

基于强化样本的伪孪生网络图像篡改定位模型

王金伟^{1,2,3,4}, 张子荷^{1,2,3}, 罗向阳^{4,5}, 马宾⁶

- (1. 南京信息工程大学数字取证教育部工程研究中心, 江苏 南京 210044;
2. 南京信息工程大学计算机学院, 江苏 南京 210044; 3. 南京信息工程大学网络空间安全学院, 江苏 南京 210044;
4. 数字工程与先进计算国家重点实验室, 河南 郑州 450001; 5. 信息工程大学, 河南 郑州 450001;
6. 齐鲁工业大学网络空间安全学院, 山东 济南 250353)

摘要: 随着互联网不断发展, 网络上的篡改图像越来越多, 掩盖篡改痕迹的手段越来越丰富。而现在大多数检测模型没有考虑到图像后处理操作对篡改检测算法的影响, 限制了其在实际生活中的应用。为了解决上述问题, 提出了一种通用的基于强化样本的伪孪生网络图像篡改定位模型。所提模型利用伪孪生网络, 一方面学习真实图像中的篡改特征; 另一方面通过约束卷积, 抑制图像内容, 从而能够更加关注篡改残留的痕迹信息。网络的两分支结构可以达到充分利用图像特征信息的目的。模型利用强化样本, 可以自适应地生成当前最需要学习的篡改类型图片, 实现对模型有针对性地训练, 使得模型在各个方向上学习收敛, 最终得到全局最优模型。利用数据增强思路, 自动生成丰富的篡改图像以及其对应的掩膜, 这很好地解决了篡改数据集有限的问题。在 4 个数据集上的大量实验证明了所提模型在像素级操作检测方面的可行性和有效性。尤其是在 Columbia 数据集上, 算法的 $F1$ 值提高了 33.5%, Matthews correlation coefficient (MCC) 得分提高了 23.3%, 说明所提模型利用深度学习模型的优点, 显著提高了篡改定位的检测效果。

关键词: 强化样本; 篡改定位; 伪孪生网络; 数据增强; 篡改图像

中图分类号: TP391

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024010

Pseudo-siamese network image tampering localization model based on reinforced samples

WANG Jinwei^{1,2,3,4}, ZHANG Zihe^{1,2,3}, LUO Xiangyang^{4,5}, MA Bin⁶

1. Engineering Research Center of Digital Forensics, Ministry of Education,
Nanjing University of Information Science and Technology, Nanjing 210044, China
2. Department of Computer Science, Nanjing University of Information Science and Technology, Nanjing 210044, China
3. Department of Cyberspace Security, Nanjing University of Information Science and Technology, Nanjing 210044, China

收稿日期: 2023-08-25; **修回日期:** 2023-11-08

通信作者: 王金伟, wjwei_2004@163.com

基金项目: 国家自然科学基金 (62072250, 62172435, U1804263, U20B2065)

Foundation Item: The National Natural Science Foundation of China (62072250, 62172435, U1804263, U20B2065)

引用格式: 王金伟, 张子荷, 罗向阳, 等. 基于强化样本的伪孪生网络图像篡改定位模型[J]. 网络与信息安全学报, 2024, 10(1): 33-47.

Citation Format: WANG J W, ZHANG Z H, LUO X Y, et al. Pseudo-siamese network image tampering localization model based on reinforced samples[J]. Chinese Journal of Network and Information Security, 2024, 10(1): 33-47.

4. State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China

5. Force Information Engineering University, Zhengzhou 450001, China

6. School of Cyberspace Security, Qilu University of Technology, Jinan 250353, China

Abstract: With the continuous development of the internet, an increasing number of images have been tampered with on the network, accompanied by a growing range of techniques to cover up tampering traces. However, most current detection models neglect the impact of image post-processing on tamper detection algorithms, limiting their real-life applications. To address these issues, a general image tampering location model based on enhanced samples and the pseudo-twin network was proposed. The pseudo-twin network enabled the model to learn tampering features in real images. On one hand, by applying convolution constraints, the image content was suppressed, allowing the model to focus more on residual trace information of tampering. The two-branch structure of the network facilitated the comprehensive utilization of image feature information. By utilizing enhanced samples, the model could dynamically generate the most crucial pictures for learning tamper types, enabling targeted training of the model. This approach ensured that the model converged in all directions, ultimately obtaining the global optimal model. The idea of data enhancement was employed to automatically generate abundant tampered images and corresponding masks, effectively resolving the limited tampering dataset issue. Extensive experiments were conducted on four datasets, demonstrating the feasibility and effectiveness of the proposed model in pixel-level tamper detection. Particularly on the Columbia dataset, the algorithm achieves a 33.5% increase in $F1$ score and a 23.3% increase in MCC score. These results indicate that the proposed model harnesses the advantages of deep learning models and significantly improves the effectiveness of tamper location detection.

Keywords: enhanced sample, tampering positioning, pseudo-siamese network, data augmentation, tampering image

0 引言

随着图像处理技术的发展、篡改工具的升级,人们篡改图像的难度越来越低,数字图像篡改事件层出不穷,这对社会产生了很大的负面影响,对国家信任体系与安全构成了威胁。因此,研究篡改图像检测技术迫在眉睫。

在早期的研究工作中,研究者提出了许多检测篡改图像的方法,这些研究的成果对图像的真实性形成了很好的保护作用。检测方法主要分为传统取证和深度学习两种。

传统取证方法是检测篡改图像领域的研究基础。2012年, Ferrara 等^[1]将局部不存在彩色滤光阵列(CFA)伪影视为篡改的证据。2016年, Li 等^[2]为了能够充分利用彩色图像的全部信息,将离散四元余弦变换(QDCT, quaternion discrete cosine transform)引入复制粘贴伪造检测中,提出了一种基于 QDCT 的高效鲁棒检测方法。2017年, Ambili 等^[3]提出了一种适用于图像被篡改模糊情况下的拼接区域定位算法。2017年, 黄维等^[4]发现, 重获取图像会引入与原图不相关的背

景信息, 该图像进行多次压缩会产生块效应网格偏移。虽然传统取证方法在早期取得了一定的成绩, 但是人为设置的一些参数会在一定程度上影响检测效果。深度学习方法可以很好地解决传统方法出现的问题。近年来, 深度学习方法在图像篡改检测领域取得了很好的进展, 越来越多的研究者选择使用深度学习的方法来进行篡改检测。

深度学习可以将特征提取和分类过程整合起来, 实现以分类结果为指导的端到端特征学习方式。

2018年, Bayar 等^[5]提出了一种新的基于卷积神经网络的通用取证方法, 称之为约束卷积层。2018年, Wu 等^[6]提出了一个端到端深度神经网络(DNN, deep neural network)解决方案, 将经典解决方案中广泛使用的3个步骤在单个端到端 DNN 中实现。2021年, 蒋小玉等^[7]提出了一种基于篡改边缘两侧与篡改区域内外不一致性引导下的图像篡改检测网络。2021年, 朱新同等^[8]提出了一种基于图像纹理特征的篡改与伪造图像分类检测算法。2021年, 李小红等^[9]提出了一种适用于

移动设备端的篡改检测方法, 考虑到卷积神经网络层与参数的复杂性过高, 设计了剪枝压缩策略。2021年, 何家豪等^[10]提出了基于渐进式监督注意力机制嵌套 U-net 的图像复制粘贴伪造检测网络。2021年, Zhuang 等^[11]提出了一种全卷积编码器-解码器架构, 并实现了 SOTA 级别的定位性能。2022年, Wei 等^[12]提出了合成对抗性网络和混合密集 U-Net。2023年, Lu 等^[13]考虑深度学习方法的可解释性, 提出了一种可解释的图像篡改检测方法。

以上基于深度学习的篡改检测方法虽然在一定程度上达到了很好的检测效果, 但是还有一些急需解决的问题。例如, 在提取图像特征方面, 对图像内容的利用还不够充分; 现有的检测网络多数采用深度网络结构, 但现有的公开篡改数据集有限, 训练特征不够丰富, 尤其是模型的后处理操作特征不丰富, 模型的通用性不强, 不足以训练通用的深层网络, 限制了其在实际中的应用。

