HRDA-Net：面向真实场景的图像多篡改检测与定位算法

朱叶[1,2]，余宜林[1]，郭迎春[1]

（1. 河北工业大学人工智能与数据科学学院，天津 300401；2. 深圳市媒体信息内容安全重点实验室，广东 深圳 518060）

摘 要：针对主流篡改数据集单幅图像仅包含一类篡改操作，且对真实图像定位存在“伪影”问题，构建面向真实场景的多篡改数据集（MM Dataset），每幅篡改图像包含拼接和移除 2 种篡改操作。针对多篡改检测与定位任务，提出端到端的高分辨率扩张卷积注意力网络（HRDA-Net），利用自顶向下扩张卷积注意力（TDDCA）模块融合图像 RGB 域和 SRM 域特征。最后，采用混合扩张卷积模块（MDC）分别提取拼接、移除和篡改检测任务特征，实现篡改区域定位和篡改置信度预测。为提高网络训练效率，提出余弦相似度损失函数作为辅助损失。实验结果表明，在 MM Dataset 下，与主流语义分割方法相比，HRDA-Net 具有较优的性能和较强的稳健性；在单篡改数据集 CASIA 和 NIST 下，与主流单篡改定位方法相比，HRDA-Net 的 F1 和 AUC 分数均较优。

关键词：深度学习；多篡改检测与定位；多篡改数据集；余弦相似度损失函数

中图分类号：TP391

文献标志码：A

DOI: 10.11959/j.issn.1000−436x.2022016

HRDA-Net: image multiple manipulation detection and location algorithm in real scene

ZHU Ye[1,2], YU Yilin[1], GUO Yingchun[1]

1. School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China

2. Shenzhen Key Laboratory of Media Security, Shenzhen 518060, China

Abstract: Aiming at the problems that the fake image just contains one tampered operation in mainstream manipulation datasets and the artifact is a common problem in manipulation location. The multiple manipulation dataset (MM Dataset) was constructed for real scene, which contained both splicing and removal in each images. Based on this, an end-to-end high-resolution representation dilation attention network (HRDA-Net) was proposed for multiple manipulation detection and localization, which fused the RGB and SRM features through the top-down dilation convolutional attention (TDDCA). Finally, the mixed dilated convolution (MDC) would respectively extract the features of splicing and removal, which could realize multiple manipulation location and confidence prediction. The cosine similarity loss was proposed as auxiliary loss to improve the efficiency of network. Experimental results on MM Dataset indicate that the performance and robustness of HRDA-Net is better than semantic segmentation methods. Furthermore, the scores of F1 and AUC are greater than state-of-the-art manipulation location methods in CASIA and NIST datasets.

Keywords: deep learning, multiple manipulation detection and location, MM Dataset, cosine similarity loss function

---

收稿日期：2021−09−27；**修回日期**：2021−12−22

通信作者：郭迎春，gyc@scse.hebut.edu.cn

基金项目：国家自然科学基金资助项目（No.62102129, No.61806071, No.91746207）；河北省自然科学基金资助项目（No.F2021202030, No.F2020202025, No.F2019202381, No.F2019202464）；河北省高等学校科学技术研究基金资助项目（No.QN2019207, No.QN2020185）

Foundation Items: The National Natural Science Foundation of China (No.62102129, No.61806071, No.91746207), The Natural Science Foundation of Hebei Province (No.F2021202030, No.F2020202025, No.F2019202381, No.F2019202464), The Sci-Tech Research Projects of Higher Education of Hebei Province (No.QN2019207, No.QN2020185)

0 引言

图像编辑软件的日益普及，使图像编辑越来越容易，甚至单张图像存在多类篡改操作。因此，多篡改图像的检测与定位任务的研究至关重要。

主流图像篡改盲取证方法可分为基于图像级的篡改检测和基于像素级的篡改定位[1]。基于双通道 R-FCN（region-based fully convolutional network）模型在篡改检测任务中表现了较高性能[2]。基于空间光照一致性[3]的图像篡改检测算法，对噪声后处理稳健性较高。随着卷积网络的发展，基于预训练卷积神经网络（CNN, convolutional neural network）[4]和基于分块 CNN[5]的卷积网络训练方式被提出以检测篡改图像。针对像素级图像单一篡改，Liu 等[6]提出基于全卷积网络和条件随机场的拼接篡改像素级定位框架，但是受数据集限制容易陷入过拟合。继而，文献[7]提出环形残差 U-net 网络，加强 CNN 的学习能力。文献[8]使用 U 型检测网络和全连接条件随机场进行篡改检测。

针对像素级的篡改定位，文献[9]提出基于串行模型的篡改分支和相似度分支融合的复制−粘贴定位算法，即将篡改区域和源区域定位后再进行分类；文献[10]提出基于自适应注意力机制和残差细化模块的卷积网络框架，进一步提升篡改区域定位精度。针对图像移除篡改取证，Li 等[11]利用不同方向高通滤波器提出增强篡改痕迹特征，进行图像移除篡改定位。另外，针对包含多类篡改的数据集，文献[12]采用多尺度卷积得到篡改概率图，并与分割结果进行融合，实现篡改区域的像素级定位。Bappy 等[13]采用长短期记忆网络和 CNN 的混合网络学习边缘矛盾特征，实现了端到端的篡改定位网络。基于 RGB 流+噪声流的双流网络[14]、ManTra-net[15]、空间金字塔注意力网络（SPAN, spatial pyramid attention network）[16]陆续被提出，实现了对拼接、复制−粘贴和图像移除篡改的像素级定位。

综上所述，主流篡改检测和定位方法都是针对单一类别篡改取证问题，其应用场景非常局限。主流图像篡改数据集仅包含单一篡改操作，且训练数据集仅包含篡改数据，对真实图像定位存在“伪影”现象，即篡改数据集训练的篡改定位网络在真实图像上仍能定位到篡改区域。并且，目前主流篡改数据集单幅图像仅包含单一种类篡改操作。为此，本文构建面向真实场景的多篡改数据集（MM Dataset, multiple manipulation dataset），每幅篡改图像同时包含拼接和移除 2 种篡改操作。以此为基础，本文提出面向真实场景的多篡改检测与定位算法，即高分辨率扩张卷积注意力网络（HRDA-Net, high-resolution representation dilation attention network），以高分辨率网络（HRNet, high-resolution representation network）[17]为基线，同时完成篡改检测和定位任务。本文的主要贡献如下。

