	0.271	0.188	0.294	0.24	0.26	0.2	0.404	0.338
CAT-Net	0.136	0.080	0.555	0.42	0.2	0.12	0.179	0.1	0.054	0.03	0.225	0.15
OSN	0.509	0.465	0.707	0.608	0.336	0.308	0.312	0.255	0.411	0.312	0.455	0.390
EITLNet	0.519	0.476	0.778	0.726	0.341	0.233	0.301	0.237	0.418	0.334	0.471	0.401
Trufor	0.537	0.494	0.760	0.701	0.261	0.210	0.322	0.263	0.225	0.193	0.421	0.372
SparseViT	0.543	0.501	0.800	0.765	0.275	0.224	0.346	0.289	0.238	0.206	0.440	0.397
DA-CMTL	0.532	0.491	0.798	0.753	0.349	0.243	0.313	0.244	0.43	0.343	0.484	0.414

注: 每个测试集上的最佳结果以加粗字体标出

3.3 消融实验

3.3.1 整体性消融

为更深入理解各模块贡献, 本文在双分支基线上逐步加入 HSA 与 MCFM 进行消融。除报告数值

增益外, 还分析二者的功能差异。表 3 显示: HSA 主要强化“跨维(空间—通道)耦合”, 对 Columbia、CASIA-v1 等以边界伪迹为主的数据集提升更为明显。

表 3 所提出方法的消融实验分析(F1 和 IoU)。

Table 3 Ablation study of the proposed method (F1 and IoU).

Methods	CASIA-v1		Columbia		DSO		NIST		IMD		Average	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
Baseline	0.519	0.4755	0.7777	0.7256	0.3407	0.2332	0.3006	0.237	0.418	0.334	0.4712	0.4011
Baseline+HSA	0.5304	0.4882	0.7978	0.7573	0.3398	0.2332	0.3034	0.2323	0.4181	0.3346	0.4779	0.4091
Baseline+MCFM	0.5289	0.4897	0.7792	0.7362	0.3569	0.25	0.3105	0.2438	0.4251	0.3412	0.4801	0.4122
Baseline+HSA+MCFM	0.532	0.4909	0.7979	0.7525	0.3487	0.2427	0.3128	0.2435	0.4296	0.3426	0.4842	0.4144

注: 每个测试集上的最佳结果以加粗字体标出

相较之下, MCFM 带来更大增益, 尤其在具备强方向依赖痕迹的数据集(如 NIST16、DSO)上; 其多方向滤波与跨模态门控可按局部取证上下文在

RGB 与 SRM 线索间自适应权衡。二者同时启用时, 网络受益于 HSA 的联合强化与 MCFM 的上下文自适应多模态融合, 获得整体最优表现。

从消融结果还可验证“空间—通道联合建模”的必要性:仅启用 HSA (Baseline+HSA) 时,平均 F1 与 IoU 均较基线稳步提升,且在边界不一致占主导的数据集上更显著。这说明 HSA 并非通用注意力模块的简单替代。相反,提升集中于“边缘与通道残差统计同时显著”的场景,与本文将 HSA 视作 $p(S, C | F)$ 的近似相吻合。此外,仅加 MCFM (Baseline+MCFM) 时, NIST16 与 DSO 上增益最大,验证了条件式跨模态门控能抑制模态专属噪声并强调 RGB 与 SRM 一致的区域。

3.3.2 HSA 合理性消融

为验证所提出 HSA 并非对经典注意力模块的简单叠加或等价替代,本文在相同双分支基线框架下进行“模块替换式”消融:将 HSA 插入位置与 Baseline+HSA 保持一致(均位于第 3 阶段之后),并分别使用经典通道/空间注意力模块 SE、ECA 与 CBAM 对该位置进行替换。除注意力结构外,所有对比版本均采用相同训练集、相同输入分辨率与相同优化参数进行训练与测试,以确保比较公平。实验结果如表 4 所示。

从整体结果来看,与不使用注意力增强的 Baseline 相比,引入任意注意力模块均可带来一定增益,说明在篡改定位任务中“显著性增强”是必要的;但不同模块的增益幅度存在明显差异。具体而言, HSA 在五个数据集上的平均性能最优,达到 Average F1=0.4842、Average IoU=0.4144,优于 SE (0.4752/0.4077)、ECA (0.4803/0.4116) 与 CBAM (0.4839/0.4134)。这一结果表明:仅进行通道重标

定(SE/ECA)或简单的空间-通道串行增强(CBAM)仍不足以稳定刻画“结构异常+噪声统计偏移”共同构成的弱残差证据,而 HSA 的分层空间-通道耦合机制能够更有效地抑制语义相关干扰并增强与篡改一致的取证激活。

