

袁国龙, 张玉金. 基于多层次特征优化的图像拼接篡改取证网络[J]. 智能计算机与应用, 2024, 14(9): 99-103. DOI: 10.20169/j. issn. 2095-2163. 240915

基于多层次特征优化的图像拼接篡改取证网络

袁国龙, 张玉金

(上海工程技术大学 电子电气工程学院, 上海 201620)

摘要: 随着数字化的普及, 数字图像篡改现象日益普遍, 给社会公正和诚信带来了巨大的挑战。针对目前图像拼接篡改取证网络中存在的篡改区域定位精度低的问题, 本文提出一种基于多层次特征优化的图像拼接篡改取证网络。采用双流编-解码器结构, 使用 SoftPool 池化方法减少编码器阶段边缘等重要特征的丢失; 同时, 将深、浅层的特征进行融合, 并引入 SE 注意力机制进行重要特征筛选, 使网络各层能够提取更优的篡改特征信息, 更准确地定位图像篡改区域。实验结果表明, 本文提出的取证网络在检测性能上优于当前主流的图像拼接篡改取证网络, 能更准确地定位拼接图像的篡改区域。

关键词: 拼接篡改取证; 篡改区域定位; 编-解码器; 特征优化; 注意力机制

中图分类号: TN929.5

文献标志码: A

文章编号: 2095-2163(2024)09-0099-05

Image splicing forensics network based on multi-level feature optimization

YUAN Guolong, ZHANG Yujin

(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: With the popularization of digital technology, the phenomenon of digital image tampering is becoming more and more common, which brings great challenges to social justice and integrity. Aiming at the problem of low positioning accuracy of tampered areas in the current image splicing forgery forensics network, an image splicing forgery forensics network based on multi-level feature optimization is proposed. The network adopts a dual-stream encoder-decoder structure, and uses the SoftPool pooling method to reduce the loss of important features such as the edges of the encoder stage. At the same time, the deep and shallow features are fused, and the SE (Squeeze-and-Excitation) attention mechanism is introduced to screen important features, so that each layer of the network can extract better tampering feature information and more accurately locate image tampering areas. Experimental results show that the proposed forensic network is superior to the current mainstream image splicing forgery forensics networks in terms of detection performance, and can more accurately locate the tampered area of the splicing image.

Key words: splicing tamper forensics; localization of tampered regions; encoder-decoder; feature optimization; attention mechanism

0 引言

当今信息时代, 图像作为一种重要的信息载体, 其篡改和伪造现象越来越普遍, 给图像的可信度和可靠性带来了严重威胁, 图像篡改取证技术的研究和应用日益重要。

拼接图像的篡改区域来自于其他图像, 所以真实区域和篡改区域往往会有不同的手工特征, 这些特征差异是拼接图像检测和篡改区域定位的重要依据。传统的基于手工特征的拼接方法主要检测噪声、纹理等单一特征, 泛化性和检测效果不佳^[1-2]。

基于深度学习的取证算法通过深度模型训练自动学习期望的特征, 无需提取显式特征, 逐渐成为图像篡改取证领域的主流方法。Cun 等^[3]认为拼接区域不仅与局部特征相关, 还与全局特征高度相关, 提出了基于半全局网络 and 全连通条件随机场的图像拼接伪造定位算法; Liu 等^[4]提出了一个基于卷积神经网络 (Convolutional Neural Network, CNN) 框架的拼接图像取证网络, 通过构建新的融合网络来检测大尺寸图像的真实区域和篡改区域的属性差异, 但该网络对小尺寸图像检测失效; Xiao 等^[5]提出一种级联卷积神经网络和自适应聚类算法相结合的图像