1) 构建面向真实场景的 MM Dataset，为真实场景下多篡改操作检测与定位提供支持。

2) 提出 HRDA-Net 模型，同时进行多篡改检测与定位任务，即检测图像是否篡改，同时定位拼接和移除篡改区域。

3) 创新性地引入了余弦相似度损失作为辅助的损失函数，有效加快网络更好收敛。

1 相关工作

1.1 图像篡改数据集

基于深度学习的图像篡改取证方法离不开大规模数据集的支持，目前公开的篡改数据集有 CASIA[18]、NIST[19]、COVERAGE[20]、DEFACTO[21]等。其中，CASIA 包含拼接篡改和复制−粘贴篡改；NIST 包含拼接、复制−粘贴和移除篡改；COVERAGE 包含 100 幅复制−粘贴篡改图像；DEFACTO 从 COCO[22]数据集中挑选了 149 000 幅图像，并且自动对图像进行拼接、复制−粘贴和移除操作。但是，以上公开数据集篡改图像仅包含单一种类篡改操作，目前没有同时包含多类篡改操作的公开数据集。本文提出面向真实场景的 MM Dataset，包含 1 000 组篡改图像和真实图像，每一幅图像都包含拼接和移除 2 种篡改操作。

1.2 注意力模块

计算机视觉领域最早应用的注意力模块是由 Hu 等[23]提出的 SE（squeeze-and-excitation）模块，同时提取特征图的空间与通道信息，在视觉领域的各种任务中得到广泛应用。Woo 等[24]提出 CBAM（convolutional block attention module）模块，在 SE 模块的基础上，利用串联结构提取通道与空间信息。Zhang 等[25]则使用 shuffle 单元整合空间与通道信息，提出 SA-net，参照人类视觉系统自上而下的特点，将每个尺度的输出特征图进行融合。

但是，相比于其他视觉识别任务，篡改取证特征较难识别[26]。因此，本文提出改进的自顶向下扩张卷积注意力（TDDCA, top-down dilation convolutional attention）模块，利用扩张卷积显著增强特征提取能力。

1.3　损失函数

损失函数是深度学习模型训练中非常重要的部分，可在训练中集中于正确的特征集合。计算机视觉任务大多选择不同的损失函数。图像分类任务中常用的损失函数有交叉熵损失（CEL, cross entropy loss）函数[27]。目标检测任务中常用的损失函数有为解决类别不平衡问题的 focal loss[28]。图像识别领域（包括行人再识别、人脸识别等）中常用的损失函数有 CosFace[29]等。篡改检测领域大多采用交叉熵损失，但优化过程中没有关注向量的方向。为解决这个问题，本文设计余弦相似度损失作为辅助损失函数，更好地优化网络输出向量的方向，从而让网络可以更快更好地收敛到最优位置。

2　网络结构

本文提出一个端到端的多篡改检测与定位模型，输入一幅篡改图像，输出该图像篡改的置信度，并定位拼接和移除篡改区域。如图 1 所示，以 HRNet 和密集网络（DenseNet）为基线，提取图像 RGB 域和 SRM（steganalysis rich model）域双流特征信息，利用自顶向下扩张卷积注意力 TDDCA 模块融合双流多尺度特征，最后利用扩张卷积（MDC, mixed dilated convolution）模块[30]训练完成篡改检测和定位任务。

2.1　双分支主干网络

主干网络由高分辨率 HRNet 分支和 SRM 密集分支（SRM DB, SRM dense branch）组成，如图 1 所示，其中 HRNet 分支通过并行连接高分辨率到低分辨率卷积提取篡改图像 RGB 流特征。为最大限度地利用 SRM 流的信息，本文借鉴 DenseNet[31]设计思路，提出 SRM DB，利用密集连接 SRM 流信息，不仅增强篡改特征提取能力，而且避免网络传播过程中特征丢失，降低梯度消失的风险。SRM DB 中第 l 层的特征图 X_l 为

$$X_l = H_l([X_0, X_1, \cdots, X_{l-1}]) \tag{1}$$

其中，$[X_0,X_1,\cdots,X_{l-1}]$表示从第 0 层到第 $l-1$ 层特征图的通道连接，$H_l(\cdot)$表示第 l 层卷积模块。

本文选取 SRM DB 中 4、8、16 和 32 倍下采样的特征图与 HRNet 对应特征图进行通道连接，如图 1 通道连接模块所示。

2.2　自顶向下扩张卷积注意力模块

文献[26]提出的自顶向下的注意力网络，模仿人类视觉注意力特点。但是，受其感受野限制，在篡改检测与定位任务中存在缺陷。基于此，本文提出自顶向下扩张卷积注意力模块，如图 2 所示。在传统的注意力模块中加入扩张卷积模块，使其感受野更加丰富，并利用残差连接，在增强信息传递的同时还能避免梯度消失。本文对图像流和 SRM 流中 4 个尺度的特征图进行注意力，专注于篡改特征的提取，在多篡改检测任务上发挥更好的效果。