分数据集分析进一步揭示了 HSA 的优势来源:在边界伪迹与噪声统计偏移更为典型的 Columbia 数据集上, HSA 达到 F1=0.7979、IoU=0.7525,相比 SE (0.7769/0.7375) 与 ECA (0.7879/0.7437) 均有提升;在真实退化更强、弱残差更稀疏的 NIST16 上, HSA 获得 F1=0.3128、IoU=0.2435,同样优于 SE (0.3031/0.2357)、ECA (0.3084/0.2404) 与 CBAM (0.3094/0.2351)。在 IMD 上, HSA 也取得最高 F1=0.4296、IoU=0.3426,体现出其在复杂纹理与混合篡改场景下的更强泛化能力。需要说明的是,在以方向残差扰动占比较高的 DSO 上, ECA/CBAM 的 F1 与 IoU 略高于 HSA,这与 DSO 样本中方向性伪迹更强、对后续方向建模与跨模态融合(MCFM)更敏感的特点一致;但从五个数据集的综合表现看, HSA 仍提供了更稳定、更一致的整体收益。

由此本文可以证明: HSA 并非传统注意力模块 (SE/ECA/CBAM) 的简单组合,其通过分层地联合建模空间结构异常与通道噪声统计偏移,能够更有效地恢复深层语义抑制下的弱残差信号,并在真实退化与跨数据集评测中带来更稳健的提升,为后续 MCFM 的跨模态自适应融合提供更可靠的取证表征基础。

表 4 经典注意力模块替换式消融实验 (SE/ECA/CBAM vs HSA, F1 与 IoU)

Table 4 Replacement ablation of attention modules (SE/ECA/CBAM vs proposed HSA) in the same insertion position (F1 and IoU).

	CASIA-v1		Columbia		DSO		NIST		IMD		Average	
	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU	F1	IoU
Baseline	0.519	0.4755	0.7777	0.7256	0.3407	0.2332	0.3006	0.237	0.418	0.334	0.4712	0.40106
+SE	0.5332	0.4914	0.7769	0.7375	0.3434	0.2383	0.3031	0.2357	0.4196	0.3357	0.47524	0.40772
+ECA	0.5156	0.4757	0.7879	0.7437	0.3608	0.2556	0.3084	0.2404	0.4289	0.3428	0.48032	0.41164
+CBAM	0.5302	0.4899	0.7953	0.7503	0.3606	0.2524	0.3094	0.2351	0.4241	0.3392	0.48392	0.41338
+HSA	0.532	0.4909	0.7979	0.7525	0.3487	0.2427	0.3128	0.2435	0.4296	0.3426	0.4842	0.41444

3.3.3 HSA 与经典注意力的对比热力图

为更直观地分析不同注意力机制对篡改区域表

征的影响,本节对注意力增强后的特征响应进行可视化展示。具体而言,在与前述实验一致的设置下,

仅替换注意力模块类型,其余网络结构、插入位置、输入尺度与训练策略均保持一致。对于同一测试图像,提取注意力作用后的特征响应,并采用统一的归一化方式将其映射为热力图,随后以相同的颜色映射与透明度叠加到原图上,以保证不同方法间具有可比性。

图5给出了不同注意力模块的响应可视化结果。从左到右分别为:输入图像、真值掩码、SE、ECA、CBAM 与本文方法 HSA 的注意力响应热力图。所有热力图采用相同归一化方式与可视化参数生成。从整体上看,SE 与 ECA 作为典型的通道注意力方法,均通过对通道维进行重标定来增强显著线索,但其空间选择性相对有限,在复杂纹理背景下仍容易出现较为弥散的激活。其中,ECA 由于采用局部通道交互,在一定程度上提升了有效通道的筛

选能力,但对细粒度篡改边界的刻画仍不够稳定。CBAM 进一步引入空间注意力后,对目标区域的响应更加突出,但仍可观察到部分背景结构被同时激活,且边界处高响应区域的连续性不足。相比之下,本文提出的 HSA 在篡改区域尤其是其边界附近产生了更加集中且连续的高响应,同时对非篡改背景的抑制更明显,能够更有效地突出与篡改一致的取证线索。上述现象表明,HSA 的分层空间-通道耦合机制有助于在强语义与复杂纹理干扰下强化弱残差信号,从而提升篡改定位的精细程度与鲁棒性。

上述注意力响应差异表明:在进入跨模态融合之前,对深层特征中的弱残差进行结构化强化是必要的;因此,下一节将进一步从结构拆解的角度分析 MCFM 在方向建模、跨模态交互与自适应融合中的关键组成与贡献来源。