基金项目: 上海市自然科学基金项目 (17ZR1411900)。

作者简介: 袁国龙 (1997-), 男, 硕士研究生, 主要研究方向: 图像处理, 图像篡改取证。Email: yuang9826@163.com

收稿日期: 2023-05-18

哈尔滨工业大学主办 ◆ 系统开发与应用

拼接检测框架,引入图像级 CNN 网络以降低模型的时间复杂度,但仅能粗略定位篡改区域;Zhou 等^[6]使用空域富模型(Steganalysis Rich Model, SRM)隐写滤波器提取图像噪声,并将图像噪声特征和 RGB 像素篡改特征输入到 Faster RCNN 网络中检测篡改区域,但该网络只能对篡改区域进行框定位;Ahmed 等^[7]通过训练一个受监督的 Mask-RCNN 网络来学习图像拼接篡改所产生的层次特征,在原始网络引入残差结构和迁移学习策略以提高网络检测性能;吴鹏等^[8]提出了一个双流 Fast R-CNN 的拼接图像篡改网络,增加一个全卷积分支以进行像素级定位,并采用错误等级分析噪声模型替代传统的富隐写分析模型,用于提取噪声信息。为了进一步提高篡改区域定位精度,Ding 等^[9]提出了一种双通道 U 型网络模型,将高通滤波器提取的残差图像与原始篡改图像相结合,以提取更多的篡改特征;Mazumdar^[10]采用双流编-解码器网络,同时检测 RGB 像素和噪声层面的篡改特征,将检测到的多种特征进行融合后定位篡改区域,一定程度上取得了良好的检测效

果。

以上网络虽然在拼接篡改检测中取得一定的成果,但还存在拼接篡改检测精度低、篡改区域边缘定位效果差的问题。为此,在文献[10]的基础上,本文提出了一种基于多层次特征优化的图像拼接篡改取证网络,在双流编-解码器中使用 SoftPool 池化方法,减少边缘等重要特征的丢失,将对应的深、浅层特征信息相融合并在解码器阶段引入 SE 注意力机制筛选重要特征信息,使网络各个层提取更优的篡改特征,以提升取证网络的检测性能。

1 网络模型架构

本文提出的网络模型以双流编-解码器结构为基础,其中两个流使用相同结构的编-解码器,分别检测 RGB 像素上的边缘等高阶特征信息和局部噪声不一致性等低阶特征信息,网络架构如图 1 所示,其中高通滤波层(High-Pass Filter, HPF)是约束卷积层,用于提取原始图像的噪声信息,作为噪声流的输入。

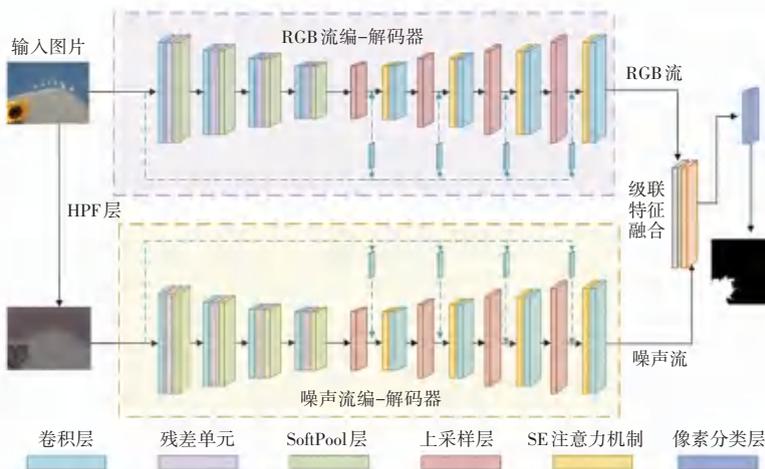


图 1 网络架构

Fig. 1 Network framework

1.1 Softpool 池化模块

为了减少网络池化过程中重要特征信息的损失,Stergiou 等^[11]提出了使用 Softmax 函数加权激活的池化方法 SoftPool。SoftPool 是一种快速有效的指数加权下采样方法,使用区域特征的 Softmax 函数进行加权池化,能够保持特征的表达性并且可微操作。以二维图像为例,SoftPool 池化计算方式如图 2 所示,其中单次池化计算区域 R 与权重矩阵 w 保持同样的维度大小。本文在 CNN 网络中引入 SoftPool,以减少池化过程中边缘等重要特征信息的丢失。