针对上述急需解决的问题, 本文展开了研究。以下的研究成果为本文研究提供了很多思路。为了解决现有公开数据集有限的问题, Zhao 等^[14]提出了一个新的图像合成方法。2022年, Chen 等^[15]使用各种不同的伪造方法生成篡改人脸图片, 该模型使用对抗网络, 进行针对性的模型训练。自监督学习^[16-18]也为研究提供了思路。为了解决对图像内容利用不充分的问题, 孪生网络^[19-20]为研究提供了思路, 利用孪生网络可以进行多分支的学习提取。为了解决检测模型的通用性问题, 强化样本、对比学习^[21]带来了很大的启发。

受到这些思路的影响, 本文针对现有主流图像篡改检测方法存在的问题, 提出了一个基于强化样本的伪孪生网络图像篡改定位模型。其中, 强化样本是指根据模型当前的学习能力, 有针对性地进行数据处理, 从而生成当前模型判断最薄弱部分的数据, 有针对性地进行加强训练; 伪孪生网络由两个对称的神经网络组成, 它们的架构相同, 但是权重不共享; 数据增强即数据增广, 是指使用旋转、缩放等方法增加训练数据集, 让数据尽可能多样化。本文所提模型采用伪孪生神经网络

结构, 一方面充分学习图像原图的特征信息; 另一方面通过约束卷积, 抑制图像内容, 让模型学习到篡改伪影等非语义信息。模型利用数据增强, 在训练过程中使网络自动生成篡改图片和掩膜, 用于丰富伪造的多样性。基于强化样本的思想, 设计的模型在训练过程中能够动态地找到当前最具挑战性的伪造。本文的贡献如下。

1) 提出了一个伪孪生网络图像篡改定位模型。该模型提取图像的语义内容和语义不可知内容, 并将两者融合学习。这有利于充分利用多个方面的图像信息。

2) 将强化样本的思想和图像篡改结合在一起, 这使得模型可以自适应地自动生成当前模型最具挑战性的伪造, 有利于模型学习当前薄弱的方向, 加快模型收敛速度。

3) 将数据增强思想进行改进后, 运用到模型训练过程中。每一轮训练都会重新生成大量的篡改图像和对应的掩膜, 丰富不可见的伪造空间, 且掩膜的同时生成节约了人工标注的成本, 很好地解决了当前篡改数据集有限的问题。

4) 本文通过在 4 个基准数据集上的大量实验, 证明了所提模型的有效性。

1 相关工作

数字图像作为人们交流的载体, 包含十分重要的信息。图像篡改的出现对图像的真实性造成了极大的威胁。目前, 已经有许多图像篡改检测方法被提出。

Cozzolino 等^[22]提出了一种基于特征的算法, 用于检测图像伪造。该方法通过计算图像残差的共现来提取局部特征, 并利用这些特征参数来区分合成图像和原始图像。然而, 该方法可能无法有效应对篡改后处理操作对伪造检测的影响。Zhou 等^[23]使用深度卷积神经网络来提取图像中的丰富特征, 其优势在于能够自动学习图像中的丰富特征, 无须手动设计, 能够捕捉到图像中的细节和篡改痕迹。但是对于使用后处理操作抹除篡改痕迹等一些复杂的篡改情况, 该方法仍存在一定的挑战。Zhao 等^[24]提出了一种基于深度学习的图像语义分割网络篡改检测方法。相较于传统

的基于局部特征的方法,该方法能够更全面地理解图像内容,从而提高了检测的准确性和鲁棒性。但是该方法仍没有考虑到后处理对图像篡改检测的影响。Ferrara 等^[1]提出了一种用于区分数码相机拍摄图片中真实区域和伪造区域的方法。利用摄像机的彩色滤光阵列去除马赛克时留下的伪影,将其视为一种数字指纹。研究人员发现,篡改操作会消除这种指纹痕迹。然而,该方法所使用的数据集都是以未经过 JPEG 压缩的 TIFF 格式存储的,这导致检测性能易受到 JPEG 压缩的影响,无法有效应对图像篡改后的后处理掩盖篡改痕迹的操作。

以上所提方法虽然在篡改检测上取得了一定的成果,但对于篡改后进行后处理操作抹除一定篡改痕迹后的图像,检测效果明显下降。Chen 等^[25]提出了一种基于信噪分离的网络,用于解决后处理图像的伪造检测问题。他们参考盲信号分离方法,将处理后的图像伪造检测问题转化为信噪分离问题。通过引入信噪分离模块,他们能够将篡改区域从具有后处理噪声的复杂背景中分离出来,从而削弱了后处理对篡改检测的负面影响,并提高了模型的鲁棒性。Lu 等^[13]提出了一种基于合作博弈的可解释图像篡改检测方法。重点是引入了合作博弈模块,通过多个合作的博弈网络对检测结果进行进一步的分析和验证,提高检测的准确性和鲁棒性,对于多种后处理操作和较小的区域都具有一定的检测效果。

2019 年, Wu 等^[26]设计了一个简单有效的自监督学习任务,巧妙地将篡改检测问题转化为局部异常检测,并利用长短期记忆(LSTM)网络模型进行异常评估。通过自监督学习任务的设计,该方法无须依赖大量标注数据,从而降低了数据获取和标注的成本。

2021 年, Zhuang 等^[11]提出了一种针对 Photoshop 篡改的方法,考虑到 Photoshop 在图像篡改中的广泛应用以及其痕迹的代表性。他们设计了一个完全卷积网络,其中包含密集卷积块,并在某些块中引入了膨胀卷积。通过利用密集连接和膨胀卷积的优势,该网络能够实现良好的篡改定位性能。然而,这种方法主要关注研究 Photoshop 这一特定的篡改软件,可能无法适用于

其他篡改软件的检测和定位。

2021 年, Chen 等^[27]提出了一种基于多视角多尺度监控的图像篡改检测机制,充分利用了图像多方面的特征。该机制采用噪声和边缘两分支学习篡改图像特征,使用的双线性在后期特征融合时是可训练的。该研究发现篡改图像的噪声和边缘信息对于语义不可知特征的学习都能做出很大的贡献。

2022 年, Wang 等^[28]提出了 ObjectFormer 进行检测和定位图像操作。ObjectFormer 能够全局地理解图像的上下文信息,通过对图像进行编码和解码来实现篡改检测和定位。但 ObjectFormer 模型的计算复杂性较高,需要更多的计算资源和时间。

以上所提方法对图像内容的利用还不够充分,应从多角度对图像特征进行提取,从而获得更丰富的特征信息;且公开篡改数据集有限,为了训练更深层的网络,应多样化扩充数据集。本文提出的扩充数据集方法集成在训练中,不仅降低了制作数据集的成本,同时有针对性地加强模型当前训练薄弱的方向。本文在后续实验中,将和以上所介绍方法进行对比,进一步证明本文所提模型的有效性。

2 本文模型

本文根据强化样本的思路,设计模型的输出结果能够指导下一轮模型的训练,加快模型训练收敛速度。利用数据增强,模型在训练过程中能够根据强化样本的指导结果自动生成篡改图片和篡改掩膜。利用伪孪生网络,更充分地利用图像特征。

2.1 整体网络模型框架

本文提出的基于强化样本的伪孪生网络图像篡改定位模型整体框架如图 1 所示。该模型首先通过数据增强模块,有针对性地生成不同篡改类型的篡改图像和其对应的掩膜;其次,网络分为上下两分支,以更充分地利用图像特征,网络上分支接收大小为 512×512 的篡改图像原图,将其输入骨干网提取特征,骨干网框架如图 2 所示。该骨干网参考 Zhao 等^[24]的研究。骨干网共有 57 个卷积层,分成了 20 个模块。其中,只有第

