

任意尺度超分辨率图像篡改掩膜生成方法

徐雄¹, 杨欣宇², 朱学康², 杜博², 粟磊², 童炳魁², 雷泽宇^{2,3}, 周吉喆^{2,4*}(1. 中国电子科技集团公司第十研究所, 四川 成都 610036; 2. 四川大学计算机学院, 四川 成都 610207;
3. 澳门大学计算机与信息科学系, 澳门 999078; 4. 教育部机器学习与产业智能工程研究中心, 四川 成都 610065)

摘要: 在图像篡改定位(IML)领域, 现有数据集数量少、质量差, 难以支撑模型的泛化与鲁棒性。为此, 提出篡改掩膜生成(MMM)框架, 引入超分辨率模块以缓解原始图像与篡改图像清晰度差异带来的噪声问题, 并通过特征嵌入拼接与上下文建模生成高质量掩膜。基于 MMM 框架, 构建包含 11 069 对原始图像、篡改图像及掩膜的篡改掩膜生成数据集(MMMD), 其中涵盖复制移动、拼接、深度伪造(Deepfake)、图像修复和风格迁移等多种篡改方式。在 CASIAv2、NIST16 和 IMD2020 数据集上的实验结果表明, MMM 框架取得了较好性能[在 CASIAv2 上 F1 值为 0.96, 交并比(IoU)为 0.90], 并在多种模型中展现出优良的泛化能力。进一步地, 使用 MMMD 预训练的 MVSS-Net 和 IML-ViT 在多个数据集上的 F1 值显著高于在传统数据集上预训练的模型, 凸显了 MMMD 在推动图像取证与篡改检测研究中的价值。

关键词: 图像篡改定位; 视觉媒体数据集; 数据集生成; 任意尺度; 超分辨率

中图分类号: TP391.41

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0252288

Manipulation Mask Manufacturer for Arbitrary-Scale Super-Resolution Images

XU Xiong¹, YANG Xinyu², ZHU Xuekang², DU Bo², SU Lei², TONG Bingkui², LEI Zeyu^{2,3}, ZHOU Jizhe^{2,4*}

(1. The 10th Research Institute of China Electronics Technology Group Corporation, Chengdu 610036, Sichuan, China;

2. College of Computer Science, Sichuan University, Chengdu 610207, Sichuan, China;

3. Department of Computer and Information Science, University of Macau, Macao 999078, China;

4. Engineering Research Center of Machine Learning and Industry Intelligence, Ministry of Education, Chengdu 610065, Sichuan, China)

【Abstract】 In the field of Image Manipulation Localization (IML), the limited quantity and poor quality of existing datasets hinder model generalization and robustness. To address this, a Manipulation Mask Manufacturer (MMM) framework that integrates a super-resolution module to mitigate clarity discrepancies between the original and tampered images is proposed. The framework generates high-quality masks by embedding and concatenating features for effective context modeling. Based on the MMM framework, a Manipulation Mask Manufacturer Dataset (MMMD) is constructed, which contains 11 069 triplets of original images, manipulated images, and corresponding masks. MMMD encompasses diverse manipulation types, including copy-move, splicing, Deepfake, image inpainting, and style transfer. Experimental results demonstrate that MMM achieves strong performance on the CASIAv2, NIST16, and IMD2020 datasets, with F1 values of up to 0.96 and an Intersection over Union (IoU) of 0.90 on CASIAv2. Furthermore, models pretrained on MMMD, such as MVSS-Net and IML-ViT, consistently outperform those pretrained on conventional datasets across multiple benchmarks, highlighting the potential of the dataset to advance research in image forensics and manipulation detection.

【Key words】 Image Manipulation Localization (IML); dataset of visual media; dataset generation; arbitrary-scale; super-resolution

0 引言

数字图像处理技术的进步^[1]使 Adobe Photoshop^[2]和 GNU 图像处理程序(GIMP)^[3]等软件变得更加强大,促进了图像篡改的广泛应用,也导致越来越多的虚假信息、修饰过的照片或编辑过的视频在社交媒体上发布。在许多情况下,这些信息

在短短几天甚至几小时内就能“病毒式传播”^[4]。虚假信息和篡改图像的激增威胁到公众的知识、信任和安全。因此,图像篡改定位(IML)应运而生,在一些文献中,它也被称为伪造检测^[5]或篡改检测^[6],其目的是判断输入图像是否经过篡改或保持真实,并通过掩膜^[7]描绘图像中被篡改的确切部分。这些部分在语义上与原始内容(篡改前的原始图像)不同,

收稿日期: 2025-04-03 修回日期: 2025-07-25

基金项目: 四川省港澳台科技创新合作项目(2024YFHZ0355)。

通信作者 E-mail: *jzzhou@scu.edu.cn

它不包括纯生成的图像(例如从纯文本生成的图像)或通过图像处理技术引入的噪声及其他非语义变化,这些变化不会改变图像的底层含义。

然而,用于算法评估和公平比较的通用基准图像数据集仍然滞后^[8]。大多数现有数据集是人工创建和标注^[9]的,篡改类型和技术有限,数据集的规模也受到限制,导致模型的泛化能力和鲁棒性较差。因此,通过从互联网上获取大量原始图像和篡改图像来创建数据集是一个很好的方向。但来自互联网的图像和视频存在压缩和清晰度问题^[10],简单地将原始图像与篡改图像相减会产生噪声图像。