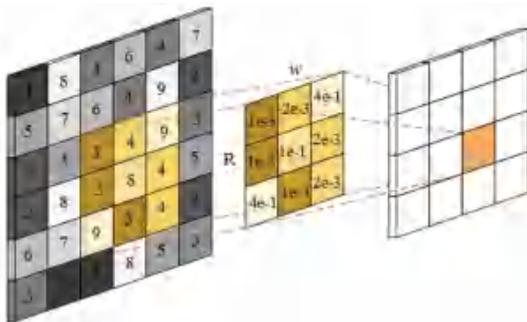


图 2 SoftPool 池化计算方式

Fig. 2 SoftPool pooling calculation method

每个位置所分配的权重为该位置特征值的 Softmax 激活输出与区域 R 内所有位置特征值的 Softmax 激活输出之和的比值。 w_i 为位置 i 对应的权重:

$$w_i = \frac{e^{a_i}}{\sum_{j \in R} e^{a_j}} \quad (1)$$

其中, e 为自然指数, a_i, a_j 为位置 i, j 的输入特征值。

SoftPool 池化操作的输出值为区域 R 内所有位置上特征值与对应权重的乘积的和, \tilde{a} 为区域 R 经过 SoftPool 池化计算的输出值:

$$\tilde{a} = \sum_{i \in R} w_i \times a_i \quad (2)$$

1.2 特征信息补充模块

在编-解码器中,解码器主要关注深层局部特征的提取,忽略了浅层全局特征在篡改区域检测中的重要性。为了使网络提取拼接图像中具有篡改鉴别性的特征信息,如边缘等,本文设计了特征信息补充模块,将浅层网络的低层次特征映射和深层网络中高层次特征映射相融合,从而使网络模型提取更加丰富的篡改区域特征信息。这一过程中,4个原始图像下采样层的卷积核大小和步长分别为 $8/8, 4/4, 2/2$ 和 $1/1$ 。

1.3 SE 注意力模块

编码器阶段提取到的特征是粗特征。同时,特征信息补充模块也会给解码器带来冗余特征信息,不利于后续网络对图像进行篡改检测和篡改区域定位。为了解决这个问题,本文在编-解码器网络中引入 SE(Squeeze-and-Excitation) 注意力机制进行特征筛选,其结构如图3所示。

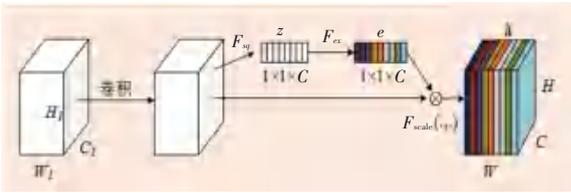


图3 SE 注意力机制结构

Fig. 3 Structure of SE attention mechanism

对于输入特征 $X = [x^1, x^2, \dots, x^{C_1}] \in \mathbb{R}^{H_1 \times W_1 \times C_1}$ 和滤波器核 $V = [v_1, v_2, \dots, v_C]$, U 是中间特征图, $U = [u_1, u_2, \dots, u_C] \in \mathbb{R}^{H \times W \times C}$, 特征 X 的变换如下:

$$u_c = X * v_c = \sum_{i=1}^{C_1} x^i * v_c^i \quad (3)$$

其中, $*$ 表示卷积, $v_c = [v_c^1, v_c^2, \dots, v_c^{C_1}]$ 。

由于卷积层提取局部特征而忽略了全局特征,

所以 SE 注意力机制使用挤压 (F_{sq}) 和激励操作 (F_{ex}) 对滤波器响应进行校准。挤压操作使用全局平均池化在网络中嵌入全局信息,通过空间维度 $H \times W$ 收缩中间特征 U , 生成统计量 $z = [z_1, z_2, \dots, z_C] \in \mathbb{R}^C$, 如下式所示:

$$z_c = F_{sq}(u_c) = \frac{1}{H \times W} \sum_{m=1}^H \sum_{n=1}^W u_c(m, n) \quad (4)$$