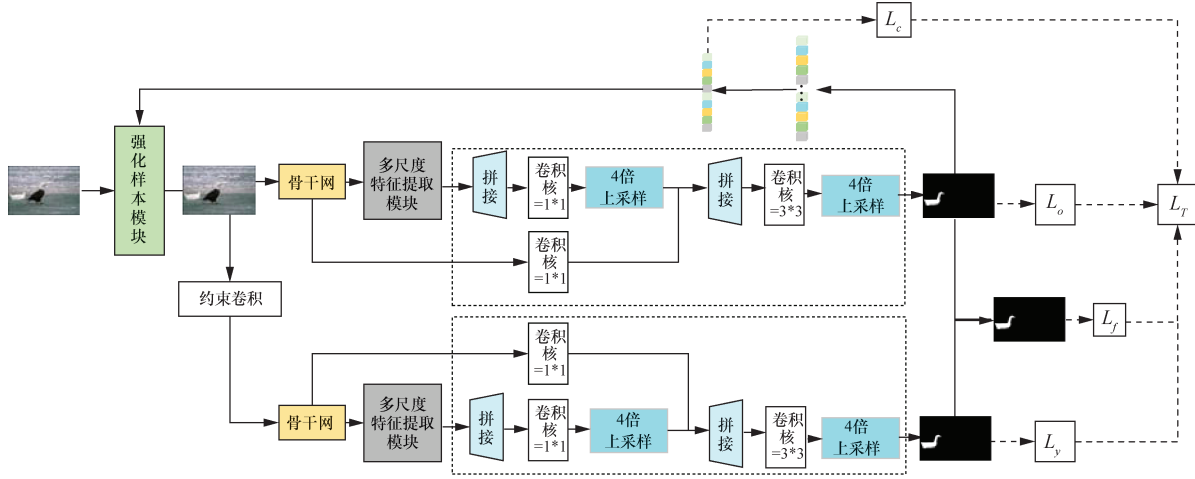


图 1 基于强化样本的伪孪生网络图像篡改定位模型整体框架

Figure 1 The overall framework of the pseudo-siamese network image tampering location model based on reinforced samples

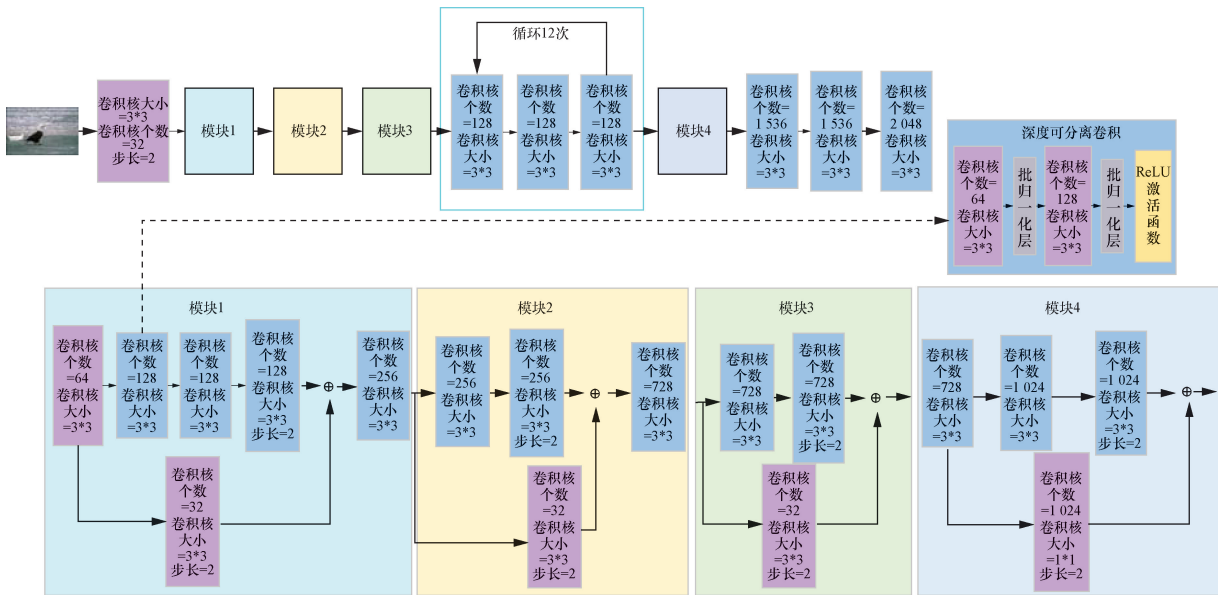


图 2 骨干网框架

Figure 2 Backbone network framework

一个模块使用的是普通卷积，其余的模块都使用深度可分离卷积，且每个深度可分离卷积后都紧跟一个批归一化（BN，batch normalization）层和一个 ReLU 激活函数。Exception 原始的池结构被步长为 2 的深度可分离卷积代替。

骨干网输出的高阶特征图进入多尺度特征提取模块，进行不同感受野大小的特征提取，并将提取的各个尺度特征图在通道维度拼接起来，得到图像更好的表征。多尺度特征提取模块如图 3 所示。由于图像的低阶特征具有丰富的细节信息，高阶特征具有丰富的语义信息。因此将高阶和低

阶两个输出特征图共同作为图像重建的特征依据。其中低阶特征采用骨干网络中 Block1 输出大小为 $128 \times 128 \times 256$ 的特征图。经过多尺度特征提取的高阶特征首先经过 1×1 卷积融合，然后使用双线性插值将特征图放大 4 倍，恢复成 128×128 大小后和骨干网输出的低阶特征拼接，然后经过 3×3 的卷积进行特征融合，再经过一个双线性插值放大 4 倍后得到该分支的预测掩膜。网络的下分支和上分支类似，不同点在于下分支接收的输入图片为经过约束卷积大小为 $512 \times 512 \times 3$ 的篡改图。最后模型将上下两分支

得到的预测结果进行通道维度上的拼接，再经过一个 1×1 的卷积学习融合，定位最终预测篡改区域。

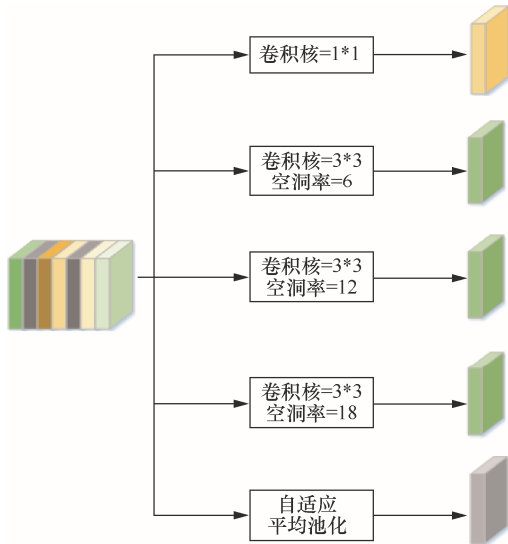


图 3 多尺度特征提取模块
Figure 3 Multi scale feature extraction module

本文设计的模型原图分支用于篡改分类预测，而预测结果用于指导下一轮强化样本模块篡改图片的生成。

2.2 强化样本模块

篡改后的图像通常会经过各种后处理操作，从而达到肉眼无法辨别的目的。这里使用的各种后处理操作，会给图像留下不同的特征痕迹。现在大多数检测模型没有考虑图像后处理操作对篡改检测算法的影响，这样训练好的模型不能对各种后处理操作都敏感。因此利用强化样本的思路，训练了一个通用的检测模型。

本文设计的模型在真实图像学习流的最后加入了一个全连接层，将篡改后处理类型进行分类预测。该分类预测结果作为强化样本篡改预测模块的输入。后处理操作类型共有 10 种，调用 PIL (Python imaging library) 模块，包括高斯模糊、均值滤波模糊、边缘增强、锐化、平滑、细节提升、改变亮度、改变色度、对比度增强、锐度增强。模型通过对这 10 种后处理操作的分类检测，计算出当前模型对各种后处理操作的检测能力。根据篡改后处理类型检测成功或失败的对应概率，有针对性地生成当前模型最具挑战性的后处理篡改图像。模型在每一个 epoch 都自适应地生

成新的篡改图像和相对应的掩膜用于网络训练，这丰富了伪造的多样性。同时根据模型的当前学习情况自适应地进行不同篡改后处理。使用对抗式训练的思维，使得模型接受当前最具挑战性的伪造检测。生成的篡改图像如图 4 所示。