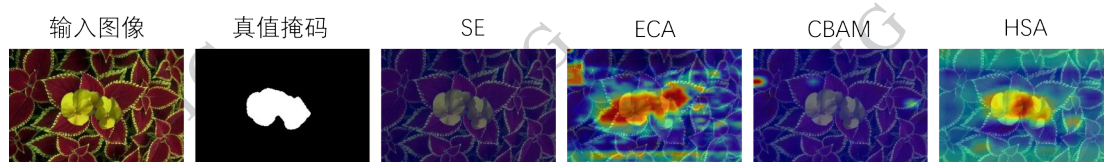


图5 不同注意力模块的响应可视化对比

Fig. 5 Visualization comparison of attention responses for different modules

3.3.4 MCFM 结构性消融

在完成注意力增强机制的定性分析后,本节进一步对跨模态融合模块 MCFM 进行结构性消融,以明确其性能增益来自何种内部机制。具体而言,在保持网络骨干、训练数据、输入分辨率及优化参数一致的前提下,构建多种 MCFM 变体,仅对其内部融合方式与交互机制进行调整,包括:简单特征拼接(a1)、等权融合(a2)、仅保留方向分支(dir_only)、移除自适应门控(no_gate)、仅采用跨模态交互(xattn_only),以及完整的 MCFM 结构(full)。实验结果如表 5 所示。

从整体结果来看,完整的 MCFM 在所有数据集及平均指标上均取得最优性能,平均 F1 与 IoU 分别达到 0.4842 和 0.4144,表明方向建模、跨模态交互与自适应融合三者的协同作用对于提升篡改定位性能是必要的。相较之下,采用简单特征拼接(a1)或等权融合(a2)的模型在多数数据集上性能明显下降,说明仅依赖静态融合方式难以有效建模 RGB 与 SRM 特征之间的互补关系。

进一步比较不同结构变体可以发现,仅保留方

向分支的 dir_only 相较于 a1 与 a2 在多数数据集上取得一定提升,表明方向先验对于刻画篡改边界与重采样伪迹具有积极作用;而仅采用跨模态交互的 xattn_only 在整体性能上进一步提升,尤其在 DSO 与 NIST 数据集上表现更为稳定,说明跨模态信息对齐有助于抑制冗余特征并增强弱残差线索。然而,上述两种变体在缺少自适应融合机制时仍存在性能上限。

当移除自适应门控机制(no_gate)时,模型性能在多个数据集上出现明显退化,尤其在 NIST 与 IMD 等真实退化更强的场景中更为显著。这一现象表明,在复杂内容与多样退化条件下,固定或等权的模态融合策略难以兼顾语义一致性与取证敏感性。相比之下,完整 MCFM 通过引入自适应门控机制,能够根据局部内容动态调节 RGB 与 SRM 特征的贡献,从而在不同数据集与场景中获得更稳健的性能提升。

由此本文发现方向先验、跨模态交互以及自适应融合机制在篡改定位任务中具有显著互补性,三者缺一不可。完整的 MCFM 能够更充分地挖掘

RGB 语义信息与 SRM 噪声统计特征之间的协同关系,为后续精细化篡改区域定位提供更加可靠的特征支撑。

为进一步直观分析 MCFM 在不同区域对多模态特征的调节行为,下一节将对其跨模态权重分布进行可视化展示。

表5 MCFM 结构性消融实验:方向分支、跨模态交互与自适应门控的贡献分析(F1 与 IoU)

Table 5 Structural ablation of MCFM: contributions of directional modeling, cross-modal interaction, and adaptive gating (F1 and IoU).

	casia1		columbia		DSO		NIST		IMD		Average	
	f1	iou	f1	iou	F1	iou	f1	iou	f1	iou	f1	iou
a1	0.5222	0.4821	0.7851	0.7384	0.3342	0.2396	0.3004	0.2348	0.4172	0.3339	0.4718	0.4058
a2	0.5295	0.4882	0.7905	0.7478	0.3345	0.2397	0.3048	0.24	0.4216	0.3391	0.4762	0.411
dir_only	0.5202	0.4806	0.7853	0.7343	0.3458	0.2434	0.3037	0.2342	0.4205	0.3333	0.4751	0.4052
no_gate	0.5317	0.4904	0.7985	0.7574	0.3053	0.2084	0.2984	0.2335	0.4152	0.3326	0.4698	0.4045
xattn_only	0.5318	0.4905	0.7909	0.7466	0.3475	0.244	0.31	0.2418	0.4268	0.3409	0.4814	0.4128
full	0.532	0.4909	0.7979	0.7525	0.3487	0.2427	0.3128	0.2435	0.4296	0.3426	0.4842	0.4144