变化检测^[11-12]涉及识别同一位置在不同时间的变化,这与检测原始图像和篡改图像之间差异的任务相似。受此启发,本文将原始图像和篡改图像视为同一图片随时间的变化,将数据集生成任务视为变化检测任务。然而,由于任务中两张图像之间的清晰度差异^[13],直接使用变化检测模型并不理想。因此,本文引入超分辨率处理模块,在生成掩膜前增强两张图像的细节。这是本文所提出的篡改掩膜生成(MMM)框架的核心思想。

MMM 框架的流程是将原始图像和篡改图像输入全卷积网络(FCN)^[14],以提取高级特征。这些特征通过最大均值差异(MMD)^[15-16]进行对齐并拼接,然后通过跨尺度局部注意力块(CSLAB)^[17]和局部频率编码块(LFEB)^[17]处理,以提升分辨率和细节。最后,特征被分割、解码、相减,从而获得掩膜。

该框架在 IMD2020^[18]、NIST16^[19]和 CASIAv2^[20]数据集上取得了出色的标注结果。此外,本文创建了篡改掩膜生成数据集(MMMD),其中包含 11 069 对原始图像、篡改图像和掩膜,且具有持续增长的潜力。该数据集涵盖多种分辨率和篡改类型,如复制移动^[21-22]、拼接^[23]、变换^[24]、深度伪造(Deepfake)^[25]、图像修复^[26]、图像变形^[27]、重建^[28]和图像风格迁移^[29]。它包括多样化的图像,如卡通、肖像、风景、室内、食物和饰品。本文使用 MMMD 训练和测试了 MVSS-Net^[30]和 IML-ViT^[31]。其他 10 个在 CASIAv2^[20]上预训练的模型在该数据集上难以取得高指标,而在该数据集上训练的模型表现出更好的泛化能力。这凸显了现有数据集的局限性。总之,本文的贡献如下:

1)提出了 MMM 框架,即使原始图像与篡改图像之间存在显著清晰度差异,该框架也能准确标注它们之间的差异。

2)利用 MMM 框架生成了大型且多样化的 MMMD,以解决图像篡改检测领域数据集短缺的

问题。

3)使用 MMMD 预训练的模型取得了较高的 F1 值,并表现出较好的泛化能力。其他预训练模型在该数据集上表现不佳,凸显了现有篡改检测数据集的局限性。该数据集更好地反映了现实世界的篡改场景。

1 相关工作

1.1 现有篡改检测数据集生成方法

当前生成图像篡改检测数据集的方法包括手动篡改,即使用 Adobe Photoshop^[2]等工具手工编辑图像,这一过程耗时且需要专业技能^[32]。自动篡改利用软件工具和脚本快速生成大量数据,但图像可能看起来不自然或带有明显的篡改痕迹。图像合成通过组合不同图像的元素创建新的视觉场景,增强了数据集的多样性,但需要复杂的技术和大量的处理时间^[33]。改进的总变分去噪方法^[9]通过自动将篡改图像与原始图像相减来获得噪声掩膜,然后进行去噪处理,但由于噪声类型多样以及传统方法泛化能力不足,该方法往往无法有效转换多种类型的篡改图像。

1.2 现有变化检测方法

近年来,基于深度学习的变化检测方法发展迅速。CHEN 等^[34]提出了双时相图像变换器(BIT),该方法将输入图像转换为少量的语义标记,使用变换器编码器在紧凑的标记时空域中建模上下文信息,然后通过解码器将富含上下文的标记投影回像素空间,以增强原始特征,最终的变化检测结果通过特征差异图像生成。ChangeFormer^[35]采用基于变换器的孪生网络架构,从一对配准的遥感图像中进行变化检测。该方法结合层次变换器编码器和多层感知机(MLP)解码器,有效提取多尺度的长距离特征差异。这与图像篡改定位类似,多尺度技术在解码器中也常被使用^[36]。在 SNUNet-CD^[37]中,ChangeFormer 通过层次变换器编码器提取双时相图像的多尺度特征,并使用轻量级 MLP 解码器融合这些特征差异,生成变化检测图。然而,由于变化检测考虑的是同一图像在不同时间且清晰度一致的情况,它无需处理图像退化问题,因此现有的变化检测模型对于数据生成任务并不有效。

1.3 现有篡改检测数据集

图像篡改检测领域的数据集发展相对缓慢。目前,被广泛认可和使用的数据集仍然是四五年前甚至十多年前的数据集。

1.3.1 传统篡改技术的数据集

几乎所有数据集都包含传统篡改方法,如拼

接^[23]、复制移动^[21-22]、移除以及各种图像增强技术,以生成“假”或“伪造”图像。Columbia^[38]使用裁剪和拼接^[23],将来自其他图像的部分嵌入到单一图像中。CASIAv1^[20]使用 Adobe Photoshop^[2]进行剪切和粘贴,包括几何变换^[24],如缩放^[39]和旋转^[40]。CASIAv2^[20]增加了更多的后期处理,并拥有更丰富的图像种类,分为 8 个类别:场景、动物、建筑、人物、植物、物体、自然和纹理。COVERAGE^[33]由 iPhone 6 前置摄像头拍摄的真实图像组成,使用 Photoshop CS4 进行处理,篡改方法包括平移、缩放、旋转、自由变换^[24]、光照变化及其组合。NIST^[19]采用局部像素修改、压缩、噪声添加、模糊和几何变换^[24]。采用拼接^[23]、复制移动^[21-22]、扭曲^[41]和对对象移除^[42]的数据集有 DEFACTo^[43],它基于 MSCOCO^[44]数据库,旨在生成具有语义意义的伪造图像。