激励操作作用于自适应地重新校准每个通道的响应,采用带有 Sigmoid 激活的门控机制,如下式所示:

$$e = F_{ex}(z, W) = \sigma(g(z, W)) = \sigma(W_2 \text{ReLU}(W_1 z)) \quad (5)$$

其中, σ 表示 Sigmoid 激活计算, g 表示用 ReLU 激活函数实现的门控机制,使用两个全连接层来形成门控机制。

最后,SE 注意力机制的输出表示为 $\tilde{X} = [\tilde{x}^1, \tilde{x}^2, \dots, \tilde{x}^C]$, 通过激活 e 来重新缩放 U , 如下式所示:

$$\tilde{x}^c = F_{scale}(u_c, e_c) = e_c u_c \quad (6)$$

其中, $e = [e_1, e_2, \dots, e_C]$, $F_{scale}(u_c, e_c)$ 表示 e_c 和 u_c 之间的通道乘法。

根据拼接图像篡改区域定位任务,其更加关注重要特征并抑制不太有用的特征,使取证网络能够更好地区分图像拼接区域。

1.4 编-解码器模块

在编码器中有4个下采样块,每个下采样块包含一个卷积层、一个残差传播层和一个 SoftPool 层组成。每个下采样块中的卷积层和残差传播层具有相同的滤波器个数。4个下采样块输出的特征图大小依次为 $128 \times 128 \times 32, 64 \times 64 \times 64, 32 \times 32 \times 128$ 和 $16 \times 16 \times 256$ 。

与编码器结构相对应,解码器由4个上采样块组成,每个上采样块由一个卷积层、一个 SE 注意力机制和一个上采样因子为2的上采样层组成。4个上采样块输出的特征图大小依次为 $32 \times 32 \times 64, 64 \times 64 \times 64, 128 \times 128 \times 32$ 和 $256 \times 256 \times 32$ 。最后两个流输出的密集特征图进行级联融合,生成 $256 \times 256 \times 64$ 大小的混合特征图。最后,混合特征图输送到一个卷积核大小为 1×1 、带有 Sigmoid 激活函数的像素预测卷积层,生成相关像素预测概率表,最终得到预测二值图。

2 实验结果与分析

2.1 数据准备

2.1.1 预训练数据集

在网络预训练方面, Hu 等^[12] 将 COCO 数据集

图像的局部区域拼接到 DRESDEN 和 NIST16 数据集的真实图片中,创建了约 27 000 张图像的自制合成数据集。本文使用其中 13 000 张拼接图像和对应的二值图作为预训练集,其中 90%的图像用于训练,10%用于测试。

2.1.2 性能验证数据集

本文使用 Columbia 和 CASIA V2.0 两个公共标准数据集来验证网络性能。

(1)Columbia 数据集有 180 张未压缩的拼接伪造彩色图像,没有经过任何的后处理操作。本文取其中 135 张图像用于训练,45 张图像用于测试。

(2)CASIA V2.0 数据集有 5 123 张彩色伪造图像,包含拼接和复制-粘贴两种篡改类型。本文取其中 715 张拼接图像用于训练,100 张拼接图像用于测试。