3.3.5 MCFM 跨模态自适应可视化

为直观分析 MCFM 在跨模态融合过程中的自适应调节行为,本节对其门控权重进行可视化展示。具体而言,在保持模型结构与推理流程一致的前提下,提取 MCFM 输出的 RGB 与 SRM 分支门控权重,并采用统一的归一化方式将权重映射为热力图;随后以相同的颜色映射与透明度参数叠加到原图上,从而反映不同空间位置上两种模态特征的相对贡献。

图六展示了一个典型样例的可视化结果。从左到右分别为:输入图像、真值掩码、模型预测结果、RGB 分支门控权重叠加图与 SRM 分支门控权重叠加图。所有权重热力图采用相同归一化与可视化参

数生成。可以观察到,MCFM 的权重分布呈现出明显的空间自适应特性:在篡改区域及其边界附近,SRM 分支的权重响应更为显著,说明模型更倾向于利用噪声统计与残差相关线索以突出弱篡改痕迹;而在大面积语义一致的背景区域,RGB 分支权重占优,有助于维持语义结构的稳定表达并抑制由残差特征带来的噪声干扰。该现象表明,MCFM 并非对两种模态进行固定或等权融合,而是能够根据局部内容与取证可信度动态调整融合比例,从而在复杂纹理与真实退化场景下获得更稳健的定位效果。

上述可视化结果从机制层面验证了 MCFM 的跨模态自适应融合特性,为后续在不同退化条件下的鲁棒性评估提供了直观依据。

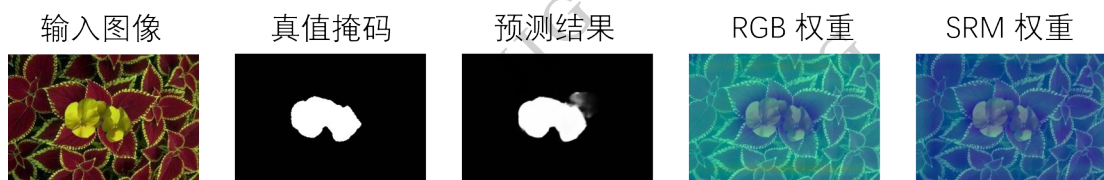


图6 MCFM 跨模态自适应门控权重可视化

Fig. 6 Visualization of cross-modal adaptive gating weights in MCFM

3.4 鲁棒性测试

社交网络中的有损处理对篡改定位挑战极大。本文遵循 OSN 协议,将四个标准数据集分别上传至 Facebook、微博、微信与 WhatsApp,并下载回本地进行测试。表 6 给出了不同方法在各平台的 F1 对

比。可见 DA-CMTL 在所有数据集与平台上均达到最佳或接近最佳的表现。

更具体地,Columbia 上 DA-CMTL 的平均 F1 达到 0.80,较 EITLNet 与 OSN 分别提升约 6.0% 与 11.8%。在 CASIA-v1 与 DSO 上,平均 F1 分别为

0.51 与 0.46, 较最优基线(EITLNet)保持 3 - 5% 的优势。即便在最具挑战性的 NIST16 上——各方法均明显降幅——DA-CMTL 仍以 0.33 - 0.35 的平均 F1 领先约 5.0%。

跨平台比较表明, 不同社交网络之间的性能差异通常较小(多数在 ± 0.02 以内), 而 DA-CMTL 始终保持绝对优势。这些结果强有力地验证了方法在面对社交平台常见的有损压缩与转码操作时的鲁棒性与泛化能力。

除了整体性能比较外, 本文进一步从“方向性破坏”的角度分析不同社交平台压缩策略对篡改伪迹的影响。现有研究表明, 微信、WhatsApp 等平台在图像压缩过程中往往采用强 JPEG 重新编码与多轮下采样, 这类处理对水平与垂直方向的周期纹理和块边界结构会造成显著破坏; 而 Twitter、Facebook 等平台使用更复杂的多步缩放与再编码策略, 可能同

时弱化对角方向的细粒度噪声模式。因此, OSN 压缩不仅降低整体图像质量, 更会引起方向性伪迹的非均匀损伤, 这对于依赖方向扰动或噪声残差的篡改定位方法影响尤为显著。

在这种背景下, DA-CMTL 在各平台下仍保持显著优势, 特别是在微信与 WhatsApp 等方向性压缩最明显的平台上, 相对于其他方法的提升幅度更为突出。具体而言, 多方向滤波的 MCFM 能够在 JPEG 重编码导致的水平/垂直方向损伤下仍保留对角扰动及跨方向残差的响应; 而跨模态门控机制则在方向特征被部分破坏时自动提高噪声残差的权重, 使模型在方向伪迹不均匀衰减的复杂条件下仍能维持稳定定位能力。这表明 DA-CMTL 不仅在整体质量退化下具有鲁棒性, 更能够适应 OSN 传输链路中典型的方向扰动差异性, 为其在真实场景应用中提供了关键支撑。

表 6 针对社交网络的鲁棒性评估结果(F1)。数据集被上传至 Facebook (Fb)、Weibo (Wb)、WeChat (Wc) 和 WhatsApp (Wa)。

Table 6 Robustness evaluation results on social networks (F1). The datasets were uploaded to Facebook (Fb), Weibo (Wb), WeChat (Wc), and WhatsApp (Wa).