1.3.2 基于深度学习的篡改数据集

现代图像篡改技术通过人工智能和深度学习,特别是生成对抗网络(GAN),实现了前所未有的真实感。这些技术包括 Deepfake^[25],能够进行面部替换、表情合成以及生成不存在的人的图像,使图像和视频篡改显得非常逼真^[45]。深度学习还在图像修复和增强方面表现出色,通过去噪、填充缺失部分和提高分辨率,使损坏的图像焕然一新,使低分辨率图像变得清晰且细节丰富。TensorFlow、PyTorch、

Keras 和 OpenCV 等工具和框架大大简化了这些技术的实现和应用。在 IMD2020^[18]中,GAN 被用于生成图像的篡改区域,修复技术^[26]被用来填充图像中缺失或损坏的部分,使其看起来自然且连贯。这也为图像篡改检测带来了更大的挑战。当前模型越来越需要多样化的数据,以更好地反映现实世界的场景。

2 提出的方法

2.1 整个框架的流程

MMM 框架整体分为 3 个模块:特征提取与拼接、超分辨率处理以及特征分离掩膜生成。本研究从网络中获取原始图像和大量篡改图像,仅保留与原始数据尺寸相同的图像。在获得原始图像和篡改图像后,将其输入 MMM 框架。该框架使用 FCN^[14]从原始图像和篡改图像中提取高级特征。这些特征通过 MMD^[15-16]进行对齐并拼接。拼接高级特征的思路受到 BIT^[34]的启发。在超分辨率处理阶段,拼接后的特征由 CSLAB^[17]和 LFEB^[17]处理,以增强图像的分辨率和细节表达。随后,该框架分离这些嵌入特征,使用解码器生成残差图像,并将其与原始图像结合。通过计算原始图像和篡改图像高分辨率特征之间的绝对差异,最终生成掩膜。整个 MMM 框架结构如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。

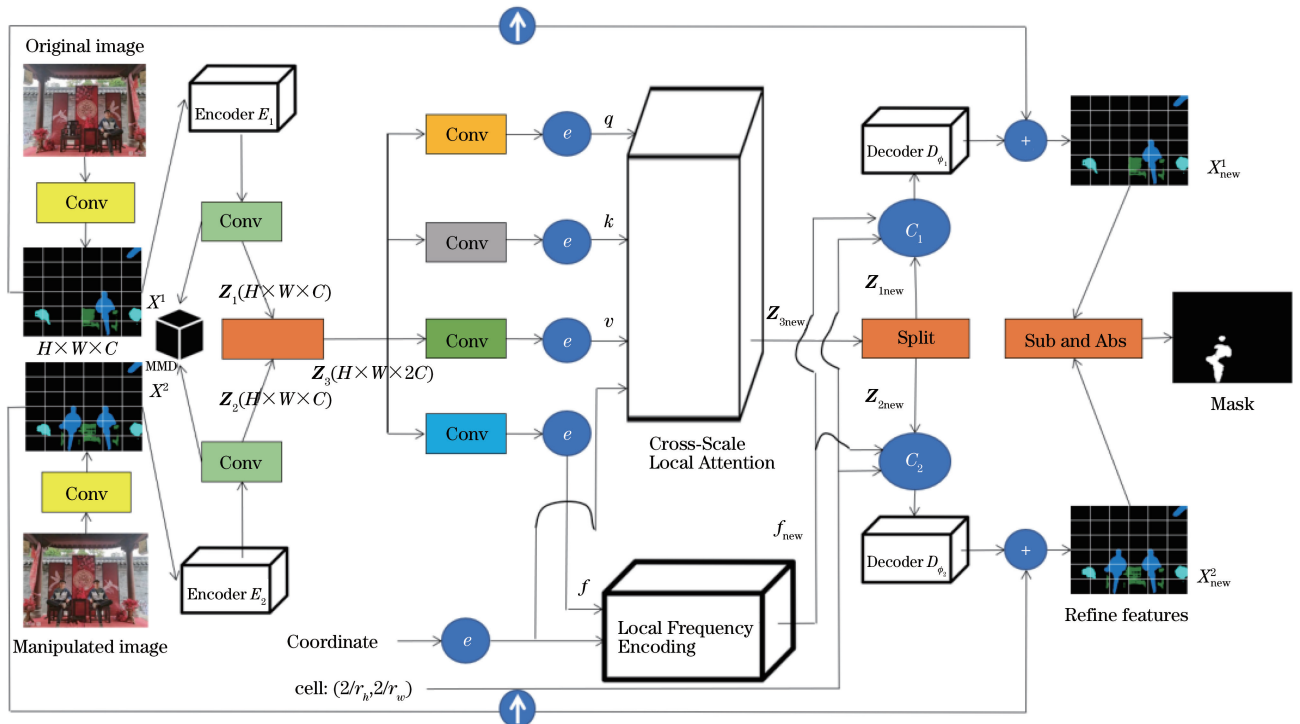


图 1 MMM 框架

Fig.1 MMM framework

2.2 特定处理算法

2.2.1 图像特征的提取与拼接

首先,本文使用 FCN^[14] 分别从原始图像和篡改图像中提取高级特征。然后,这些特征被输入到编码器和编码器中,得到特征嵌入和(以下称为“和”)。和通过 MMD^[14-38] 计算来消除数据分布的差异,使模型更专注于内容的差异。同时和被拼接成 $Z_3 \in \mathbb{R}^{H \times W \times 2C}$ 。