2.3　训练方法

多篡改检测和定位任务要求同时检测出多个

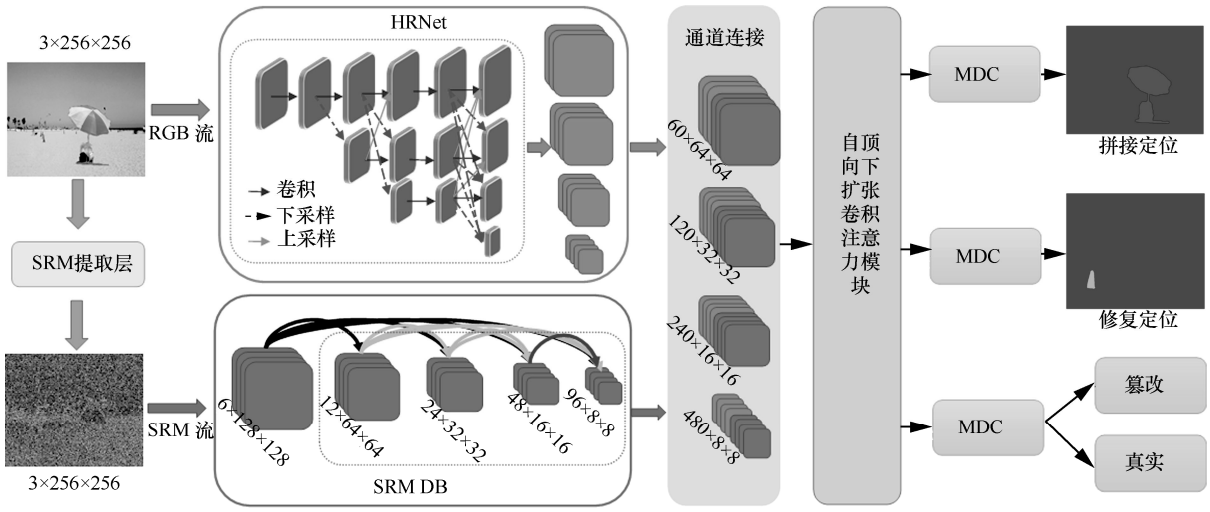


图 1　高分辨率扩张卷积注意力网络结构

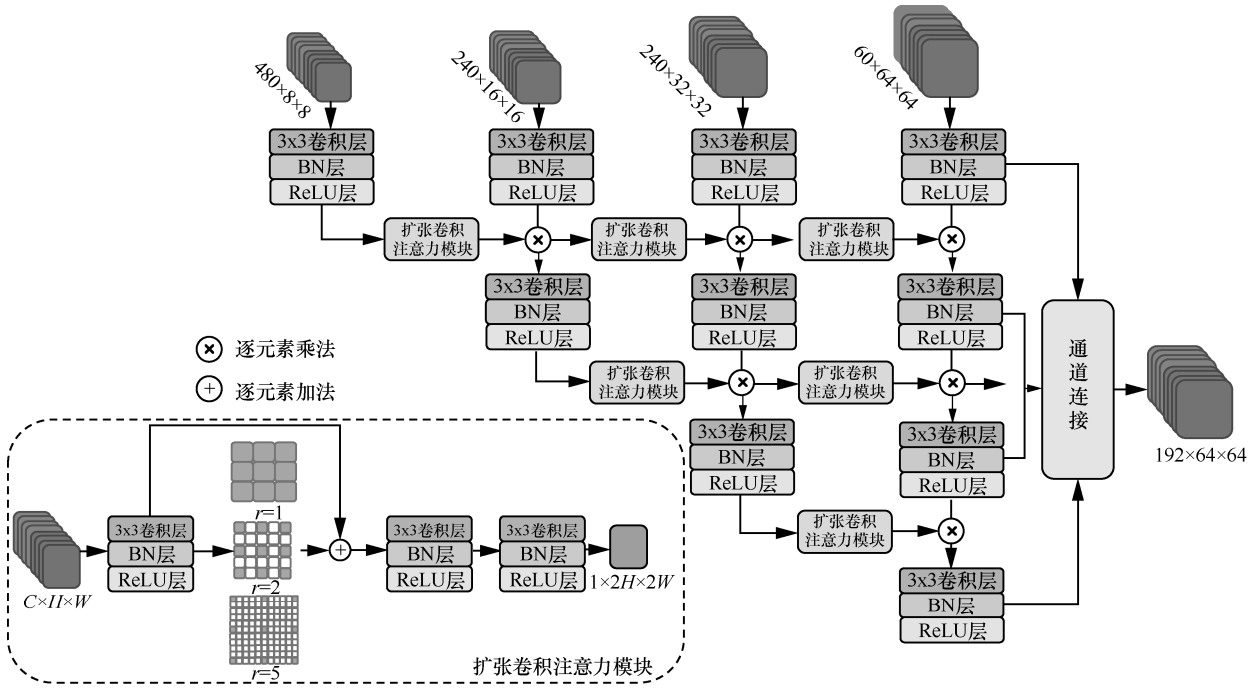


图 2　自顶向下扩张卷积注意力 TDDCA 模块

篡改类型。实验表明，不同篡改类型具有不同的篡改特征，为了能够让网络学习到不同类型所独有的特征，引入扩张卷积模块进行不同篡改任务的特征提取。

HRDA-Net 在检测模块训练时包含真实图像，但是在定位模块训练时，真实图像的加入会造成网络难以收敛的问题。为解决该问题，本文采用分步训练方式。篡改检测和定位任务的特征图基本一致，即可先训练定位模块后冻结参数，再单独训练篡改检测模块，可在不影响定位模块的基础上，让检测模块学会提取检测任务所需特征。因此，本文的训练步骤主要分为两步：第一步，使用自制拼接数据集对网络进行预训练后，采用 MM Dataset 训练拼接和移除篡改定位模块；第二步，将主干网络参数冻结，单独训练篡改检测模块。训练步骤如图 3 所示。

2.4　损失函数

主流篡改检测模型使用的损失函数通常不关注输出向量方向的优化。为解决该问题，本文提出余弦相似度函数作为辅助损失函数，在向量方向上进行优化，其目标是预测值 y_{pred} 和标签值 y_{label} 一致，即 y_{pred} 和 y_{label} 之间的夹角余弦值为 1。因此，本文设计余弦相似度损失函数 $L_{\cos}$ 如式(2)所示，取值范围为[0,1]。

$$L_{\cos}=1-\frac{y_{\text{label}}y_{\text{pred}}}{\|y_{\text{label}}\|_2\|y_{\text{pred}}\|_2} \tag{2}$$

其中，$\|\cdot\|_2$ 表示 L2 范式。

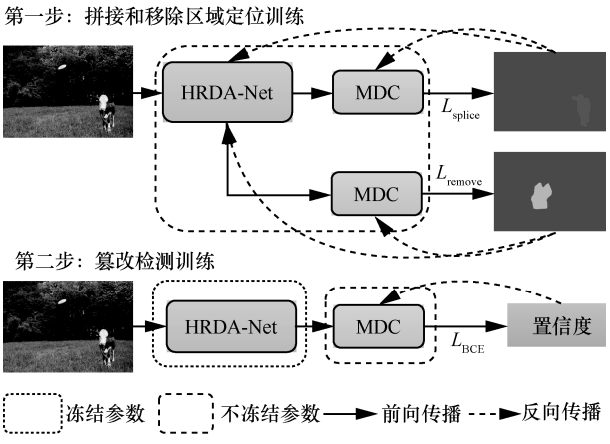


图 3　训练步骤

HRDA-Net 的训练主要分为 2 个步骤，每个步骤都有各自的损失函数，如图 3 所示。第一步为拼接和移除区域定位训练，包含 2 个并行的 MDC 模块，所以定位的损失函数 L_{loc} 包含拼接损失 L_{splice} 和移除损失 L_{remove}，如式(3)所示。