图 4 生成的篡改图像
Figure 4 Generate tampered image

为了简化，本文默认图像被篡改后，只进行一种类型的后处理操作。

强化样本模块如图 5 所示。该部分将篡改后的大小为 $512 \times 512 \times 3$ 的图像输入后处理模块中，得到对应的一幅经过后处理的 $512 \times 512 \times 3$ 大小的篡改图像和对应的一张 $512 \times 512 \times 1$ 大小的掩膜，然后将其输入骨干网进行学习训练，这可以使得篡改图像更加真实，肉眼更难辨别，更符合真实世界的篡改图像流程。更重要的是，模型会根据上一轮的篡改检测能力，按照特定的概率实现对篡改图像的后处理操作。本文规定第一个 epoch 对图像篡改后处理的各种操作概率是随机的。下面举例说明篡改图像后处理模块的实施细节：设上一轮模型对各个后处理类型的检测成功概率为 $L=[L1, L2, \dots, L9]$ ，则下一轮对图像进行各种后处理的概率为 $1-L=[1-L1, 1-L2, \dots, 1-L9]$ 。该方法可以使得检测准确率高的后处理类型在下一轮对图片进行该后处理的概率大大减小。反之，上一轮模型检测能力欠佳的后处理类型，会在下一轮大大增加对图像进行该后处理操作的概率。该方法可以自动生成对当前模型挑战性最大的篡改图像。

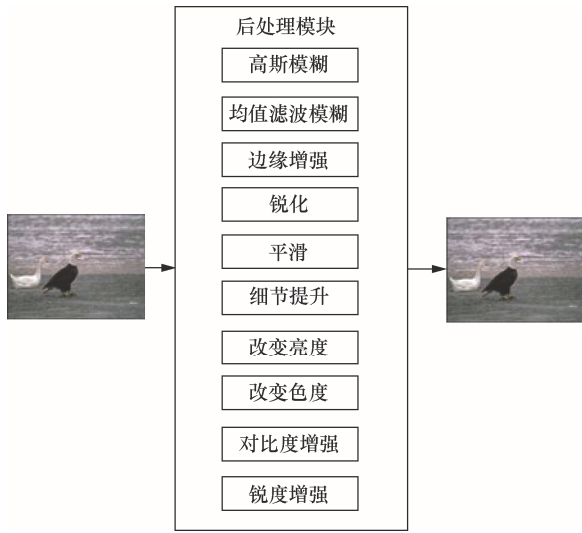


图 5 强化样本模块
Figure 5 Strengthening the sample module

2.3 伪孪生网络

针对现有大多数研究对图像的信息利用还不够充分的问题,为了更全面地利用图像特征信息,本文利用伪孪生网络的特性,使用两个分支来学习利用图像信息。伪孪生网络双分支示意如图 6 所示,它使得网络从图像原图和经过约束卷积模块后的图像中分别学习提取特征,这样能更充分地利用图像信息。网络的两分支都接收 $512 \times 512 \times 3$ 大小的图像,两分支分别得到 $512 \times 512 \times 1$ 大小的预测结果。最后模型将两条支路的预测结果在通道维度上进行拼接,得到 $512 \times 512 \times 2$ 的特征图,为了进一步融合两分支的特征,模型经过 1×1 大小的卷积核,输入通道数为 2,输出通道数为 1,融合得到 $512 \times 512 \times 1$ 大小的最终预测掩膜,该结果作为整个网络的预测输出。最终网络输出得到的像素级预测掩膜大小为 $512 \times 512 \times 1$ 。

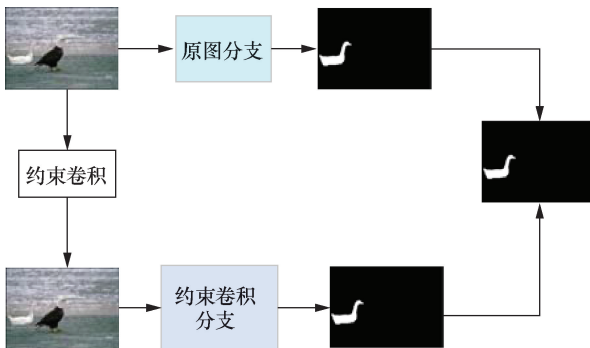


图 6 伪孪生网络双分支示意
Figure 6 Schematic diagram of dual branches in pseudo twin network

本文所提模型能够从真实图像和经过约束卷积处理后的图像中学习特征。本文的伪孪生网络权重不共享,使用两个优化器分别训练两个分支。原图分支关注图像的语义信息,约束卷积分支关注图像的非语义信息。两分支关注的信息存在较大的差异,单独的学习训练使得特征更加集中,更易提取。因此,使用原图和约束卷积的伪孪生网络在一定程度上可以提高模型的像素级检测能力,这充分说明了篡改图像中的篡改伪影等非语义信息对图像篡改检测的重要性,篡改痕迹不一定是语义的。

图像篡改检测任务在多数情况下并不需要过多地关注图像语义信息,不需要判断图像中是否存在某个语义物体。但是篡改一般是篡改图像中的语义物体,因此需要关注的是篡改痕迹引导的语义物体定位。图像篡改检测任务同样需要关注的是篡改伪影,关注篡改区域和非篡改区域的不一致性,这是语义不可知的。因此需要通过使用约束卷积模块,抑制图像内容,将篡改伪影等非语义信息突显出来,从而使得模型更好地学习图像中的非语义信息,更准确地学习到篡改痕迹特征。

2.4 强化样本制作数据集模块

当前公开篡改数据集的数据有限,为了解决这个问题,本文参考 Zhou 等^[23]的研究。

本文在 voc2012 数据集上进行篡改图像和其对应掩膜的制作,解决了当前篡改数据集有限的问题。在生成额外训练数据的基础上,将生成的篡改图像和强化样本结合起来,生成有针对性的学习数据,更好地配合模型训练。篡改数据集制作步骤如下,制作篡改图像数据集过程示例如图 7 所示。

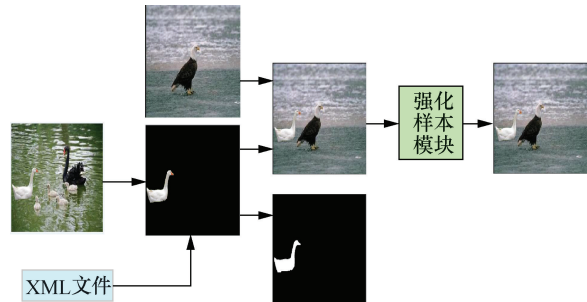


图 7 制作篡改图像数据集过程示例
Figure 7 Example of the process of creating tampered image datasets

1) 根据 voc2012 的 XML 标注文件, 随机选择一个目标图像, 然后在该图像中随机选择一个目标。

2) 将选择的目标作为篡改目标物体。

3) 对篡改目标物体做随机的旋转、缩放等变换, 每个操作发生的概率都为 50%。

4) 在 voc2012 中随机选择一个原始图像, 根据所选择的原始图像尺寸对目标物体的图像大小进行调整, 使得目标物体图像和随机选择的原始图像大小一样。之后将变换后的目标物体随机粘贴到图像中, 生成篡改图像。

5) 根据随机变换后的篡改目标生成篡改图像掩膜。

6) 将生成的篡改图像输入强化样本模块, 进行特定概率的后处理操作, 使得生成的篡改图像有针对性地对模型进行训练。

2.5 损失函数设置

为了衡量模型预测的性能, 本文为不同的分支选用不同的损失计算方式。整个模型损失函数表达式如式(1)所示。

$$L_T = L_o + L_y + L_f + L_c \quad (1)$$

其中, L_o 表示模型在原图分支上使用 BCELoss 计算的预测损失, L_y 表示图像在约束卷积分支上使用 BCELoss 计算的预测损失, L_f 表示模型融合学习两分支的预测结果后得到的最终预测结果使用 BCELoss 计算的预测损失, L_c 表示原图分支衡量预测分类损失。

BCELoss 表达式如式(2)所示。其中, t 表示标签, 即目标值, 它需要在 $[0,1]$, o 表示模型的预测值, 在计算损失之前需要经过 sigmoid 函数处理, $t[i]$ 表示标签掩膜第 i 个位置的值, $o[i]$ 表示模型预测第 i 个位置的值。

$$\begin{aligned} \text{BCELoss} = \text{Loss}(o, t) = \\ -\frac{1}{n} \sum (t[i] \cdot \log(o[i]) + (1 - t[i]) \cdot \log(1 - o[i])) \end{aligned} \quad (2)$$