2.2.2 任意尺度超分辨率处理

通过 4 个独立的卷积层进行投影,以获得 4 个潜在嵌入,分别对应查询、键、值和频率。由于 2 张图像的尺寸相同,本文使用原始图像的坐标和单元格。原始图像和篡改图像将基于任意上采样比例在二维低分辨率坐标中生成二维高分辨率坐标。接下来,二维坐标连同和被输入到 CSLAB^[7] 中,以估计局部潜在嵌入。和二维坐标也将被输入到 LFEB^[7] 中,以估计局部频率嵌入。具体来说,和表示用于执行局部坐标采样的局部网格的高度和宽度。CSLAB^[17] 和 LFEB^[17] 按如下方式估计和:

$$Z_{3\text{new}} = \text{CSLAB}(\delta x, q, k, v) \quad (1)$$

$$f_{\text{new}} = \text{LFEB}(\delta x, f) \quad (2)$$

$$\delta x = \{x_q - x^{(i,j)}\}_{i \in \{1,2,\dots,G_h\}, j \in \{1,2,\dots,G_w\}} \quad (3)$$

CSLAB 和 LFEB 借鉴了 CHEN 等^[17] 的工作,具体结构如图 2 所示。CSLAB 的主要功能是聚合跨尺度的局部特征信息,利用查询(q)、键(k)和值(v)特征来增强模型捕捉细粒度细节的能力,从而提高图像篡改检测的准确性。LFEB 负责对输入的频率特征进行局部频率编码,通过捕捉局部频率变化来增强模型对边缘和纹理信息的敏感性。

2.2.3 图像特征分离及掩膜生成

$Z_{1\text{new}}$ 和 $Z_{2\text{new}}$ 是从 $Z_{3\text{new}}$ 分割得到的。解码器 D_{ϕ_1} 和解码器 D_{ϕ_2} 分别利用这些嵌入以及提供的单元格生成残差图像。然后,对原始图像和篡改图像进行上采样,并将其与对应的残差图像逐元素相加。这生成了原始图像和篡改图像的高分辨率高级特征。本文将这两个高分辨率高级特征相减并取绝对值,以获得最终的掩膜。

损失函数模型的损失函数定义如下:

$$L = \frac{1}{H_p \times W_p} \sum_{h=1}^H \sum_{w=1}^W l(P_{hw}, M_{hw}) \quad (4)$$

式中: $H_p \times W_p$ 是图像中像素的总数; P_{hw} 是模型在位置(h, w)处的概率分布; M_{hw} 是位置(h, w)处的掩码标签; $l(P_{hw}, M_{hw})$ 是交叉熵。

$$l(P_{hw}, M_{hw}) = - \sum_c M_{hw}^{(c)} \ln(P_{hw}^{(c)}) \quad (5)$$

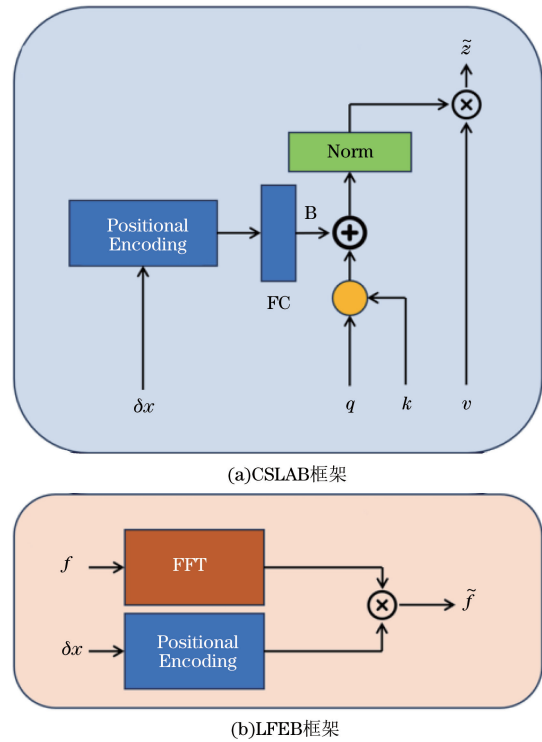


图 2 CSLAB 和 LFEB 的框架

Fig. 2 The framework of CSLAB and LFEB

式中: c 表示类别索引; $M_{hw}^{(c)}$ 是一个二元指示符(0 或 1),表示类别标签 c 是否是位置(h, w)的正确分类; $P_{hw}^{(c)}$ 是位置(h, w)类别 c 的预测概率。交叉熵衡量了真实标签分布与预测概率分布之间的差异,用于通过最小化这种差异来优化模型参数。

3 实验结果和分析

3.1 MMMD 的创建

本文从互联网上包含篡改图像的网站(如百度 PS 吧)爬取了图像。这是一个获取原始图像和篡改图像的理想来源^[9]。大多数用户会请求他人帮助修改他们提供的图片。因此,在他们的帖子下通常会有许多篡改图像。从不同的帖子中保存原始图像和篡改图像,从而能够在短时间内收集大量数据。本文将每个帖子的第一张图片视为原始图像,如图 3 的第 1 行所示,并将该帖子中所有其他相同尺寸的图片视为篡改图像,如图 3 的第 2 行所示。

表 1 是现有主要数据集的图片数量对比。观察表 1 可发现,DEFACTO 数据集中篡改图像的数量为 149 587,这一数量远远超过了其他数据集,充分体现该数据集在数据规模方面具有的优势,不过 MMMD 在数量上同样有着出色的表现,远超 Columbia(180 张)、CASIAv1(920 张)等早期数据集,表明其在数据丰富度上的潜力。