Methods	Columbia				CASIA-v1				DSO				NIST16				Average			
	Fb	Wb	Wc	Wa	Fb	Wb	Wc	Wa	Fb	Wb	Wc	Wa	Fb	Wb	Wc	Wa	Fb	Wb	Wc	Wa
MVSS-net	0.691	0.689	0.69	0.685	0.387	0.409	0.248	0.359	0.277	0.258	0.214	0.181	0.264	0.251	0.212	0.165	0.309	0.306	0.225	0.235
OSN	0.714	0.724	0.727	0.727	0.462	0.466	0.405	0.478	0.447	0.370	0.366	0.341	0.329	0.294	0.286	0.313	0.413	0.377	0.352	0.377
EITLNet	0.755	0.75	0.752	0.758	0.47	0.455	0.46	0.465	0.41	0.405	0.412	0.408	0.34	0.338	0.342	0.339	0.407	0.399	0.405	0.404
DA-CMTL	0.8	0.799	0.799	0.8	0.505	0.52	0.515	0.512	0.46	0.455	0.462	0.458	0.355	0.353	0.356	0.354	0.440	0.443	0.444	0.441

注: 每个测试集上的最佳结果以加粗字体标出

3.5 模型规模比较

如表 7 所示, DA-CMTL 在保持较高检测性能的同时, 具备出色的模型紧凑性与计算效率。相较于性能较高的 TruFor, 在相同训练分辨率(512×512)下, 参数量从 68.7M 降至 60.01M, 计算量从 236.5G 大幅降至 48.55G, 计算开销减少超过 75%。

与同分辨率的其他方法(MVSS: 147.0M/167.0G; CAT-Net: 114.0M/134.0G; EITLNet: 75.18M/62.32G)相比, DA-CMTL 在模型复杂度与推理效率上仍具明显优势。即便与较低输入分辨率(256×256)的轻量级模型(ManTraNet: 3.9M/274.0G; PSCC-Net: 3.7M/45.7G)相比, DA-CMTL 也在计算效率与表征能力之间取得更优平衡。

表 7 与 SOTA 模型在参数和 FLOPs 上进行比较

Table 7 Comparison with state-of-the-art models in terms of parameters and FLOPs.

Method	Size	Parameter	FLOPs
ManTraNet	256x256	3.9M	274.0G
PSCC-Net	256x256	3.7M	45.7G
MVSS	512x512	147.0M	167.0G
CAT-Net	512x512	114.0M	134.0G
Trufor	512x512	68.7M	236.5G
EITLNet	512x512	75.18M	62.32G
DA-CMTL	512x512	60.01M	48.55G

总体而言,在公平的跨数据集评测协议下,DA-CMTL 达到或优于当前最新水平。图7的质化对比进一步说明:无论操纵是否发生在对象层面,DA-CMTL 都能有效利用与语义无关的非语义特征,准确识别篡改区域,同时避免由语义相关性引发的误

检。进一步观察显示,部分现有模型在最后三组样例中出现明显的语义相关误检;相比之下,DA-CMTL 通过其独特的注意力机制有效抑制语义干扰,更专注于图像完整性相关的关键非语义线索。