$$L_{\text{loc}}=\alpha L_{\text{splice}}+\beta L_{\text{remove}} \tag{3}$$

其中，权重 α=0.5 和 β=0.5 以保证拼接和移除定位

任务平等。L_{splice} 和 L_{remove} 如式(4)和式(5)所示。

$$L_{\text{splice}} = \gamma_1 L_{\text{BCE}} + \delta_1 L_{\cos} \tag{4}$$

$$L_{\text{remove}} = \gamma_2 L_{\text{BCE}} + \delta_2 L_{\cos} \tag{5}$$

其中，γ_1=γ_2=1.0，δ_1=δ_2=0.4（经多组实验选取的经验值，可加快收敛速度）。L_{BCE} 表示交叉熵损失函数，如式(6)所示。

$$L_{\text{BCE}} = -y_{\text{label}} \log\left(y_{\text{pred}}\right) - \left(1 - y_{\text{label}}\right) \log\left(1 - y_{\text{pred}}\right) \tag{6}$$

其中，y_{lable} 和 y_{pred} 分别为真实和预测标签。第二步单独训练篡改检测时，使用交叉熵 L_{BCE} 作为其损失函数。

3　实验结果

为验证 HRDA-Net 多篡改检测与定位性能，本文在 MM Dataset 进行消融和稳健性实验。在对比实验中，本文首先与主流语义分割模型进行对比，将多篡改定位作为 3 类语义分割任务，即将篡改图像分类为背景、拼接和移除区域。为验证 HRDA-Net 在单篡改定位任务上的性能，本文在 CASIA 与 NIST 数据集上进行微调，与主流单篡改定位模型进行对比。

3.1　MM Dataset 与实验数据集

本文构建 MM Dataset，真实图像从 COCO[22] 数据集中选取，包含 1 000 组真实图像和篡改图像，分辨率为 640 像素×480 像素，使用 Photoshop 工具对每一幅图像进行拼接和移除篡改。拼接操作下，篡改源与目标图像的光照、对比度、尺度等信息不同，因此，本文针对目标图像的像素分布特点，对拼接源区域进行了相应的后处理操作，包括亮度、对比度、尺度变换、目标翻转等。在移除篡改操作中，利用 Photoshop 仿制图章、修复画笔、修补工具等，根据待移除区域的像素分布特点选取相对应的移除区域，如图 4 所示，其中，篡改区域中黑色为背景区域，灰色为拼接区域，白色为移除区域。

深度学习模型训练需要有大量数据作为支撑，但 CASIA、NIST 和 MM Dataset 的数据量都不足以支撑模型训练。本文首先使用程序自动生成 15 000 幅拼接篡改图像对模型进行预训练。自制数据集从 COCO 数据集中随机选择两幅图像，对其中一幅图像使用 COCO 数据集提供的标注信息随机选取一个目标，并粘贴到另一幅图像。数据集使用情况如表 1 所示。

表 1　　数据集使用情况

数据集	篡改图像数量/幅	用途
自制拼接数据集	15 000	预训练
MM Dataset	800/200	微调/测试
CASIA v2.0	5 059	微调
CASIA v1.0	908	测试
NIST	404/160	微调/测试

3.2　度量评价和实验参数设置

为了全面量化模型的性能，本文选用 4 个指标进行评估，分别是准确率 precise、召回率 recall、F1 分数和误检率 fp。其中，F1 度量多篡改区域定位的精度，误检率 fp 度量真实像素被检测为篡改像素的比例，即“伪影”问题。计算式如式(7)～式(10)所示。

$$\text{precise} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{7}$$

(a) 真实图像　(b) 篡改源图像　(c) 篡改图像　(d) 篡改区域

图 4　多篡改数据集图像示例

表 2 模型有效性消融实验结果对比

HRNet	SRM DB	TDDCA	MDC	L_{cos}-定位	L_{cos}-检测	拼接-F1	移除-F1	fp	Accuracy
√	×	×	×	√	×	0.804	0.485	0.303	0.766
√	√	×	×	√	×	0.490	0.156	0.365	0.692
√	√	√	×	√	×	0.894	0.558	0.162	0.828
√	√	×	√	√	×	0.745	0.413	0.177	0.775
√	√	√	√	×	×	0.891	0.564	0.152	0.836
√	√	√	√	√	√	0.899	0.576	0.150	0.833
√	√	√	√	√	×	**0.899**	**0.576**	**0.150**	**0.850**

$$\text{recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{8}$$

$$\text{F1}=\frac{2\text{precise}\cdot\text{recall}}{\text{precise}+\text{recall}} \tag{9}$$

$$\text{fp}=\frac{\text{FP}}{\text{FP}+\text{TN}} \tag{10}$$

其中，TP 表示预测正确的篡改像素点数目，FP 表示预测错误的篡改像素点数目，FN 表示预测错误的真实像素点数目，TN 表示预测正确的真实像素点数目。本文采用正确率 Accuracy 指标对篡改检测分类结果进行评估，如式(11)所示。

$$\text{Accuracy}=\frac{\text{correct\_num}}{\text{image\_num}} \tag{11}$$

其中，correct_num 表示预测正确的图像数量，image_num 表示测试的图像总数。另外，本文与主流篡改定位方法对比，引入 AUC（area under curve）指标，即接收者操作特征（ROC, receive operating characteristic）曲线下方与坐标轴围成的面积。

本文实验环境是 Pytorch 1.8，Torchvision 的版本为 0.4，CPU 为 Intel I5-10400f，内存大小为 16 GB，显卡为 RTX2060 6G。在整个实验过程中，学习率始终设置为 1×10^{-4}。使用 SGD 作为损失函数，weight_decay 取值为 5×10^{-4}，moment 取值为 0.9。在进行模型预训练的时候，本文使用 HRNet 官方提供的语义分割预训练模型。

