提高图像篡改检测区域选取性能的 FCR-CNN 模型

魏晓燕¹⁾, 左鑫兰¹⁾, 但志平^{1,2)}, 吴义熔^{1,3)}, 董方敏^{1,2)}, 孙水发^{1,2,3)*}

¹⁾ (三峡大学计算机与信息学院 宜昌 443002)

²⁾ (湖北省水电工程智能视觉监测重点实验室(三峡大学) 宜昌 443002)

³⁾ (智慧医疗宜昌市重点实验室 宜昌 430002)

(watersun@ctgu.edu.cn)

摘 要: 图像篡改检测不同于目标检测, 篡改检测更加关注篡改伪像这一目标而非图像内容本身, 需要学习更加丰富的特征. 为提高图像篡改检测区域选取性能, 提出一种将特征金字塔网络(feature pyramid networks, FPN)模型与级联区域卷积神经网络(cascade region-convolutional neural networks, Cascade R-CNN)模型相结合的 FCR-CNN 模型. 首先将 FPN 模型提取的多尺度篡改特征输入到区域建议网络(region proposal network, RPN), 然后由 RPN 输出篡改分类分数和区域建议框, 最后将区域建议框输入到 3 阶段 Cascade R-CNN 进行检测. 此外, 系统地分析 FCR-CNN 模型的损失函数. 基于 CASIA, Columbia 和 NC2016 数据集, 与其他算法进行对比实验, 结果表明, FCR-CNN 模型能够有效地检测与定位篡改区域; 其中, 在 CASIA 数据集上, 其与 FPN 模型和 Cascade R-CNN 模型相比, F_1 分数分别提高了 6.0% 和 7.5%.

关键词: 图像篡改检测; 特征金字塔网络; 级联区域卷积网络; 区域建议框
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18512

FCR-CNN Model to Improve Selection Performance of Image Manipulation Detection Region

Wei Xiaoyan¹⁾, Zuo Xinlan¹⁾, Dan Zhiping^{1,2)}, Wu Yirong^{1,3)}, Dong Fangmin^{1,2)}, and Sun Shuifa^{1,2,3)*}

¹⁾ (College of Computer and Information Technology, China Three Gorges University, Yichang 443002)

²⁾ (Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, China Three Gorges University, Yichang 443002)

³⁾ (Yichang Key Laboratory of Intelligent Medicine, Yichang 443002)

Abstract: Image manipulation detection is different from object detection as it pays more attention to tampered artifacts rather than image content and needs to learn to obtain more abundant features. The FCR-CNN model that combines the feature pyramid networks (FPN) model and the cascade region-convolutional neural networks (Cascade R-CNN) model is proposed to improve the performance of image manipulation detection. The FCR-CNN model first inputs the multi-scale manipulation features extracted by the FPN network to the region proposal network (RPN), then outputs the manipulation classification score and the region proposal from the RPN network, and finally the region proposal is input to the three-stage Cascade R-CNN model for detection. Additionally, the loss function of the FCR-CNN model is systematically analyzed. Comparative experiments with other detection models are also performed on the CASIA, Columbia, and NC2016 datasets. The results show that the FCR-CNN model can effectively detect and locate manipulation areas. Among

收稿日期: 2020-07-24; 修回日期: 2020-11-25. 基金项目: 国家自然科学基金(U1703261, 61871258); 湖北省自然科学基金(2019CFB215). 魏晓燕(1995—), 女, 硕士研究生, 主要研究方向为图像篡改检测; 左鑫兰(1995—), 女, 硕士研究生, 主要研究方向为图像篡改检测; 但志平(1976—), 男, 博士, 教授, 硕士生导师, 主要研究方向为智能信息处理; 吴义熔(1970—), 男, 博士, 教授, 博士生导师, 主要研究方向为人工智能; 董方敏(1965—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为智能信息处理; 孙水发(1977—), 男, 博士, 教授, 博士生导师, CCF 会员, 论文通讯作者, 主要研究方向为计算机视觉、图像处理.

them, compared with the FPN model and the Cascade R-CNN model on the CASIA dataset, the F_1 score of the FCR-CNN model increased by 6.0% and 7.5%, respectively.