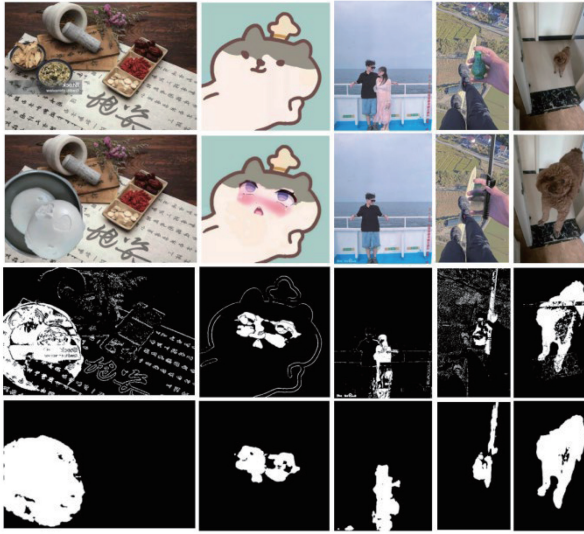


图 3 MMM 框架生成的结果图像

Fig.3 MMM framework-generated result images

表 1 现有篡改检测数据集图片数量对比

Table 1 Comparison of the number of images in existing tampering detection datasets

Dataset	Number of Images
Columbia	180
CASIAv1	920
CASIAv2	5 063
Coverage	100
NIST16	564
DEFACTO	149 587
IMD2020	2 010
MMMD(Ours)	11 069

随后对原始图像和篡改图像的灰度图像进行相减,得到包含大量噪声的掩码图像。这是因为图像在互联网传输过程中会经历不可逆的压缩和质量下降,并且在使用如 Photoshop 等图像编辑软件打开时也会经历压缩。原始图像和篡改图像包含人眼难以察觉的噪声差异,这使得直接相减得到的掩码图像不适合作为训练数据。NIST16 数据集中的图像是多尺度的,因此本文使用它来预训练 MMM 框架。本文将原始图像及其对应的篡改图像成对输入到 MMM 框架中,该框架包括 3 个步骤:特征提取与拼接、超分辨率处理以及特征分离掩码生成。这一过程最终生成了图 2 最后一行的预测掩码图像。预测图像的噪声显著减少,可用于篡改检测模型。由于假设 PS 论坛的第一张图像是原始图像,后续图像是篡改图像,但这并不总是正确的,这导致了噪声较多、几乎全白的掩码。特殊的篡改技术也会导致噪声掩码。因此,白区域超过 70%或少于 1%的掩码被视为无效并移除。

整个 MMMD 被分为 3 组:原始图像、篡改图像和预测掩码,每组包含 11 069 张图像。每张篡改图像在另外两组中都有对应的原始图像和掩码。该数据集包含不同分辨率的图像和多种篡改类型,包括复制移动^[21-22]、变换^[24]、Deepfake^[25]、图像修复^[26]、变形^[27]、重建^[28]和风格迁移。它涵盖了广泛的图像类别,如卡通、肖像、风景、室内、食物和饰品。

3.2 模型在现有数据集上的实验结果

本文创新性地使用深度学习来标注图像篡改检测数据集,目前尚无现有的可比较方法。因此,本文在 NIST16^[19]、IMD2020^[18]和 CASIAv2.0^[20]数据集上训练并验证模型,以展示其在区分原始图像和篡改图像方面的有效性。

3.2.1 实验细节

该模型基于 PyTorch 实现。训练设置学习率为 0.01,最大训练轮数为 100 个周期。学习率衰减迭代次数设置为 100。每次训练周期后进行验证,将在验证集上表现最佳的模型用于测试集的评估。实验在一台配备 16 GB 内存、i7-11800H CPU 和 RTX 3060 GPU 的机器上进行。

3.2.2 测试指标

本文使用准确率 (Accuracy, A)、精确率 (Precision, P)、召回率 (Recall, R)、F1 值 (F_1 , F_1)、交并比 (IoU, r_{IoU})来评估模型的性能。

F1 值在篡改检测 (召回率) 和避免误报 (精确率) 之间取得平衡,防止仅追求高召回率或高精度带来的偏差。F1 值的计算公式如下:

$$F_1 = 2 \times \left(\frac{P \times R}{P + R} \right) \quad (6)$$

IoU 能反映模型的定位准确性,准确率能反映模型的整体性能。IoU 的计算公式如下:

$$r_{IoU} = \frac{N_{TP}}{N_{TP} + N_{FN} + N_{FP}} \quad (7)$$

准确率的计算公式如下:

$$A = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FN} + N_{FP}} \quad (8)$$

式中: N_{TP} 、 N_{FP} 、 N_{FN} 、 N_{TN} 分别表示真正例、假正例、假负例和真负例数量。

3.2.3 实验结果

实验结果如表 2 所示。该模型在 3 个数据集上表现出色,在所有指标上均达到高水平。它在图像篡改检测数据集上展示了强大的泛化能力,特别是在涉及传统篡改方法的数据集上。性能在 IMD2020^[18]上略有下降,该数据集使用了如 GAN

和修复^[26]等深度学习技术,但总体而言,模型在各种数据集上表现出良好的适应性和性能。

表 2 MMM 模型在不同数据集上的性能

Dataset	F1	IoU	Precision	Recall	Accuracy
CASIAv2	0.96	0.90	0.95	0.96	0.99
NIST16	0.94	0.89	0.95	0.92	0.98
IMD2020	0.88	0.81	0.89	0.88	0.96