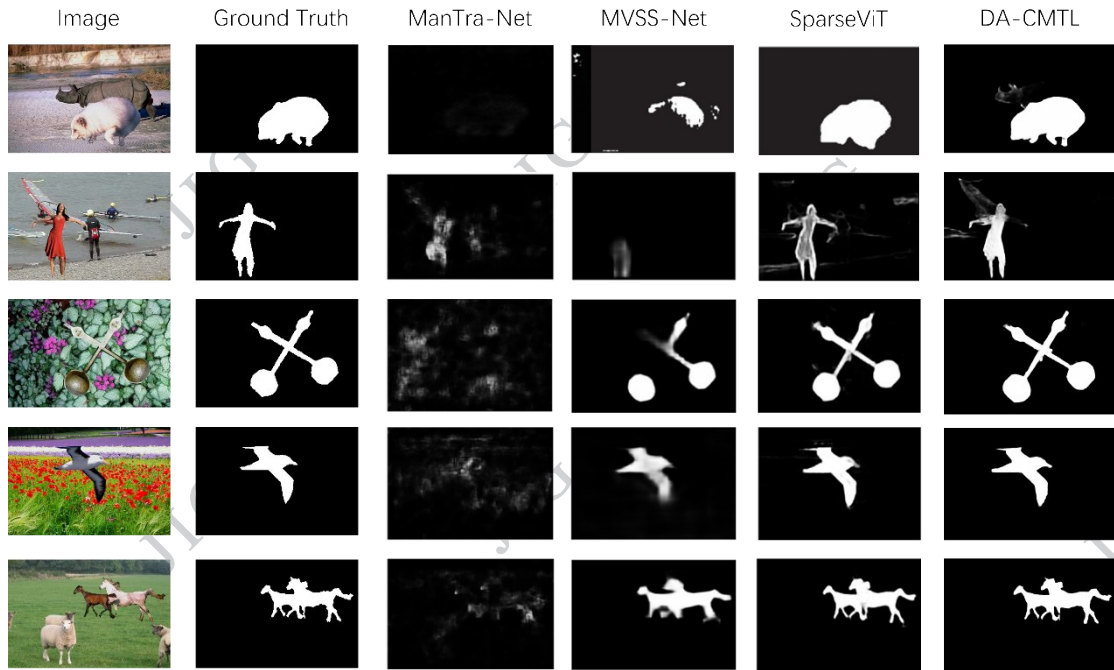


图7 不同方法的篡改定位质化对比。部分现有方法在最后三组样例中出现语义相关误检。

Fig. 7 Qualitative comparison of tampering localization results produced by different methods. Some existing methods exhibit semantic-related false positives in the last three examples.

4 结 论

本文围绕图像篡改定位中弱残差稀疏、方向依赖性强、跨模态不一致与深层平滑抑制等关键科学问题,提出了 DA-CMTL 框架,实现了方向敏感性、模态自适应性与空间-通道耦合推理的统一建模。该框架从篡改伪迹的物理机制出发,系统性弥补了现有方法在方向建模缺失、跨模态融合静态化及弱残差在深层表征中易被语义平滑弱化等三大核心缺口。具体而言,HSA 通过联合建模空间结构与通道噪声,使弱残差在高语义特征中仍保持可辨识性;MCFM 则通过多方向滤波与跨模态门控显式捕获各类篡改操作中的方向扰动,并在局部退化状态变化下动态调节 RGB 与 SRM 残差的融合策略,从而形成基于物理一致性的方向感知跨模态推理机制。

在五个公开基准与多类真实社交平台压缩链路

上,DA-CMTL 在极弱操纵、强压缩与复杂纹理等典型困难场景中依然保持领先性能。实验结果揭示出具有普适意义的两项关键规律:(1)真实退化条件下,方向扰动与噪声残差呈现显著互补性,是弱残差增强的重要统计基础;(2)高层语义平滑会显著削弱弱残差信号,而空间-通道耦合机制能够有效抑制这一现象。这些规律不仅为 DA-CMTL 的模块设计提供了验证,也揭示了弱残差驱动的跨模态推理在真实退化环境中应遵循的基本原则,对该领域未来模型设计具有理论指导意义。

尽管 DA-CMTL 在多类场景下表现稳健,但其性能仍可能受限于极弱篡改下 SRM 残差响应抑制与强压缩链路导致的方向特征退化。此外,扩散模型伪造在噪声结构、方向分布与跨尺度一致性方面与传统操纵显著不同,其统计规律仍未被充分理解。未来研究可从以下三个方向展开:(1)构建具备内容自适应能力的噪声与方向先验,使跨模态增强机制

能够根据局部复杂度与退化强度动态调节;(2)结合真实互联网压缩链路的统计特性,建立抗退化训练范式与稳健表征机制,提高模型在复杂传输环境下的可靠性;(3)系统建模扩散伪造的噪声演化规律,探索基于扩散逆过程或噪声轨迹一致性的全新取证先验,并构建覆盖扩散伪造的跨分布检验基准,以推动该方向形成具有可验证性的理论体系。

总体而言,DA-CMTL 验证并确立了弱残差取证中三类关键能力——方向敏感性、模态自适应性与空间-通道耦合推理——必须被协同建模的核心规律,提出了一种结构化、物理一致且可扩展的跨模态取证范式。该研究不仅具有司法鉴定、媒体可信度审查与内容安全治理等实际应用价值,也为构建基于物理一致性的可信视觉内容认证体系提供了方法论上的基础性支撑。

参考文献 (References)