原图分支将得到的掩膜结果放入全连接层, 得到预测的篡改分类结果。该分类结果将作为下一轮模型训练数据篡改后处理生成概率的依据。本文使用 Softmax 损失函数作为衡量预测分类结果的标准, 如式(3)所示。

$$\text{Softmax} = -\frac{1}{m} \sum_{i=1}^m y^i \log(p^i) + (1 - y^i) \log(1 - p^i) \quad (3)$$

其中, m 表示样本总数, p^i 是第 i 个样本的预测结果, y^i 是第 i 个样本的实际结果。

3 实验结果及分析

本节通过实验来验证所提模型在图像篡改检测上的有效性。本文在 4 个基准数据集上对所提模型进行了评估, 并且与之前的研究成果进行了比较。实验同时证明了使用数据增强、强化样本制作篡改图像的方法加入训练后, 对训练模型的正向作用, 证明了强化样本制作数据集的有效性。

3.1 实验设置

3.1.1 数据集介绍

为了评估所提模型, 本文在 CASIA^[29]、Columbia^[30]、DSO-1、NIST16 这 4 个数据集上进行实验, 还加入了一定比例的使用数据增强制作的篡改图像。

本文将公开的篡改数据集按照 0.8:0.2 的比例划分为训练集和测试集, 之后将使用数据增强制作的篡改图像和划分出的训练集按照一定的比例混合。本文实验采取的公开数据集和制作数据集的混合比例为 0.90:0.10、0.85:0.15、0.80:0.20、0.75:0.25。

1) CASIA 数据集中有拼接图像和复制粘贴篡改图像, 但没有提供标签。所有的篡改区域都是精心挑选并且经过了滤波和模糊操作淡化操作痕迹的。

2) Columbia 是一个全拼接图像的数据集, 所有图像都没有经过后处理操作, 该数据集提供篡改标签。

3) DSO-1 数据集是一个复制粘贴篡改图像数据集, 该数据集提供标签, 而且篡改后都进行了后处理操作。

4) NIST16 是一个很具有挑战性的数据集, 它包括拼接、复制粘贴和擦除等篡改, 该数据集提供标签。所有的篡改图像都经过了后处理操作来隐藏篡改痕迹。

3.1.2 网络参数设置

本文设计的网络模型接收的图像大小为

512×512×3，网络利用翻转、随机角度旋转等数据增强手段扩大数据的多样性，这样可以得到更多不同的篡改图像，之后将这些篡改图像放入强化样本模块进行后处理。篡改图像是内部生成且不可见的，从而可以进一步扩大数据伪造空间，探索更多的伪造可能。

本文训练网络设置的初始学习率为 1×10^{-4} 。训练过程中共设置 300 个 epoch，每 5 个 epoch 学习率下降 50%。Batch-size 设置为 3，本文的所有实验均在单个 NVIDIA GeForce GTX 1080 Ti GPU 上进行。

3.2 后处理方法选择

JPEG 压缩是最常见的后处理操作，但本文没有将其引入后处理模块中，这是因为 JPEG 压缩是一个不可导的过程，会产生很大的信息损失。并且压缩质量是由质量因子决定的，而质量因子有很多组，因此在训练过程中对于质量因子的选择是一个值得思考的问题。若将所有质量因子都经过强化样本模块进行训练，会使得该模块过于庞大和复杂。本文希望该模型的训练规模较小，希望能在较小规模的前提下同样达到很好的检测效果。而高斯模糊等后处理手段同样对图像造成高频损失、细节损失，与低通类似。因此选择高斯模糊等后处理操作代替 JPEG 压缩，其能够达到相似的效果，且缩小了后处理模块的规模。为了求证最终生成模型可以检测经过 JPEG 压缩后的图像，验证这些后处理类操作类型可以替代 JPEG 压缩，本文将测试集篡改图像经过 JPEG 压缩等后处理后，放入模型进行检测。测试集经过后处理后的篡改定位效果如表 1 所示，实验表明高斯模糊等后处理可以替代 JPEG 压缩，达到相似的效果，且缩小了强化样本模块的规模。

表 1 测试集经过后处理后的篡改定位效果
Table 1 The tampering localization effect of the test set after post-processing

后处理操作类型	CASIA 测试集 $F1$ 值
JPEG 压缩 (质量因子 $Q=80$)	0.532
高斯模糊 (半径 $r=2$)	0.534
边缘增强 (Filteraegs=(3,3), 2,0)	0.537
改变色度 (增强因子=0.7)	0.541
对比度增强 (增强因子=1.5)	0.535

3.3 模块分析

3.3.1 强化样本模块不同后处理参数消融实验

在本文实验采用的强化样本模块中，后处理操作类型一共有 10 种，包括高斯模糊（半径设置为 2）、均值滤波模糊（滤波半径设置为 6）、边缘增强、锐化、平滑、细节提升（这 4 种后处理都采用 ImageFilter 中的默认设置参数）、改变亮度（增强因子设置为 0.8）、改变色度（增强因子设置为 0.7）、对比度增强（增强因子设置为 1.5）、锐度增强（增强因子设置为 3.0）。为了探究设置不同后处理参数对最终网络检测性能的影响，本文改变强化样本模块中的后处理参数，进行消融训练实验，实验结果如表 2 所示。

实验结果发现，使用不同的后处理参数对网络检测性能的影响较小，网络仍能很好地进行篡改定位检测，这是因为设计的网络在强化样本模块中，使用不同后处理操作的图像会加入模型的训练过程。强化样本模块后处理参数的改变，同样会在模型训练时对训练数据进行相应的改变，也就是说，所提模型可以对不同的后处理操作参数有效，这也是所提模型的优势。

3.3.2 自制数据集有效性消融实验

为了验证所制作篡改数据集的有效性，证明可以降低人为制作篡改数据集的成本，能够加入训练集进行模型训练，本文进行了如下实验。训练集中共使用 5 516 幅篡改图像，包括不同比例的制作篡改图像和 CASIA 数据集中的篡改图像。测试集全部为 CASIA 数据集中经过后处理的篡改图像。通过不断地增加训练集中自制数据集的占比，使用越来越多的自制数据集代替 CASIA 篡改训练集，训练完成后，在 CASIA 测试集中测试模型的篡改检测效果。

训练集中自制数据集和 CASIA 数据集的不同占比对测试集 $F1$ 值的影响效果评估如表 3 所示。结果表明，随着自制数据集占比不断扩大，CASIA 测试集的像素级 $F1$ 值基本保持稳定。这说明本文制作的数据集能够有效补充公开标准数据集，达到成功训练模型的目的。该方法可以极大地减少人为手动制作篡改数据集的时间和精力。而且经过强化样本模块后，输入图像还可以有针对性生成当前模型最需要学习的篡改后处

理类型图像, 加快模型的训练收敛速度, 减少深度学习训练的时间成本。

表 2 强化样本模块不同后处理参数对模型性能的影响
Table 2 The effect of different postprocessing parameters on model performance in the reinforcement sample module

后处理操作	CASIA 测试集 $F1$ 值
高斯模糊 ($r=2$)	0.542
均值滤波 ($r=6$)	
边缘增强 (Filteraegs=(3,3),2,0)	
锐化 (Filteraegs=(3,3),16,0)	
平滑 (Filteraegs=(5,5),100,0)	
细节提升 (Filteraegs=(3,3),1,255)	
改变亮度 (增强因子=0.8)	
改变色度 (增强因子=0.7)	
对比度增强 (增强因子=1.5)	
锐度增强 (增强因子=3.0)	
高斯模糊 ($r=4$)	0.532
均值滤波 ($r=10$)	
边缘增强 (Filteraegs=(3,3),2,0)	
锐化 (Filteraegs=(3,3),16,0)	
平滑 (Filteraegs=(5,5),100,0)	
细节提升 (Filteraegs=(3,3),1,255)	
改变亮度 (增强因子=0.9)	
改变色度 (增强因子=0.8)	
对比度增强 (增强因子=1.6)	
锐度增强 (增强因子=3.1)	
高斯模糊 ($r=6$)	0.538
均值滤波 ($r=12$)	
边缘增强 (Filteraegs=(3,3),2,0)	
锐化 (Filteraegs=(3,3),16,0)	
平滑 (Filteraegs=(5,5),100,0)	
细节提升 (Filteraegs=(3,3),1,255)	
改变亮度 (增强因子=1.0)	
改变色度 (增强因子=0.9)	
对比度增强 (增强因子=1.7)	
锐度增强 (增强因子=3.2)	