3.3 生成数据集在现有模型上的实验结果

MDL-BenCo^[46]重现了主流的 IML 模型,本文的模型实验基于此框架进行。本文使用 MMMD 训练了 2 个主要模型 MVSS-Net^[30]和 IML-ViT^[31],并验证了它们的泛化能力,结果如表 3 所示。

表 3 在 MMMD 上预训练的各种模型在不同数据集上的 F1 值

Model	Columbia	NIST16	IMD2020	COVERAGE
MVSS-Net	0.39/0.49	0.25/0.30	0.28/0.32	0.25/0.35
IML-ViT	0.77/0.81	0.30/0.34	0.33/0.37	0.42/0.30

如表 3 中所示,使用 MMMD 训练的 MVSS-Net 和 IML-ViT 在各种数据集上的指标高于在 CASIAv2 上预训练的模型。它们在 MMMD 上表现出更高的泛化和鲁棒性。由于 MMMD 比常用数据集大得多,且覆盖了更广泛的篡改类型和场景,它有助于提升现有篡改检测模型的泛化能力。在 MMMD 上预训练的模型有望在其他数据集上获得更好的性能。

实验结果通过图 4 进行可视化,展示了 MMMD 预训练模型在 F1 值上的整体优势。柱状图清晰表明,MMMD 预训练的 MVSS-Net 和 IML-ViT 在 Columbia 和 IMD2020 等复杂数据集上的 F1 值显著高于 CASIAv2 预训练模型。

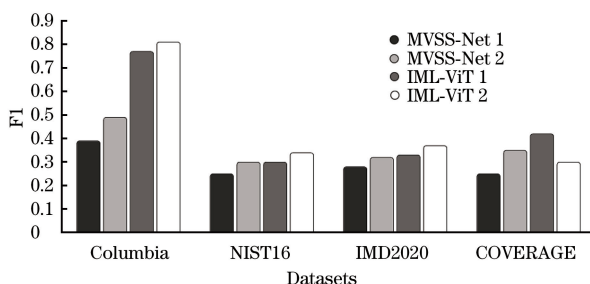


图 4 在 MMMD 上预训练的各种模型在不同数据集上的 F1 值可视化图

Fig.4 F1 values visualization chart of various models pre-trained by MMMD on different datasets

本文使用 MMMD 验证了图像篡改检测模型 ManTra-Net^[47]、MVSS-Net^[30]、CAT-Net^[48]、ObjectFormer^[49]、NCL-IML^[50]、TruFor^[51]、IML-ViT^[31]和 PSCC-Net^[52],这些模型均在 CASIAv2^[20]上预训练,并发现了现有数据集与 MMMD 相比的不足之处,结果如表 4 所示。

表 4 在 CASIAv2 上预训练的各种模型在不同数据集上的 F1 值

Model	COVERAGE	Columbia	CASIAv1	MMMD (Ours)
Mantra-Net	0.08	0.46	0.12	0.08
MVSS-Net	0.27	0.40	0.53	0.27
CAT-Net	0.29	0.55	0.58	0.30
ObjectFormer	0.30	0.34	0.44	0.32
NCL-IML	0.22	0.45	0.50	0.26
TruFor	0.40	0.88	0.71	0.30
IML-ViT	0.43	0.76	0.72	0.24
PSCC-Net	0.23	0.60	0.38	0.32

如表 4 所示,在 CASIAv2^[20]上预训练的各种模型在 MMMD 上难以取得高指标。CASIAv2 是近年来用于篡改检测的较大数据集之一,涵盖了广泛的篡改类型,常被用作模型的预训练数据集。然而,在 CASIAv2 上预训练的模型在 MMMD 上的指标低于其他数据集,这凸显了传统数据集的局限性。与 MMMD 相比,这些数据集规模较小,覆盖的篡改类型较少,且与现实世界的篡改图像仍存在一定距离。因此,在这些数据集上训练的模型难以在现实生活中的篡改图像上表现出色。

本文采用图 5 的分组柱状图方式进行可视化呈现,以此来展示各个模型的 F1 值。从图中可以看出,在 MMMD 数据集上, ObjectFormer 和 PSCC-Net 的 F1 值处于最高水平,然而 ManTra-Net 的表现却是最差的,这种情况或许和模型架构的特征提取机制存在关联。ObjectFormer 以及 PSCC-Net 把多尺度特征提取与上下文建模相结合,可更为有效地捕捉 MMMD 中的复杂篡改特征(如图像修复^[14]、风格迁移^[24]),但适应性较差,对于多样化的篡改类型适应能力欠佳,图 5 中 ManTra-Net 的柱状图高度是最低的,这很直观地体现出了它的性能存在不足。

当直接处理图像而不使用超分辨率模块时,对于某些严重退化的图像,仍然会得到未经超分辨率处理的噪声图像。这些图像不适合用作篡改检测模型的训练数据。

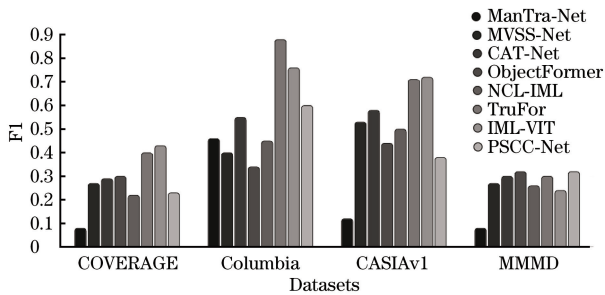


图 5 在 CASIAv2 上预训练的各模型在各个数据集上的 F1 值可视化图

Fig.5 F1 values visualization chart of various models pre-trained by CASIAv2 on different datasets