为了使网络更有效地学习不同篡改类型的相关特征,并减少过拟合,对相关训练集上的图像进行随机水平或垂直翻转、图像 180°旋转和随机裁剪等数据增强操作。

2.2 实验环境和评价指标

本文使用单个 NVIDIA TATAN RTX GPU 训练网络模型,在 Win10 系统上使用 TensorFlow 定义深度学习框架。在训练过程中每个批次取 8 张图像及对应 Ground-Truth 图,设置阈值为 0.5,使用 Adam 优化器将损失最小化,初始学习率设为 0.000 05。本文选择精确率(Precision)、召回率(Recall)以及 F1 分数作为衡量网络性能的评估指标。

2.3 实验结果分析

2.3.1 对比试验结果

为了验证本文提出的网络在拼接图像篡改取证工作中的性能,将本文提出的网络与现存的主流取证网络 Matra-Net^[13]、C2R-Net^[14]、TED-Net^[10] 在 Columbia 和 CASIA V2.0 数据集上的实验结果做对比实验,结果见表 1 和表 2。

表 1 在 Columbia 数据集上实验结果对比

Table 1 Comparison of experimental results on Columbia dataset

对比网络	Precision	Recall	F1
Matra-Net ^[13]	0.716	0.549	0.584
C2R-Net ^[14]	0.804	0.612	0.695
TED-Net ^[10]	0.896	0.793	0.841
本文网络	0.931	0.840	0.883

表 2 在 CASIA V2.0 数据集上实验结果对比

Table 2 Comparison of experimental results on CASIA V2.0 dataset

对比算法	Precision	Recall	F1
Matra-Net ^[13]	0.631	0.673	0.651
C2R-Net ^[14]	0.581	0.808	0.676
TED-Net ^[10]	0.609	0.704	0.653
本文网络	0.636	0.735	0.682

在相关数据集上的实验结果比较中,本文提出的网络的精确率、召回率和 F1 值都有着最优的表现。本文提出的网络在 Columbia 数据集上的 F1 值比 C2R-Net、TED-Net 网络分别高出了 18.8 和 4.2 个百分点,在 CASIA V2.0 数据集上的 F1 值也分别高出了 1.4 和 2.9 个百分点。在 Columbia 数据集上,C2R-Net 的精确率和召回率比本文提出网络分别低了 12.7 和 22.8 个百分点,这是因为 C2R-Net 使用聚类方法对网络检测结果进行后处理,这仅可以对篡改区域进行粗略定位,使得定位精度较差。综上,本文提出的网络在图像拼接篡改检测性能上优于主流的网络。

2.3.2 消融实验

本文在 Columbia 和 CASIA V2.0 数据集上进行了消融实验,分别验证本文网络中各模块对于网络检测性能提升的有效性,相关实验结果见表 3。其中,Base 指原始的双流编-解码器网络,Base-SP、Base-SP-SC 和 Base-SP-SC-SE 表示在 Base 模型中逐次引入 SoftPool 池化模块、特征信息补充模块和 SE 注意力机制,其中 Base 网络池化层使用最大池化方法进行池化计算。显然,随着网络相关模块的使用,网络在 Columbia 数据集上的 F1 值逐步增高,各模块均能有效提升网络的拼接篡改检测性能。

表 3 在相关数据集上的消融实验 F1 值结果

Table 3 Results of ablation experiments F1 values on related datasets

模型	Columbia	CASIA V2.0
Base	0.841	0.652
Base-SP	0.852	0.659
Base-SP-SC	0.866	0.657
Base-SP-SC-SE	0.883	0.682

2.3.3 篡改区域定位结果

为了进一步验证本文提出的网络在图像篡改区域定位中的有效性,在相关数据集上与对比网络的篡改区域定位结果进行了比较,定位结果如图 4 所示。实验结果显示,本文提出的网络能够更加准确