3.3.3 消融实验

为了说明强化样本模块和伪孪生网络的有效性, 本文使用 CASIA、DSO-1、Columbia 这 3 个基准数据集进行消融实验。原图流、约束卷积流、

强化样本模块的有无在 CASIA、DSO-1、Columbia 数据集上对 $F1$ 值的影响如表 4 所示。

表 3 训练集中自制数据集和 CASIA 数据集的不同占比对测试集 $F1$ 值的影响效果评估

Table 3 Evaluation of the impact of different proportions of the production dataset and CASIA dataset in the training set on the $F1$ of the test set

训练集中 CASIA: 自制数据集	训练集中 CASIA 图像 数量	训练集中 制作图像 数量	测试集 CASIA $F1$ 值
0.90:0.10	4 965	551	0.541
0.85:0.15	4 689	827	0.535
0.80:0.20	4 413	1 103	0.539
0.75:0.25	4 137	1 379	0.537
0.70:0.30	3 861	1 655	0.532
0.60:0.40	3 309	2 207	0.526
0.40:0.60	2 206	3 310	0.531
0.30:0.70	1 654	3 862	0.527

表 4 原图流、约束卷积流、强化样本模块的有无在 CASIA、DSO-1、Columbia 数据集上对 $F1$ 值的影响
Table 4 The influence of the presence or absence of original graph flow, constrained convolutional flow, and reinforcement sample modules on $F1$ scores on CASIA, DSO-1, and Columbia datasets

数据集	原图流	约束 卷积流	强化 样本模块	$F1$ 值
CASIA	√			0.458
	√	√		0.492
	√	√	√	0.542
DSO-1	√			0.598
	√	√		0.634
	√	√	√	0.683
Columbia	√			0.846
	√	√		0.894
	√	√	√	0.973

实验发现, 约束卷积流和强化样本模块都存在时的模型性能比没有强化样本模块时性能好, $F1$ 值得到了很大的提升, 平均提升了 5.9%。尤其在 Columbia 数据集上, $F1$ 值提升了 12.7%。为了更直观地看到伪孪生网络和强化样本模块对模型训练的影响, 本文绘制了折线图表示模型各个模块对模型训练收敛速度和 $F1$ 值的影响。在 CASIA 数据集上的评测结果如图 8 所示, 在 DSO-1 数据集上的评测结果如图 9 所示, 在

Columbia 数据集上的评测结果如图 10 所示。实验结果发现, 添加了语义不可知特征提取视图和强化样本模块后, 像素级篡改定位检测效果得到了提升, 它可以更加精准地定位篡改区域。这是因为在篡改检测任务中, 图像的篡改痕迹更多地存在于语义不可知特征, 所以使用约束卷积抑制图像的语义特征变得十分重要。

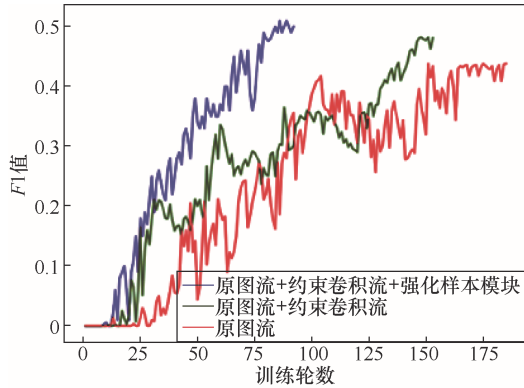


图8 原图流、约束卷积流、强化样本模块的有无在 CASIA 数据集上对模型训练收敛速度的评测结果

Figure 8 The evaluation results of rate of convergence of model training on CASIA dataset for the presence or absence of original image flow, constrained convolution flow and enhanced sample module

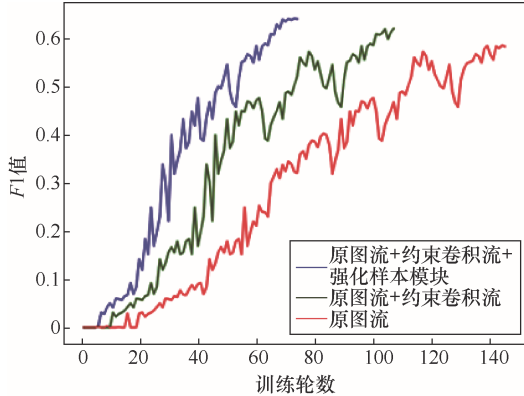


图9 原图流、约束卷积流、强化样本模块的有无在 DSO-1 数据集上对模型训练收敛速度的评测结果

Figure 9 The evaluation results of Rate of convergence of model training on DSO-1 data set for the presence of original graph flow, constrained convolution flow and enhanced sample module

因此, 本文设计采用伪孪生网络, 使用双分支分别学习图像的语义特征和语义不可知特征。这样可以充分地利用图像篡改特征, 从而达到更好的检测效果。加入强化样本的网络模型, 在模型训练时, 收敛速度得到了极大的提升。这是因为强化样本模块可以有针对性地生成当前模型最需要学习训练的篡改特征, 可以对网络进行特定的训练。模型会根据上一轮的训练结果, 得知当

前网络最需要训练的篡改后处理类型, 从而自动地生成相应类型的篡改图。这在一定程度上解决了深度学习训练模型耗时长的的问题, 提高了模型检测定位的能力。实验结果证明了所提模型在加快训练速度和提高定位精度上的能力。

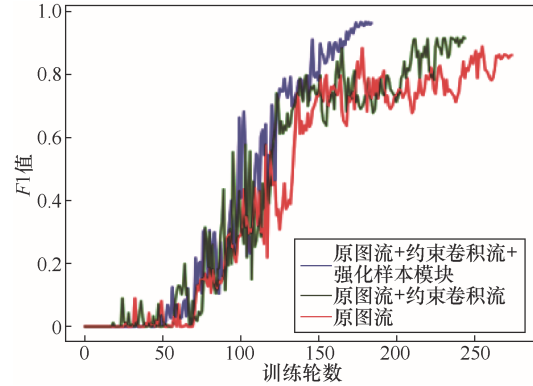


图10 原图流、约束卷积流、强化样本模块的有无在 Columbia 数据集上对模型训练收敛速度的评测结果

Figure 10 Whether the original image flow, constrained convolution flow and enhanced sample module have evaluated the Rate of convergence of model training on the Columbia dataset

3.3.4 骨干网循环次数消融实验

构建神经网络的规模大小在很大程度上会影响模型的检测效果, 如果模型参数数量过多, 模型过大, 很容易导致训练过拟合。反之, 模型参数不足, 会导致训练效果不佳, 训练欠拟合。所以, 找到一个平衡点对模型的训练检测效果至关重要。本文实验测试了骨干网中循环模块不同循环次数对模型检测性能的影响。本文使用 CASIA 数据集作为实验标准, 将数据集按照 0.8:0.2 的比例划分为训练集和测试集。在训练集上进行模型的训练, 训练模型收敛之后, 在划分出的测试集上进行测试, 计算得到测试集 $F1$ 值。骨干网络中循环模块在不同循环次数下的模型训练的 $F1$ 值变化如图 11 所示。实验结果表明, 当设置骨干网的循环模块循环 12 次时, 模型可以达到最佳的检测效果。因此, 本文进行的实验都采用 12 次循环。

3.3.5 强化样本指导消融实验

强化样本的预测模块采用原图流的预测分类输出作为输入, 实验时对选择哪一个分支的预测分类作为强化样本模块的指导流输入也进行了消融实验, 结果如表 5 所示。实验时分为原图流预测分类、约束卷积流预测分类、融合流预测分类 3 种情况。实验结果发现, 使用原