4 结束语

本文创新性地提出了一个篡改掩膜生成框架,用于生成图像篡改检测数据集。它解决了图像篡改检测领域中数据集规模小、质量差以及篡改检测类型有限的问题。该框架通过拼接图像特征嵌入、执行上下文建模并捕捉像素之间的长距离关系,利用所提出的数据集消除数据分布的差异。大量实验验证了所提方法的有效性。本文展示了该框架在现有数据集上的强大性能。本文提出的数据集能更好地贴合篡改检测模型在现实世界中必须面对的篡改场景,这将有助于这些模型在实际应用中提升泛化能力和鲁棒性。本文还证明了该框架在训练效果上优于其他数据集。

参考文献

- [1] 骆伟祺, 黄继武, 丘国平. 鲁棒的区域复制图像篡改检测技术[J]. 计算机学报, 2007, 30(11): 1998-2007.
LUO W Q, HUANG J W, QIU G P. Robust detection of region-duplication forgery in digital image [J]. Chinese Journal of Computers, 2007, 30 (11): 1998-2007. (in Chinese)
- [2] FAULKNER A, CHAVEZ C. Adobe Photoshop CC classroom in a book[M]. [S. l.]: Adobe Press, 2013.
- [3] GIMP Development Team. GIMP: GNU image manipulation program[EB/OL]. [2024-11-01]. <https://www.gimp.org/>.
- [4] FITZPATRICK N. Media manipulation 2.0: the impact of social media on news, competition, and accuracy[J]. Athens Journal of Mass Media and Communications, 2018, 4(1): 45-62.
- [5] FARID H. Image forgery detection [J]. IEEE Signal Processing Magazine, 2009, 26(2): 16-25.
- [6] WANG W, DONG J, TAN T N. A survey of passive image tampering detection[M]//HO A T S, SHI Y Q, KIM H J, et al. Digital watermarking. Berlin, Germany: Springer, 2009: 308-322.
- [7] SU L, MA X C, ZHU X K, et al. Can we get rid of handcrafted feature extractors? Sparsevit: nonsemantics-centered, parameter-efficient image manipulation localization through sparse-coding transformer[EB/OL]. [2024-11-01]. <https://arxiv.org/pdf/2307.14863v2>.
- [8] DANG-NGUYEN D T, PASQUINI C, CONOTTER V, et al. RAISE: a raw images dataset for digital image forensics[C]//Proceedings of the 6th ACM Multimedia Systems Conference. New York, USA: ACM, 2015: 219-224.
- [9] YANG X Y, ZHOU J Z. Manipulation mask generator: high-quality image manipulation mask generation method based on modified total variation noise reduction [C] // Proceedings of the IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML). Washington D. C., USA: IEEE Press, 2023: 218-223.
- [10] BOUDIER T, SHOTTON D M. Video on the Internet: an introduction to the digital encoding, compression, and transmission of moving image data[J]. Journal of Structural Biology, 1999, 125(2/3): 133-155.
- [11] RENSINK R A. Change detection [J]. Annual Review of Psychology, 2002, 53: 245-277.
- [12] LU D, MAUSEL P, BRONDIZIO E, et al. Change detection techniques [J]. International Journal of Remote Sensing, 2004, 25(12): 2365-2401.
- [13] TAN Y, ZHENG H T, ZHU Y H, et al. CrossNet++: cross-scale large-parallax warping for reference-based super-resolution[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(12): 4291-4305.
- [14] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE Press, 2015: 3431-3440.
- [15] SEJDINOVIC D, SRIPERUMBUDUR B, GRETTON A, et al. Equivalence of distance-based and RKHS-based statistics in hypothesis testing[J]. The Annals of Statistics, 2013, 41(5): 2263-2291.
- [16] LONG M S, CAO Y, WANG J M, et al. Learning transferable features with deep adaptation networks [C] // Proceedings of International Conference on Machine Learning. [S. l.]: PMLR, 2015: 97-105.
- [17] CHEN H W, XU Y S, HONG M F, et al. Cascaded local implicit transformer for arbitrary-scale super-resolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, Canada: IEEE Press, 2023: 18257-18267.
- [18] NOVOZAMSKY A, MAHDIAN B, SAIC S. IMD2020: a large-scale annotated dataset tailored for detecting manipulated images [C] // Proceedings of the IEEE Winter Applications of Computer Vision Workshops (WACVW). Snowmass Village, USA: IEEE Press, 2020: 71-80.
- [19] GUAN H Y, KOZAK M, ROBERTSON E, et al. MFC datasets: large-scale benchmark datasets for media forensic challenge evaluation [C] // Proceedings of the IEEE Winter Applications of Computer Vision Workshops (WACVW). Waikoloa Village, USA: IEEE Press, 2019: 63-72.
- [20] DONG J, WANG W, TAN T N. CASIA image tampering detection evaluation database [C] // Proceedings of the IEEE China Summit and International Conference on Signal and Information Processing. Beijing, China: IEEE Press, 2013: 422-426.
- [21] AMERINI I, BALLAN L, CALDELLI R, et al. A SIFT-based forensic method for copy-move attack detection and transformation recovery [J]. IEEE Transactions on Information Forensics and Security, 2011, 6(3): 1099-1110.
- [22] CHRISTLEIN V, RIESS C, JORDAN J, et al. An evaluation of popular copy-move forgery detection approaches [J]. IEEE Transactions on Information Forensics and Security, 2012, 7(6): 1841-1854.
- [23] HSU Y F, CHANG S F. Detecting image splicing using geometry invariants and camera characteristics consistency [C] // Proceedings of the IEEE International Conference on Multimedia and Expo. Toronto, Canada: IEEE Press, 2006: 549-552.