3.3 消融实验

1) 模型有效性实验

为验证 HRNet、SRM DB、TDDCA、MDC 和 L_{cos} 损失有效性，本节在 MM Dataset 上进行消融实验，如表 2 所示，其中，加粗字体表示性能最优值，√表示实验模型包含对应模块，×表示实验模型不包含对应模块。本文使用不同结构分支分别提取 RGB 流与 SRM 流特征，产生特征空间位置偏移，如表 2 第二行所示，直接进行双流特征融合后，实验结果较单流 HRNet 提取 RGB 特征相比，拼接-F1 降低 0.31，移除-F1 降低 0.33，误检率 fp 升高 0.06，Accuracy 降低 0.07。因此，为消除特征流空间位置偏移，本文提出 TDDCA+MDC 进行特征融合，实验结果如表 2 第三行和第四行所示。实验结果表明，采用 TDDCA+MDC 的特征融合，F1 性能最优，且 fp 最低。通过损失函数消融实验分析，在多篡改区域定位任务中，添加 L_{cos} 性能提升，但在多篡改检测任务中添加 L_{cos}，误检率 fp 保持不变，但由于过拟合 F1 性能稍微下降。因此，本文在多篡改检测任务损失函数设计中，仅采用交叉熵损失，没有添加余弦相似度损失。

2) 训练方法有效性实验

为验证本文提出的分步训练方法有效性，本节分别对多篡改定位单任务、多篡改检测单任务，分步训练是否冻结参数进行消融实验，实验结果如表 3 所示，其中，—表示没有相对应的实验数据。

表 3 训练方法有效性消融实验结果对比

训练方法	拼接-F1	移除-F1	fp	Accuracy
多篡改定位	0.899	0.576	0.150	—
多篡改检测	—	—	—	0.892
分步训练−参数不冻结	0.537	0.221	0.246	**0.913**
分步训练−参数冻结	**0.899**	**0.576**	**0.150**	0.850

实验结果表明，本文提出的多篡改定位和检测分步训练方法在冻结参数后，加入真实图像进行篡改检测任务训练时不影响多篡改定位性能，但较多篡改检测单任务 Accuracy 降低 0.02。若训练篡改检测任务时不进行参数冻结，受真实图像影响，拼接-F1 和移除-F1 分别下降 0.36 和 0.35，误检率 fp 升高

0.11，但篡改检测任务 Accuracy 提升 0.06。因此，本文选择分步训练时，冻结多篡改定位任务参数，保证多篡改定位和检测 2 个任务性能指标较优。

3.4　稳健性实验

真实场景下，篡改图像大多经过各类后处理操作，如网络传输压缩、噪声等。因此，抗后处理操作稳健性的篡改检测与定位框架尤其重要。本文设计 6 种后处理操作的稳健性实验，分别是 JPEG 压缩、高斯噪声、高斯模糊、亮度、对比度和色彩平衡，具体的参数设置如表 4 所示。

表 4　稳健性实验后处理操作及其参数设置

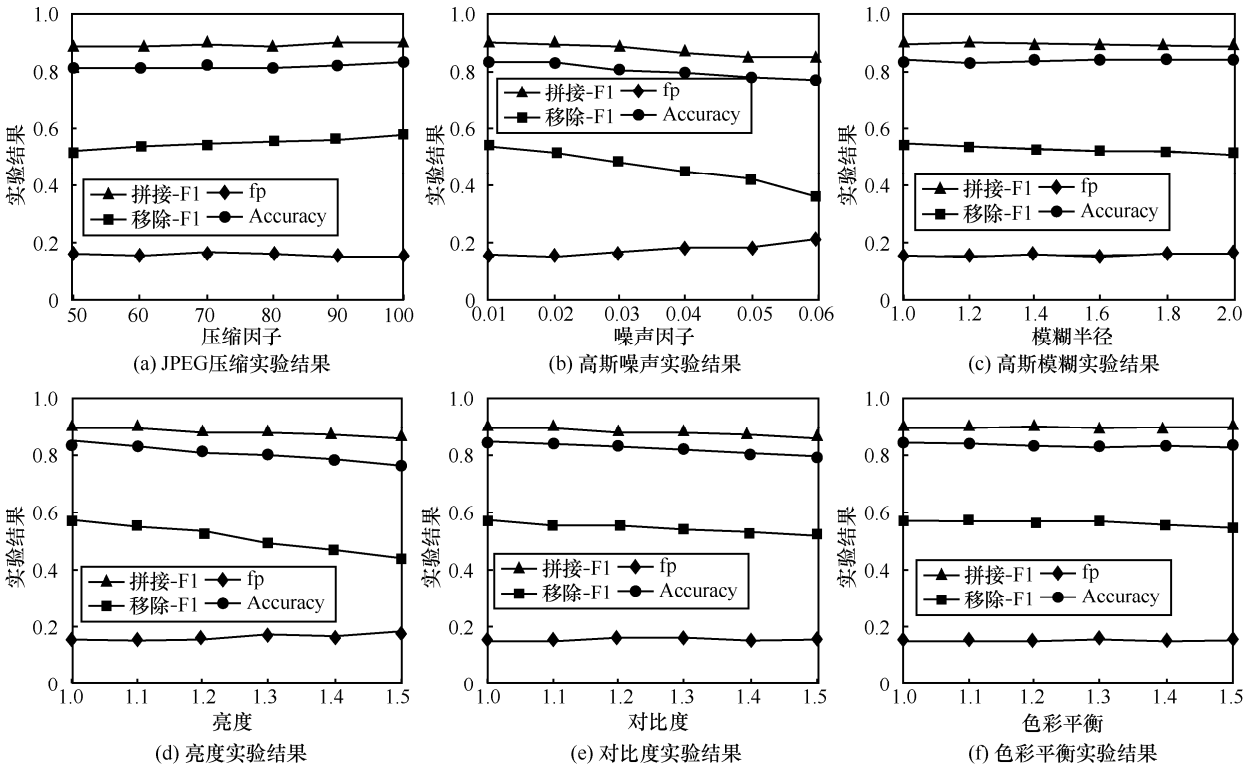
后处理手段	参数名称	参数取值
JPEG 压缩	压缩因子	{100,90,80,70,60,50}
高斯噪声	噪声参数	{0.06,0.05,0.04,0.03,0.02,0.01}
高斯模糊	半径	{1.0, 1.2,1.4,1.6,1.8,2.0}
亮度	—	{1.0, 1.1, 1.2, 1.3, 1.4, 1.5}
对比度	—	{1.0, 1.1, 1.2, 1.3, 1.4, 1.5}
色彩平衡	—	{1.0, 1.1, 1.2, 1.3, 1.4, 1.5}