- Avraami O, Fried O and Lischinski D. 2023. Blended latent diffusion [J]. *ACM Transactions on Graphics*, 42(4): 1-11 [DOI: 10.1145/3592450]
- Bayar B and Stamm M C. 2016. A deep learning approach to universal image manipulation detection using a new convolutional layer [C]// *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*. New York, NY, USA: ACM: 5-10 [DOI: 10.1145/2909827.2930786]
- Bianchi T and Piva A. 2012. Image forgery localization via block-grained analysis of JPEG artifacts [J]. *IEEE Transactions on Information Forensics and Security*, 7(3): 1003-1017 [DOI: 10.1109/TIFS.2012.2187516]
- Christlein V, Riess C, Jordan J, Riess C and Angelopoulou E. 2012. An evaluation of popular copy-move forgery detection approaches [J]. *IEEE Transactions on Information Forensics and Security*, 7(6): 1841-1854 [DOI: 10.1109/TIFS.2012.2218597]
- Dong C, Chen X, Hu R, Cao J and Li X. 2022. MVSS-Net: multi-view multi-scale supervised networks for image manipulation detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3): 3539-3553 [DOI: 10.1109/TPAMI.2022.3180556]
- Dong J, Wang W and Tan T. 2013. CASIA image tampering detection evaluation database [C]// *2013 IEEE China Summit and International Conference on Signal and Information Processing*. Piscataway, NJ, USA: IEEE: 422-426 [DOI: 10.1109/ChinaSIP.2013.6625374]
- Fridrich J and Kodovsky J. 2012. Rich models for steganalysis of digital images [J]. *IEEE Transactions on Information Forensics and Security*, 7(3): 868-882 [DOI: 10.1109/TIFS.2012.2190402]
- Gafni O and Wolf L. 2020. Wish you were here: context-aware human generation [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE: 7837-7846 [DOI: 10.1109/CVPR42600.2020.00786]
- Guillaro F, Cozzolino D, Sud A, Dufour N and Verdoliva L. 2023. Tru-For: leveraging all-round clues for trustworthy image forgery detection and localization [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE: 20606-20615 [DOI: 10.1109/CVPR52729.2023.01974]
- Guo K, Zhu H and Cao G. 2024. Effective image tampering localization via enhanced transformer and co-attention fusion [C]// *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing*. Piscataway, NJ, USA: IEEE: 4895-4899 [DOI: 10.1109/ICASSP48485.2024.10446332]
- Hsu Y F and Chang S F. 2006. Detecting image splicing using geometry invariants and camera characteristics consistency [C]// *2006 IEEE International Conference on Multimedia and Expo*. Piscataway, NJ, USA: IEEE: 549-552 [DOI: 10.1109/ICME.2006.262447]
- Hu X, Zhang Z, Jiang Z, Chaudhuri S, Yang Z and Nevatia R. 2020. SPAN: spatial pyramid attention network for image manipulation localization [C]// *European Conference on Computer Vision*. Cham, Switzerland: Springer: 312-328 [DOI: 10.1007/978-3-030-58589-1_19]
- İmamoğlu M B, Ulutaş G and Ulutaş M. 2013. Detection of copy-move forgery using Krawtchouk moment [C]// *2013 8th International Conference on Electrical and Electronics Engineering*. Piscataway, NJ, USA: IEEE: 311-314 [DOI: 10.1109/ELECO.2013.6713851]
- Jiang X Y and Liu C X. 2021. Image splicing forgery detection network guided by edge and region inconsistency [J]. *Journal of Image and Graphics*, 26(10): 2411-2420 [蒋小玉, 刘春晓. 2021. 边缘与区域不一致性引导下的图像拼接篡改检测网络 [J]. *中国图象图形学报*, 26(10): 2411-2420 [DOI: 10.11834/jig.200298]
- Johnson M K and Farid H. 2007. Exposing digital forgeries through specular highlights on the eye [C]// *International Workshop on Information Hiding*. Berlin, Germany: Springer: 311-325 [DOI: 10.1007/978-3-540-77370-2_21]
- Kanagavalli N and Latha L. 2017. A survey of copy-move image forgery detection techniques [C]// *2017 International Conference on Inventive Systems and Control*. Piscataway, NJ, USA: IEEE: 1-6 [DOI: 10.1109/ICISC.2017.8068703]
- Korus P and Huang J. 2016. Multi-scale analysis strategies in PRNU-based tampering localization [J]. *IEEE Transactions on Information Forensics and Security*, 12(4): 809-824 [DOI: 10.1109/TIFS.2016.2636089]
- Kwon M J, Yu I J, Nam S H and Lee H K. 2021. CAT-Net: compression artifact tracing network for detection and localization of image splicing [C]// *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. Piscataway, NJ, USA: IEEE:

- 375-384 [DOI: 10.1109/WACV48630.2021.00042]
- Lin C, Huang K, Wen Y M and Lu W. 2024. Copy-move forgery detection based on matching-pair clustering [J]. *Journal of Image and Graphics*, 29(12): 3595-3611 (蒯聪, 黄轲, 温雅敏, 卢伟. 2024. 匹配对聚类的图像复制粘贴篡改检测[J]. *中国图象图形学报*, 29(12): 3595-3611) [DOI: 10.11834/jig.230454]
- Lin T Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J et al. 2014. Microsoft COCO: common objects in context [C]//*European Conference on Computer Vision*. Cham, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- Ma X, Du B, Jiang Z, Al Hammadi A Y and Zhou J. 2023. IML-ViT: benchmarking image manipulation localization by vision transformer [EB/OL]. [2026-01-01]. <https://arxiv.org/abs/2307.14863> [DOI: 10.48550/arXiv.2307.14863]
- Nichol A, Dhariwal P, Ramesh A, Shyam P, Mishkin P, McGrew B et al. 2021. GLIDE: towards photorealistic image generation and editing with text-guided diffusion models [EB/OL]. [2026-01-02]. <https://arxiv.org/abs/2112.10741> [DOI: 10.48550/arXiv.2112.10741]
- Popescu A C and Farid H. 2004. Exposing digital forgeries by detecting duplicated image regions [R]. TR2004-515. Dartmouth College
- Su L, Ma X, Zhu X, Niu C, Lei Z and Zhou J Z. 2025. Can we get rid of handcrafted feature extractors? SparseViT: nonsemantics-centered, parameter-efficient image manipulation localization through sparse-coding transformer [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press: 7024-7032 [DOI: 10.1609/aaai.v39i7.32754]
- Tomioka Y and Kitazawa H. 2011. Digital camera identification using CFA pattern noise [C]//*2011 IEEE International Conference on Multimedia and Expo*. Piscataway, NJ, USA: IEEE: 1-4 [DOI: 10.1109/ICME.2011.6012060]
- Vaccari C and Chadwick A. 2020. Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news [J]. *Social Media + Society*, 6(1): 2056305120903408 [DOI: 10.1177/2056305120903408]
- Wang W, Dong J and Tan T. 2010. Tampered region localization of digital color images based on JPEG compression noise [C]//*International Workshop on Digital Watermarking*. Berlin, Germany: Springer: 120-133 [DOI: 10.1007/978-3-642-18405-5_10]
- Wu H, Zhou J, Tian J, Liu J and Qiao Y. 2022. Robust image forgery detection against transmission over online social networks [J]. *IEEE Transactions on Information Forensics and Security*, 17: 443-456 [DOI: 10.1109/TIFS.2022.3144878]
- Wu H, Zhou J, Tian J et al. 2022. Robust image forgery detection over online social network shared images [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE: 13440-13449 [DOI: 10.1109/CVPR52688.2022.01308]
- Wu Y, AbdAlmageed W and Natarajan P. 2019. ManTra-Net: manipulation tracing network for detection and localization of image forgeries [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE: 9543-9552 [DOI: 10.1109/CVPR.2019.00977]
- Yang C, Li H, Lin F, Jiang B and Zhao H. 2019. Constrained R-CNN: a general image manipulation detection model [EB/OL]. [2026-01-02]. <https://arxiv.org/abs/1911.08217> [DOI: 10.48550/arXiv.1911.08217]
- Zampoglou M, Papadopoulos S and Kompatsiaris Y. 2017. Large-scale evaluation of splicing localization algorithms for web images [J]. *Multimedia Tools and Applications*, 76(4): 4801-4834 [DOI: 10.1007/s11042-016-3795-2]
- Zanardelli M, Guerrini F, Leonardi R and Adami N. 2023. Image forgery detection: a survey of recent deep-learning approaches [J]. *Multimedia Tools and Applications*, 82(12): 17521-17566 [DOI: 10.1007/s11042-022-13797-w]
- Zhou P, Chen B C, Han X, Najibi M, Shrivastava A, Lim S et al. 2020. Generate, segment and refine: towards generic manipulation segmentation [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press: 13058-13065 [DOI: 10.1609/aaai.v34i07.7007]
- Zhou P, Han X, Morariu V I and Davis L S. 2018. Learning rich features for image manipulation detection [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ, USA: IEEE: 1053-1061 [DOI: 10.1109/CVPR.2018.00116]

作者简介

薛峻凜,男,硕士研究生,主要研究方向为图像处理与计算机视觉。E-mail:1283458936@qq.com

唐云祁,通信作者,男,教授,博士生导师,主要研究方向为人工智能模式识别、刑事科学技术等。E-mail:tangyunqi@pps.edu.cn

吴思慷,男,硕士研究生,主要研究方向为图像处理与计算机视觉。E-mail:1609902054@qq.com

余博雅,男,硕士研究生,主要研究方向为电子签名检验。E-mail:2534235825@qq.com