地定位图像篡改区域,更注重篡改区域的边缘细节信息,具有更好的篡改区域定位能力。

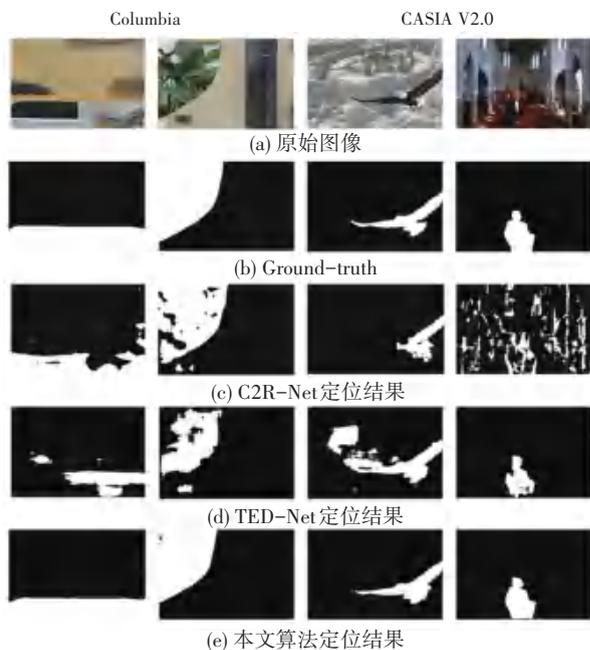


图4 在各数据集上的定位结果对比

Fig. 4 Comparison of positioning results on each dataset

3 结束语

本文提出了一种基于多层次特征优化的图像篡改取证网络,用于拼接图像的篡改检测和篡改区域像素级定位。在相关拼接数据集上的实验结果表明,本文提出的取证网络在检测性能上优于当前流行的图像拼接篡改取证网络,同时能更准确地定位拼接图像的篡改区域。未来可以进一步完善其网络结构,使其能够更高效检测复合篡改类型图像。

参考文献

[1] ZHU N, LI Z. Blind image splicing detection via noise level function[J]. Signal Processing: Image Communication, 2018, 68: 181-192.
 [2] RHEE K H. Detection of spliced image forensics using texture analysis of median filter residual[J]. IEEE Access, 2020, 8: 103374-103384.

[3] CUN X, PUN C M. Image splicing localization via semi-global network and fully connected conditional random fields [C]//Proceedings of the European Conference on Computer Vision (ECCV) Workshops. Cham: Springer, 2018: 473-481.
 [4] LIU B, PUN C M. Deep fusion network for splicing forgery localization [C]//Proceedings of the European Conference on Computer Vision (ECCV) Workshops. Cham: Springer, 2018: 580-592.
 [5] XIAO B, WEI Y, BI X, et al. Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering[J]. Information Sciences, 2020, 511: 172-191.
 [6] ZHOU P, HAN X, MORARIU V I, et al. Learning rich features for image manipulation detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 1053-1061.
 [7] AHMED B, GULLIVER T A, ALZAHIR S. Image splicing detection using mask-RCNN [J]. Signal, Image and Video Processing, 2020, 14: 1035-1042.
 [8] 吴鹏,陈北京,郑雨鑫,等. 基于双流 Faster R-CNN 的像素级图像拼接篡改定位算法[J]. 电子测量与仪器学报, 2021, 35(4): 154-160.
 [9] DING H, CHEN L, TAO Q, et al. DCU-Net: A dual-channel U-shaped network for image splicing forgery detection[J]. Neural Computing and Applications, 2023, 35(7): 5015-5031.
 [10] MAZUMDAR A, BORA P K. Two-stream encoder-decoder network for localizing image forgeries [J]. Journal of Visual Communication and Image Representation, 2022, 82: 103417.
 [11] STERGIU A, POPPE R, KALLIATAKIS G. Refining activation downsampling with SoftPool [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 10357-10366.
 [12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.
 [13] 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. New York: IEEE, 2019: 9543-9552.
 [14] XIAO B, WEI Y, BI X, et al. Image splicing forgery detection combining coarse to refined convolutional neural network and adaptive clustering[J]. Information Sciences, 2020, 511: 172-191.