本文在 MM Dataset 分别对 HRDA-Net 的篡改检测（Accuracy）和篡改定位（拼接-F1、移除-F1、fp）2 个任务进行实验，如图 5 所示。实验结果表明，在篡改定位和篡改检测任务中，HRDA-Net 在各个参数的后处理操作下，实验结果平稳，稳健性较好。移除篡改定位任务在高斯噪声和亮度后处理操作中，随着参数增大性能有所降低，但在 JPEG 压缩、高斯模糊、对比度和色彩平衡后处理操作中稳健性较优。

3.5　对比实验与分析

1) 语义分割模型对比实验

训练语义分割模型时，将多篡改任务当成一个三分类的语义分割任务，损失函数全部使用多分类的交叉熵损失。本次实验中与 HRDA-Net 进行对比的模型包含 FCN[32]、Deeplabv3[33]、PSPNet[34]、DANet[35]、RRU-net[7]和 HRNet[17]，实验结果如表 5 所示，其中，加粗字体表示最高值。在拼接篡改定位、移除篡改定位和篡改检测任务中，HRDA-Net 相较其他模型有明显优势，除拼接篡改定位任务的 precise 指标低于 FCN[32]，本文 HRDA-Net 的 precise、recall 和 F1 均最优。由此，HRDA-Net 相对于传统语义分割模型更加适合于多篡改检测与定位任务。另外，相较于拼接篡改定位，移除篡改定位分数较低，一方面是因为在预训练只使用拼接篡改图像，另一方面是移除篡改的特征相对于拼接篡改来说更难提取。HRDA-Net 在 MM Dataset 的实验结果示例如图 6 所



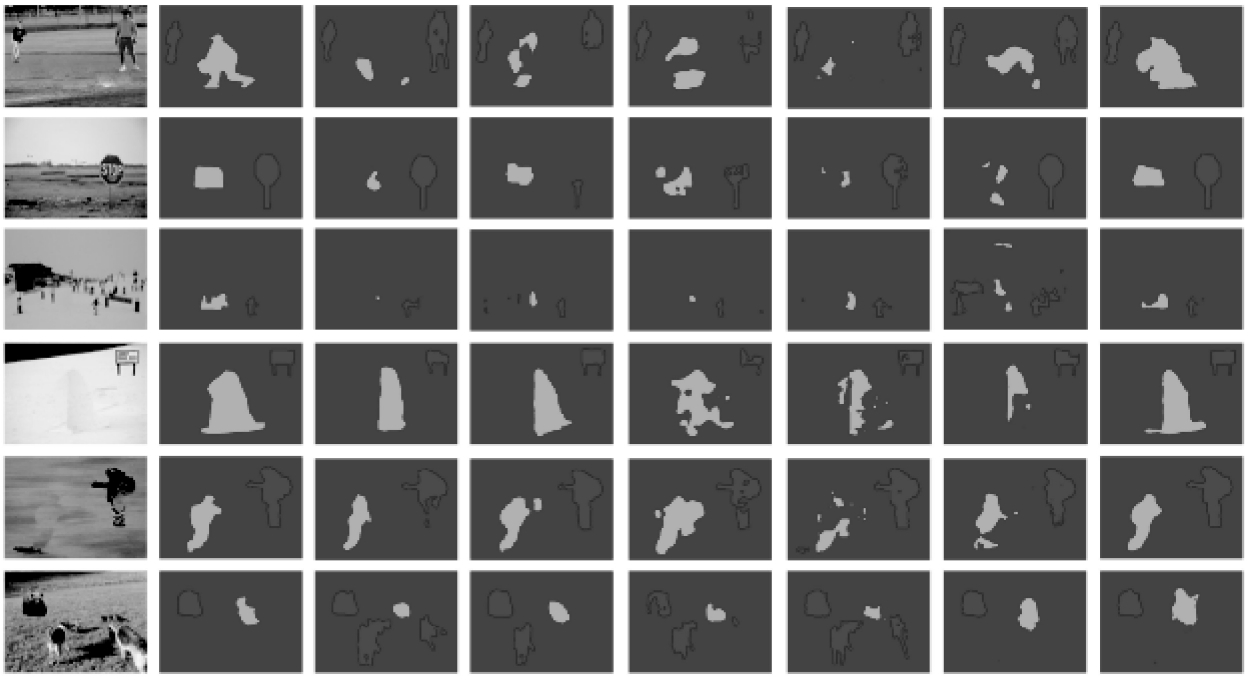
(a) JPEG压缩实验结果　(b) 高斯噪声实验结果　(c) 高斯模糊实验结果

(d) 亮度实验结果　(e) 对比度实验结果　(f) 色彩平衡实验结果

图 5　稳健性实验结果

表 5　　HRDA-Net 与主流语义分割模型对比实验结果

模型名称	拼接篡改定位			移除篡改定位		
	precise	recall	F1	precise	recall	F1
FCN	**0.962**	0.616	0.636	0.580	0.225	0.305
Deeplabv3	0.831	0.727	0.770	0.760	0.385	0.507
PSPNet	0.808	0.685	0.734	0.581	0.327	0.407
DANet	0.718	0.799	0.751	0.717	0.228	0.344
RRU-Net	0.690	0.793	0.727	0.457	0.224	0.286
HRNet	0.867	0.768	0.804	0.700	0.388	0.485
HRDA-Net	0.888	**0.913**	**0.899**	**0.764**	**0.481**	**0.576**



(a) 篡改图像　(b) Ground-Truth　(c) FCN　(d) Deeplabv3　(e) PSPNet　(f) DANet　(g) RRU-Net　(h) HRDA-Net