- [24] PETROU M M P, PETROU C. Image processing: the fundamentals[M]. [S. l.]: John Wiley & Sons, Inc. 2010.
- [25] MIRSKY Y, LEE W K. The creation and detection of deepfakes[J]. *ACM Computing Surveys*, 2022, 54(1): 1-41.
- [26] ELHARROUSS O, ALMAADEED N, AL-MAADEED S, et al. Image inpainting: a review [J]. *Neural Processing Letters*, 2020, 51(2): 2007-2028.
- [27] WOLBERG G. Image morphing: a survey[J]. *The Visual Computer*, 1998, 14(8): 360-372.
- [28] DEMOMENT G. Image reconstruction and restoration: overview of common estimation structures and problems[J]. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1989, 37(12): 2024-2036.
- [29] JING Y C, YANG Y Z, FENG Z L, et al. Neural style transfer: a review [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2020, 26(11): 3365-3385.
- [30] DONG C B, CHEN X R, HU R H, et al. MVSS-Net: multi-view multi-scale supervised networks for image manipulation detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(3): 3539-3553.
- [31] MA X C, DU B, LIU X G, et al. IML-ViT: Image manipulation localization by vision transformer [EB/OL]. [2024-11-01]. <https://arxiv.org/pdf/2307.14863>.
- [32] THAKUR R, ROHILLA R. Recent advances in digital image manipulation detection techniques: a brief review[J]. *Forensic Science International*, 2020, 312: 110311.
- [33] WEN B H, ZHU Y, SUBRAMANIAN R, et al. COVERAGE—a novel database for copy-move forgery detection [C] // *Proceedings of the IEEE International Conference on Image Processing (ICIP)*. Phoenix, USA: IEEE Press, 2016: 161-165.
- [34] CHEN H, QI Z P, SHI Z W. Remote sensing image change detection with transformers [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-14.
- [35] BANDARA W G C, PATEL V M. A transformer-based Siamese network for change detection[C]//*Proceedings of the 2022 IEEE International Geoscience and Remote Sensing Symposium*. Kuala Lumpur, Malaysia: IEEE Press, 2022: 207-210.
- [36] ZHU X K, MA X C, SU L, et al. Mesoscopic insights: orchestrating multi-scale & hybrid architecture for image manipulation localization[EB/OL]. [2024-11-01]. <https://arxiv.org/pdf/2412.13753>.
- [37] FANG S, LI K Y, SHAO J Y, et al. SNU-Net-CD: a densely connected Siamese network for change detection of VHR images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.
- [38] SHI J B, MALIK J. Normalized cuts and image segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(8): 888-905.
- [39] ANDREADIS I, AMANATIADIS A. Digital image scaling[C]//*Proceedings of the IEEE Instrumentation and Measurement Technology Conference*. Ottawa, Canada: IEEE Press, 2005: 2028-2032.
- [40] GINSBERG R H. Image rotation[J]. *Applied Optics*, 1994, 33(34): 8105_1.
- [41] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. *IEEE Transactions on Image Processing*, 2004, 13(9): 1200-1212.
- [42] GLASBEY C A, MARDIA K V. A review of image-warping methods[J]. *Journal of Applied Statistics*, 1998, 25(2): 155-171.
- [43] MAHFOUDI G, TAJINI B, RETRAINT F, et al. DEFACTo: image and face manipulation dataset [C] // *Proceedings of the 27th European Signal Processing Conference (EUSIPCO)*. A Coruna, Spain: IEEE Press, 2019: 1-5.
- [44] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C] // *Proceedings of ECCV 2014*. Berlin, Germany: Springer, 2014: 740-755.
- [45] ZHUANG P Y, LI H D, TAN S Q, et al. Image tampering localization using a dense fully convolutional network [J]. *IEEE Transactions on Information Forensics and Security*, 2021, 16: 2986-2999.
- [46] MA X C, ZHU X K, SU L, et al. IMDL-BenCo: a comprehensive benchmark and codebase for image manipulation detection & localization [EB/OL]. [2024-11-01]. <https://arxiv.org/pdf/2406.10580>.
- [47] WU Y, ABDALMAGEED W, NATARAJAN P. ManTra-Net: manipulation tracing network for detection and localization of image forgeries with anomalous features[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, USA: IEEE Press, 2019: 9535-9544.
- [48] KWON M J, NAM S H, YU I J, et al. Learning JPEG compression artifacts for image manipulation detection and localization [J]. *International Journal of Computer Vision*, 2022, 130(8): 1875-1895.
- [49] WANG J K, WU Z X, CHEN J J, et al. ObjectFormer for image manipulation detection and localization [C] // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, USA: IEEE Press, 2022: 2354-2363.
- [50] ZHOU J Z, MA X C, DU X, et al. Pre-training-free image manipulation localization through non-mutually exclusive contrastive learning [C] // *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE Press, 2023: 22289-22299.
- [51] GUILLARO F, COZZOLINO D, SUD A, et al. TruFor: leveraging all-round clues for trustworthy image forgery detection and localization[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, Canada: IEEE Press, 2023: 20606-20615.
- [52] LIU X H, LIU Y J, CHEN J, et al. PSCC-Net: progressive spatio-channel correlation network for image manipulation detection and localization[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(11): 7505-7517.