图流预测分类训练的模型效果较好。这是因为原图流的篡改图像并没有经过约束卷积，因此保留的语义信息更多，模型提取的图像颜色、纹理等特征更丰富，更有利于分类识别。因此，模型最终采用原图流的预测分类结果作为强化样本模块的指导流输入。

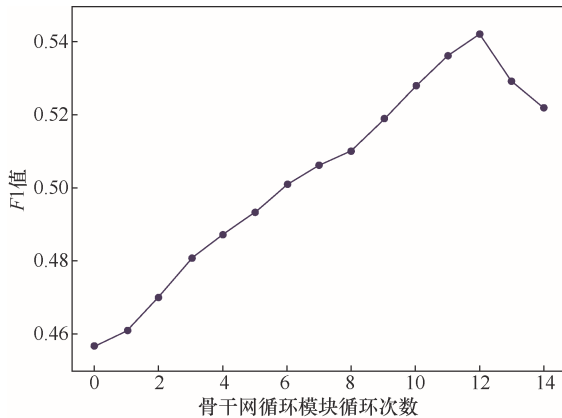


图 11 骨干网中循环模块在不同循环次数下的模型训练的 $F1$ 值变化

Figure 11 Model training $F1$ change of cycle module in backbone network under different cycle times

表 5 强化样本模块不同指导流输入对测试集分数的影响

Table 5 The influence of different guidance flow inputs on test set scores in the reinforcement sample module

指导流	CASIA $F1$ 值
强化样本+原图流预测分类	0.542
强化样本+约束卷积流预测分类	0.528
强化样本+融合流预测分类	0.534

3.4 模型通用性

为了验证所提模型的通用性，证明所提模型在不同数据集上的训练效果。本文增加使用 NIST16 和自制数据集进行不同比例的混合实验。NIST16 共有 564 幅图像，按照 0.8:0.2 的比例划分为训练集和测试集。训练集有 451 幅图像，测试集有 113 幅图像。由于数据集本身数量较少，所以加入自制数据集进行补充，前面 CASIA 和自制数据集不同比例的训练实验能够说明自制数据集可以作为补充数据加入训练。因此，最终训练数据集共有 1 503 幅图像，包括不同比例的 NIST16 和自制数据集。

训练集中自制数据集和 NIST16 数据集的不同占比对测试集 $F1$ 值的影响效果评估如表 6 所

示。结果表明，随着制作数据集占比的不断扩大，NIST16 测试集的像素级 $F1$ 值基本保持稳定，这说明所提模型对不同数据集训练都有一定的检测效果，可以证明所提模型的通用性。

表 6 训练集中自制数据集和 NIST16 数据集的不同占比对测试集 $F1$ 值的影响效果评估

Table 6 Evaluation of the impact of different proportions of the production dataset and NIST16 dataset in the training set on the $F1$ of the test set

训练集中 NIST16: 自制数据集	训练集中 NIST16 图像数量	训练集中制作图像数量	测试集 NIST16 $F1$ 值
0.10:0.90	150	1353	0.628
0.20:0.80	300	1203	0.631
0.30:0.70	451	1052	0.625

3.5 模型性能评估

本文在 4 个基准数据集上对本文模型进行了实验测试。表 7 为模型在 CASIA、DSO-1、NIST16 和 Columbia 数据集上的评测结果。本文分别计算了 $F1$ 值、MCC、召回率和精确率进行模型性能的评估。从实验结果可以看到所提模型在篡改检测定位上的有效性。尤其在 Columbia 数据集上 $F1$ 值达到了 0.973。分析认为，Columbia 数据集的篡改区域没有进行后处理操作，篡改特征较明显，这有利于约束卷积分支对篡改伪影的提取学习。实验证明了所提模型对图像篡改检测任务的有效性。

表 7 模型在 CASIA、DSO-1、NIST16 和 Columbia 数据集上的评测结果

Table 7 Evaluation results of the model on the CASIA, DSO-1, NIST16 and Columbia datasets

数据集	$F1$ 值	MCC	召回率	精确率
CASIA	0.542	0.512	0.506	0.584
DSO-1	0.683	0.573	0.678	0.689
NIST16	0.635	0.572	0.639	0.631
Columbia	0.973	0.947	0.957	0.988

3.6 对比实验

本节将所提模型和现有的研究模型进行了比较：NOI3^[22]、CFA2^[1]、SNIS^[25]、RGB-N^[23]、Zhao^[24]、ManTra^[26]、MVSS-Net^[27]、Lu^[13]、DenFCN^[11]、Wang^[28]，在 CASIA、Columbia 数据集上计算像素级 $F1$ 值和 MCC 值，以进行模型性能评估。如表 8 所示，“—”表示文献中未提供相应的结果。

从表 8 可以看出, 本文提出的模型在检测效果方面有一定程度上的提高。特别是在 Columbia 数据集上, 模型的 $F1$ 值提高了 33.5%, MCC 得分提高了 23.3%, 这是由于 Columbia 数据集的篡改图像没有经过任何的篡改后处理, 篡改操作留下的篡改特征痕迹没有得到破坏, 这有利于模型对该篡改特征的学习。对于在 CASIA 数据集上相对较差的表现, 本文认为该数据集中的篡改图像都是精心挑选并且经过了滤波和模糊操作, 淡化了篡改操作痕迹的, 这导致大部分篡改特征痕迹消失。数据集中很多篡改后的图像肉眼已经无法区分, 这对模型训练造成了阻碍, 不足以很好地学习篡改痕迹特征。

对于 CFA2^[1]在 CASIA 上的表现, 本文分析认为由于该模型使用的数据集都是未经过 JPEG 压缩的 TIFF 格式, 所以该模型的检测性能会受到 JPEG 压缩的影响。CASIA 数据集中有经过 JPEG 压缩的篡改图像, 所以检测效果较差。

对于 MVSS-Net^[27]在 CASIA 和 Columbia 两个数据集上相对较差的表现, 本文分析认为该模型虽然兼顾了图像的语义和非语义信息, 但是其为了兼顾模型在篡改图像检测上的灵敏度和在真实未篡改图像上的特异度, 将最后的检测结果掩膜进行全局最大池化, 并将其加入损失函数计算, 这在一定程度上会使得模型具有约束性。

表 8 在两个基准数据集上的 $F1$ 值和 MCC 值
Table 8 $F1$ and MCC values on two benchmark datasets

模型	CASIA		Columbia	
	$F1$ 值	MCC	$F1$ 值	MCC
NOI3 ^[22]	0.176	0.089	0.448	0.207
CFA2 ^[1]	0.212	0.162	0.503	0.328
SNIS ^[25]	0.368	—	0.681	—
RGB-N ^[23]	0.408	0.287	0.697	—
Zhao ^[24]	0.423	0.427	0.733	0.714
ManTra ^[26]	0.46	0.45	0.793	0.742
MVSS-Net ^[27]	0.452	0.404	0.638	—
Lu ^[13]	0.51	0.49	—	—
DenFCN ^[11]	0.559	0.516	—	—
Wang ^[28]	0.579	0.518	—	—
本文模型	0.542	0.512	0.973	0.947

对于 Wang 等^[28]在 CASIA 上的表现, 本文认为该模型不仅利用了篡改原图提取特征, 还对输入图像进行 DCT 变换, 得到高频信息发现伪造痕迹; 另外, 该模型中的物体编码器和解码器可以对图像中的不同物体相互作用, 使得图像中相应区域的特征更具判别性, 有效地建模了篡改图像物体层面的一致性信息。

3.7 实验结果可视化

图 12 展示了本文所提模型的篡改检测定位效果, 第一列为篡改图像原图, 第二列为标注掩膜标签, 第三列为模型输出预测结果。从图中可以看出, 本文所提模型可以完成篡改区域的定位检测, 可以较好地分割出篡改区域。本文提出的模型是一个端到端的模型, 无须任何预处理和后处理。实验证明了所提模型在篡改检测方面有着不错的检测定位性能。