图 6　HRDA-Net 在 MM Dataset 的实验结果示例

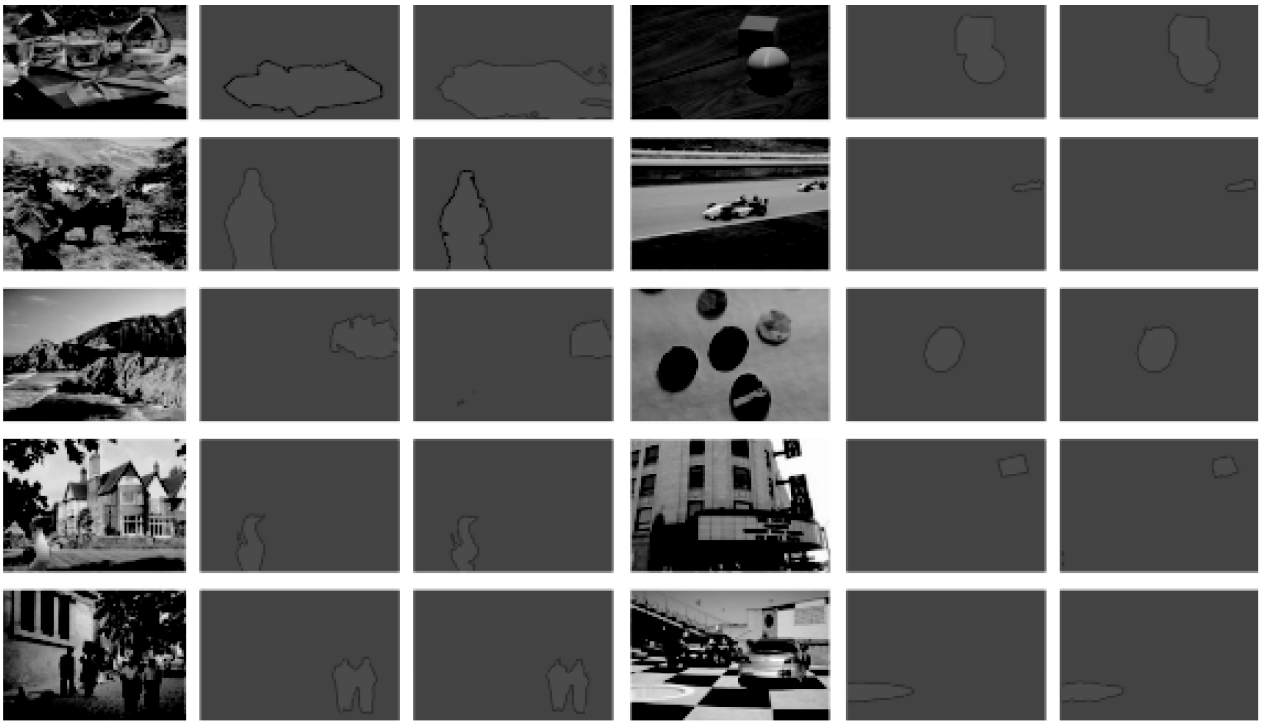
示，其中，黑色表示背景区域，灰色表示拼接篡改区域，白色表示移除篡改区域。

2) 单篡改定位模型对比实验

本文在 CASIA 与 NIST 数据集上与主流单篡改定位模型进行对比，实验结果如表 6 所示，其中，加粗部分表示相应列中的最高指标，—表示在发表的论文中没有相对应的数据。实验结果证明，HRDA-Net 在 NIST 数据集上的 F1 和 AUC 分数均最优，其中 F1 和 AUC 分数分别比次优 SEINet 高了近 6%和 1.3%；在 CASIA 数据集上的 F1 分数达到最优，比 SEINet 高 0.8%，且 AUC 分数与 GSCNet 持平。由此证明，HRDA-Net 泛化性较好，进行单篡改定位任务时性能仍较优秀。HRDA-Net 在 CASIA 和 NIST 数据集上的可视化结果如图 7 所示，其中，黑色表示背景区域，灰色表示拼接篡改区域。

表 6　CASIA 和 NIST 数据集中单篡改定位对比实验

模型	F1		AUC	
	NIST	CASIA	NIST	CASIA
NoI	0.285	0.263	0.487	0.612
CFA	0.174	0.207	0.501	0.522
RGB-N	0.722	0.408	0.937	0.795
LSTM-En	—	0.391	0.793	0.762
GSCNet	0.837	0.471	0.917	**0.833**
SEINet	0.891	0.488	0.980	0.801
HRDA-Net	**0.951**	**0.496**	**0.993**	**0.833**



(a) 篡改图像 (CASIA)　(b) Ground-Truth (CASIA)　(c) HRDA-Net (CASIA)　(d) 篡改图像 (NIST)　(e) Ground-Truth (NIST)　(f) HRDA-Net (NIST)

图 7　HRDA-Net 在 CASIA 和 NIST 数据集上的可视化结果

4　结束语

针对真实场景下图像篡改通常包含多类篡改操作问题，本文提出 MM Dataset 及多篡改检测和定位模型 HRDA-Net。对比实验结果证明，HRDA-Net 模型具有较强的篡改检测和定位性能，并且在 6 种后处理操作中都具有较好的稳健性和泛化性。本文是多篡改取证任务的一次初步尝试，所提出的 MM Dataset 目前只包含拼接和移除 2 种篡改手段。在今后的工作中，作者将会继续完善，包含更多篡改操作，如复制–粘贴等，并提出更加具有泛化性以及拥有更强检测性能的多篡改检测与定位算法。

参考文献：

[1] 乔通, 姚宏伟, 潘彬民, 等. 基于深度学习的数字图像取证技术研究进展[J]. 网络与信息安全学报, 2021, 7(5): 13-28.
QIAO T, YAO H W, PAN B M, et al. Research progress of digital image forensic techniques based on deep learning[J]. Chinese Journal of Network and Information Security, 2021, 7(5): 13-28.

[2] 田秀霞, 李华强, 张琴, 等. 基于双通道 R-FCN 的图像篡改检测模型[J]. 计算机学报, 2021, 44(2): 370-383.
TIAN X X, LI H Q, ZHANG Q, et al. Dual-channel R-FCN model for image forgery detection[J]. Chinese Journal of Computers, 2021, 44(2): 370-383.

[3] 张旭, 胡晰远, 陈晨, 等. 基于透视投影下空间光照一致性分析的图像拼接篡改检测[J]. 自动化学报, 2019, 45(10): 1857-1869.
ZHANG X, HU X Y, CHEN C, et al. Image splicing detection based on spatial lighting consistency analysis under perspective projection[J]. Acta Automatica Sinica, 2019, 45(10): 1857-1869.

[4] RAO Y, NI J Q. A deep learning approach to detection of splicing and copy-move forgeries in images[C]//Proceedings of 2016 IEEE International Workshop on Information Forensics and Security (WIFS). Piscataway: IEEE Press, 2016: 1-6.

[5] ROTA P, SANGINETO E, CONOTTER V, et al. Bad teacher or unruly student: can deep learning say something in Image Forensics analysis?[C]//Proceedings of 2016 23rd International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2016: 2503-2508.

[6] LIU B, PUN C M. Locating splicing forgery by fully convolutional networks and conditional random field[J]. Signal Processing: Image Communication, 2018, 66: 103-112.

[7] BI X L, WEI Y, XIAO B, et al. RRU-net: the ringed residual U-net for image splicing forgery detection[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2019: 1-10.

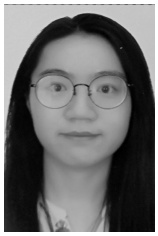
[8] 王珠珠. 基于 U 型检测网络的图像篡改检测算法[J]. 通信学报, 2019, 40(4): 171-178.
WANG Z Z. Image forgery detection algorithm based on U-shaped detection network[J]. Journal on Communications, 2019, 40(4): 171-178.

[9] CHEN B J, TAN W J, COATRIEUX G, et al. A serial image copy-move forgery localization scheme with source/target distinguishment[J]. IEEE Transactions on Multimedia, 2021, 23: 3506-3517.

[10] ZHU Y, CHEN C F, YAN G, et al. AR-net: adaptive attention and residual refinement network for copy-move forgery detection[J]. IEEE

Transactions on Industrial Informatics, 2020, 16(10): 6714-6723.
[11] LI H D, HUANG J W. Localization of deep inpainting using high-pass fully convolutional network[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2019: 8301-8310.
[12] LIU Y Q, GUAN Q X, ZHAO X F, et al. Image forgery localization based on multi-scale convolutional neural networks[C]//Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. New York: ACM Press, 2018: 85-90.
[13] BAPPY J H, SIMONS C, NATARAJ L, et al. Hybrid LSTM and encoder-decoder architecture for detection of image forgeries[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2019, 28(7): 3286-3300.
[14] ZHOU P, HAN X T, MORARIU V I, et al. Learning rich features for image manipulation detection[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 1053-1061.
[15] WU Y, ABDALMAGEED W, NATARAJAN P. ManTra-net: manipulation tracing network for detection and localization of image forgeries with anomalous features[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 9543-9552.
[16] HU X F, ZHANG Z H, JIANG Z Y, et al. SPAN: spatial pyramid attention network for image manipulation localization[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2020: 312-328.
[17] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 5693-5703.
[18] DONG J, WANG W, TAN T N. CASIA image tampering detection evaluation database[C]//Proceedings of 2013 IEEE China Summit and International Conference on Signal and Information Processing. Piscataway: IEEE Press, 2013: 422-426.
[19] GUAN H Y, KOZAK M, ROBERTSON E, et al. MFC datasets: large-scale benchmark datasets for media forensic challenge evaluation[C]//Proceedings of 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW). Piscataway: IEEE Press, 2019: 63-72.
[20] WEN B H, ZHU Y, SUBRAMANIAN R, et al. COVERAGE—a novel database for copy-move forgery detection[C]//Proceedings of 2016 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2016: 161-165.
[21] MAHFOUDI G, TAJINI B, RETRAINT F, et al. DEFACTO: image and face manipulation dataset[C]//Proceedings of 2019 27th European Signal Processing Conference (EUSIPCO). Piscataway: IEEE Press, 2019: 1-5.
[22] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//European Conference on Computer Vision. Berlin: Springer, 2014: 740-755.
[23] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.
[24] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//European Conference on Computer Vision. Berlin: Springer, 2018: 3-19.
[25] ZHANG Q L, YANG Y B. SA-net: shuffle attention for deep convolutional neural networks[C]//Proceedings of 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2021: 2235-2239.
[26] PANG B, LI Y Z, LI J F, et al. TDAF: top-down attention framework for vision tasks[J]. arXiv Preprint, arXiv: 2012.07248, 2020.
[27] SHORE J, JOHNSON R. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy[J]. IEEE Transactions on Information Theory, 1980, 26(1): 26-37.
[28] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2980-2988.
[29] WANG H, WANG Y T, ZHOU Z, et al. CosFace: large margin cosine loss for deep face recognition[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 5265-5274.
[30] HW A, CL A, ML A, et al. Optimized HRNet for image semantic segmentation[J]. Expert Systems with Applications, 2020, 174: 114532.
[31] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 4700-4708.
[32] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2015: 3431-3440.
[33] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv Preprint, arXiv:1706.05587, 2017.
[34] ZHAO H S, SHI J P, QI X J, et al. Pyramid scene parsing network[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 2881-2890.
[35] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 3146-3154.
[36] MAHDIAN B, SAIC S. Using noise inconsistencies for blind image forensics[J]. Image and Vision Computing, 2009, 27(10):1497-1503.
[37] FERRARA P, BIANCHI T, ROSA A D, et al. Image forgery localization via fine-grained analysis of CFA artifacts[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(5):1566-1577.
[38] SHI Z N, SHEN X J, CHEN H P, et al. Global semantic consistency network for image manipulation detection[J]. IEEE Signal Processing Letters, 2020, 27: 1755-1759.
[39] ZHU Y, QI N, GUO Y, et al. SEINet: semantic-edge interaction network for image manipulation localization[C]//Proceedings of the fourth Chinese Conference on Pattern Recognition and Computer Vision. [S.l.:s.n.], 2021: 1-13.

[作者简介]



朱叶（1989− ），女，山东菏泽人，博士，河北工业大学讲师、硕士生导师，主要研究方向为图像安全取证、图像处理与模式识别。

余宜林（1998− ），男，福建南平人，河北工业大学硕士生，主要研究方向为图像安全取证。

郭迎春（1970− ），女，河北张家口人，博士，河北工业大学副教授、硕士生导师，主要研究方向为图像处理与模式识别、人工智能等。

收录声明

本刊对发表的文章，拥有出版电子版、网络版版权，并拥有和其他网站交换信息的权利。本刊支付的稿酬中已经包含上述费用。

Journal on Communications has the copyright to publish electronic edition, online edition of the published articles, and has the right to exchange information with other sites. The expenses have been included in the fee paid by editorial department.

道德声明

本刊发表的论文是作者独立取得的原创性研究成果，无一稿多投；论文内容不涉及国家机密；未曾以任何形式用任何文种在国内外公开发表过；论文内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题，论文作者将承担全部责任。

The authors of *Journal on Communications* guarantee that their submitted articles are original and contain nothing confidential. The said article is only submitted to *Journal on Communications*. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.

通信学报

Journal on Communications

发行代号：国内2-676 / 国外M395　2022年1月25日出版　定价：98.00元

ISSN 1000-436X

9 771000 436229 01>