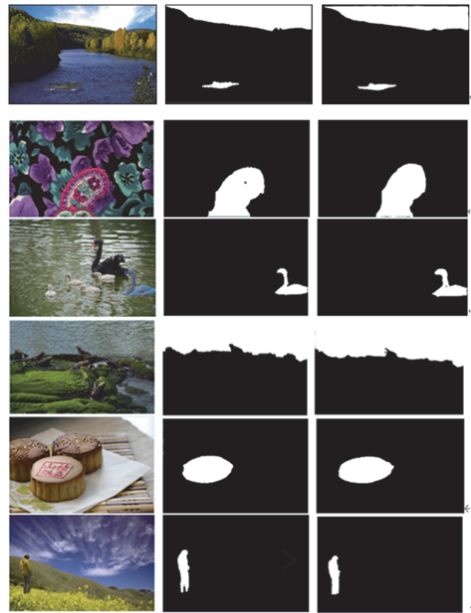


图 12 本文所提模型的篡改检测定位效果
Figure 12 The tamper detection and localization effect of the model proposed in this article is demonstrated

4 结束语

本文提出了一个基于强化样本的伪孪生网络图像篡改定位模型。为了更充分地提取利用图像篡改特征, 本文模型利用伪孪生网络, 使用两个分支对篡改图的语义特征和非语义特征分别学习提取, 最后进行融合学习。为了得到更多的、具

有针对性的篡改图, 本文使用强化样本和数据增强相结合的方法, 在生成新的篡改图像的基础上进行针对性的后处理操作, 使得模型能够对各种后处理操作都敏感, 不会因为后处理操作掩盖了篡改伪影而受到限制。同时, 制作篡改图像解决了目前篡改数据集有限的问题。该模型由于在每一个 epoch 都有针对性地生成经过后处理的篡改图像, 使得图像在各个方向上收敛训练, 最终得到全局最优模型。本文在不同数据集上的大量实验证明了所提模型的可行性和有效性。

虽然本文在图像篡改定位领域取得了较好的效果, 但仍需在未来的工作中针对后处理模块的方法进行补充, 增强其完备性。同时, 考虑将所提模型封装为系统进行实际的应用, 根据反馈意见进一步进行完善。

参考文献:

- [1] FERRARA P, BIANCHI T, DE ROSA A, 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.
- [2] LI C, MA Q, XIAO L M, et al. An image copy move forgery detection method using QDCT[C]//Proceedings of the International Conference on Internet Multimedia Computing and Service. 2016: 5-8.
- [3] AMBILI B, GEORGE N. A robust technique for splicing detection in tampered blurred images[C]//Proceedings of 2017 International Conference on Trends in Electronics and Informatics (ICEI). 2017: 897-901.
- [4] 黄维, 黄添强, 张雪莉, 等. 基于块效应网格偏移的重获取 JPEG 图像篡改检测[J]. 网络与信息安全学报, 2017, 3(12): 24-30.
HUANG W, HUANG T Q, ZHANG X L, et al. JPEG recapture image tamper detection method based on block effect grid offset[J]. Chinese Journal of Network and Information Security, 2017, 3(12): 24-30.
- [5] BAYAR B, STAMM M C. Constrained convolutional neural networks: a new approach towards general purpose image manipulation detection[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2691-2706.
- [6] WU Y, ABD-ALMAGEED W, NATARAJAN P. Image copy-move forgery detection via an end-to-end deep neural network[C]//Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). 2018: 1907-1915.
- [7] 蒋小玉, 刘春晓. 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, 2021, 26(10): 2411-2420.
JIANG X Y, LIU C X. Edge and region inconsistency-guided image splicing tamper detection network[J]. Journal of Image and Graphics, 2021, 26(10): 2411-2420.
- [8] 朱新同, 唐云祁, 耿鹏志. 基于特征融合的篡改与深度伪造图像检测算法[J]. 信息安全学报, 2021(8): 70-81.
- [9] ZHU X T, TANG Y Q, GENG P Z. Detection algorithm of tamper and deepfake image based on feature fusion[J]. Netinfo Security, 2021(8): 70-81.
- [9] 李小红, 王晓霞. 一种剪枝压缩 CNN 移动设备端图像篡改检测方法[J]. 计算机仿真, 2021, 38(3): 83-86, 91.
LI X H, WANG X X. An image tamper detection method for CNN mobile devices based on pruning compression[J]. Computer Simulation, 2021, 38(3): 83-86, 91.
- [10] 何家豪. 图像复制移动篡改检测研究及实现[D]. 大连: 大连理工大学, 2021.
HE J H. Research and implementation of image copy mobile tampering detection[D]. Dalian: Dalian University of Technology, 2021.
- [11] 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.
- [12] WEI Y, MA J F, WANG Z Z, et al. Image splicing forgery detection by combining synthetic adversarial networks and hybrid dense U-net based on multiple spaces[J]. International Journal of Intelligent Systems, 2022, 37(11): 8291-8308.
- [13] LU W, XU W B, SHENG Z Q. An interpretable image tampering detection approach based on cooperative game[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(2): 952-962.
- [14] ZHAO T C, XU X, XU M Z, et al. Learning self-consistency for deepfake detection[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021: 15003-15013.
- [15] CHEN L, ZHANG Y, SONG Y B, et al. Self-supervised learning of adversarial example: towards good generalizations for deepfake detection[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 18689-18698.
- [16] SOLIMAN M, LEHMAN C, ALREGIB G. S6: semi-supervised self-supervised semantic segmentation[C]//Proceedings of 2020 IEEE International Conference on Image Processing (ICIP). 2020: 1861-1865.
- [17] XU H, LI X, GONG W Y. Research on recognition of landslides with remote sensing images based on extreme learning machine[C]//Proceedings of 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC). 2017: 740-747.
- [18] HALIASSOS A, MIRA R, PETRIDIS S, et al. Leveraging real talking faces via self-supervision for robust forgery detection[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022: 14930-14942.
- [19] KRISHNA V A, REDDY A, NAGAJYOTHI D. Signature recognition using Siamese neural networks[C]//Proceedings of 2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNC). 2021: 1-4.
- [20] WANG Y Y, ZHANG W S, ZHANG L M, et al. Depthwise over-parameterized Siamese network for visual tracking[C]//Pro-

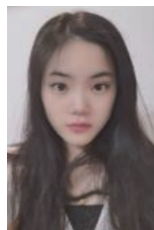
- ceedings of 2021 International Conference on Information Technology and Biomedical Engineering (ICITBE). 2022: 58-62.
- [21] SUN K, YAO T P, CHEN S, et al. Dual contrastive learning for general face forgery detection[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(2): 2316-2324.
- [22] COZZOLINO D, POGGI G, VERDOLIVA L. Splicebuster: a new blind image splicing detector[C]//Proceedings of 2015 IEEE International Workshop on Information Forensics and Security (WIFS). 2015: 1-6.
- [23] 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. 2018: 1053-1061.
- [24] ZHAO Q, CAO G, ZHOU A T, et al. Image tampering detection via semantic segmentation network[C]//Proceedings of 2020 15th IEEE International Conference on Signal Processing (ICSP). 2020: 165-169.
- [25] CHEN J X, LIAO X, WANG W, et al. SNIS: a signal noise separation-based network for post-processed image forgery detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(2): 935-951.
- [26] 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). 2019: 9535-9544.
- [27] CHEN X R, DONG C B, JI J Q, et al. Image manipulation detection by multi-view multi-scale supervision[C]//Proceedings of International Conference on Computer Vision. 2021.
- [28] WANG J, WU Z, CHEN J, et al. ObjectFormer for image manipulation detection and localization[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2022.
- [29] 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. 2013: 422-426.
- [30] HSU Y F, CHANG S F. Detecting image splicing using geometry invariants and camera characteristics consistency[C]//Proceedings

of IEEE International Conference on Multimedia and Expo. 2006.

[作者简介]



王金伟（1978—），男，内蒙古呼伦贝尔人，南京信息工程大学教授、博士生导师，主要研究方向为人工智能安全、多媒体取证和信息隐藏。



张子荷（1999—），女，河北廊坊人，南京信息工程大学硕士生，主要研究方向为信息安全、多媒体取证。



罗向阳（1978—），男，湖北荆门人，信息工程大学教授、博士生导师，主要研究方向为图像隐写、隐写分析技术。



马宾（1973—），男，山东济宁人，齐鲁工业大学教授、博士生导师，主要研究方向为可逆信息隐藏、多媒体取证、隐写与隐写分析。