Key words: image manipulation detection; feature pyramid networks; cascade region-convolutional neural networks; region proposal

随着数字相机的普及和数字图像处理技术的快速发展,借助一些对用户友好的图像编辑软件,如 Photoshop 和美图秀秀等,用户可以随时随地非常轻松地对图像进行编辑,并共享到各大社交平台,如微信和微博等,但其中不乏一些恶意的编辑掩盖了图像的真实性,混淆视听。因此,有必要开展检测图像是否发生篡改的相关研究,这对于数字图像取证、防伪以及军事防卫具有重要意义。

传统图像篡改检测算法根据篡改所引起的图像统计特性的变化进行检测,如局部噪声特性^[1]、颜色滤波阵列(color filter array, CFA)特性^[2]、JPEG 压缩特性^[3]等。孙鹏等^[4]提出基于色温估计的拼接篡改检测方法,克服了传统的图像取证技术过分依赖人工经验的局限性。然而,上述传统篡改检测算法大多只能检测某一特定类型的篡改。Bayar 等^[5]首先提出一种基于约束卷积层的深度学习算法,用于解决多种篡改类型的检测问题。约束卷积层的设计是为了防止卷积神经网络(convolutional neural networks, CNN)学习代表图像内容的特征,专门用于抑制图像内容和学习篡改检测特征。Bappy 等^[6]提出一种将长短期记忆(long-short term memory, LSTM)网络与 CNN 结合的篡改检测与定位模型,其基于图像补丁同时开展分类与分割,用来定位图像中的篡改区域,利用分类和分割二者之间的依赖性以有效地提高算法的准确性。Bartusiak 等^[7]提出一种基于生成对抗网络(generative adversarial networks, GAN)的图像取证方法,通过生成对抗样本的方式对抗判别器,从而提高判别器的鲁棒性和对篡改图像的检测能力。Zhou 等^[8]基于更快速的区域卷积神经网络(faster region-convolutional neural networks, Faster R-CNN)提出了一种双流 Faster R-CNN 模型,即 RGB-N 模型,其中 RGB 流用于提取如非自然篡改边界的篡改特征;噪声流利用隐写分析模型(steganalysis rich model, SRM)^[9]从输入图像中提取噪声信息;然后利用噪声流分析篡改图像中的噪声特征;双流互补,最终检测出篡改痕迹并且能够鉴别篡改类型。Wei 等^[10]基于 RGB-N 提出一种结合边缘检测的 Faster R-CNN 图像篡改检测模型,即 RGB-E 模型,它使用边缘检

测算子捕捉真实区域与篡改区域边缘之间的一致性,通过学习篡改图像的特征进行篡改检测与定位。

本文提出了一种将特征金字塔网络(feature pyramid networks, FPN)^[11]模型与级联区域卷积神经网络(cascade region-convolutional neural networks, Cascade R-CNN)^[12]模型结合的 FCR-CNN 模型。该模型将 FPN 提取到的多尺度篡改特征输入到区域建议网络(region proposal network, RPN)^[13]层,得到目标区域建议框;并将区域建议框输入到 Cascade R-CNN 的 3 阶段检测器进行检测,最终获得分类结果。利用 FCR-CNN 模型对图像拼接和复制-移动伪造 2 种篡改类型进行了检测与定位实验,均得到了较为理想的结果。

1 相关工作

1.1 FPN 模型

FPN 模型^[11]是一个横向连接的自上而下的结构,在多尺度目标检测任务中表现出色,尤其是小目标检测任务。深层卷积网络不同的深度对应应有不同层次的语义特征以及分辨率,FPN 模型正是基于此将较深层的低分辨率高语义信息与较浅层的高分辨率低语义信息通过自上而下的方式连接,构建出各种尺度的高级语义特征。FPN 模型接收任意尺寸的单尺度图像作为输入,图像经过自下而上的卷积神经网络以完全卷积的方式生成适当大小的特征图,得到特征金字塔,最后在特征金字塔层次结构的每一层都进行目标预测。FPN 模型结构如图 1 所示。

自下而上的途径是 CNN 的前馈过程。在前馈过程中,抽取每个阶段的最后一层输出作为特征构成特征金字塔。例如,在残差网络(residual neural networks, ResNet)中,将每个阶段的最后一个残差块的特征激活输出表示为 $\{C_1, C_2, C_3, C_4\}$,分别对应于 Conv2, Conv3, Conv4 和 Conv5 的输出,其步长相对于输入图像分别为 $\{4, 8, 16, 32\}$ 。

自上而下的途径通过上采样完成。具体来说,

图 1 FPN 模型结构

对特征图进行上采样,使来自高层特征金字塔分辨率较低但语义特征较强的特征图获得高分辨率的特征.横向连接将 CNN 每一层输出的特征图与经过上采样得到的特征进行融合,得到新的特征图,表示为 $\{P_1, P_2, P_3, P_4\}$. 新的特征图因融合了不同层的特征,从而达到增强特征的目的.

FPN 模型借助卷积网络的多尺度金字塔结构提取多层篡改特征,但没有考虑后续检测器的性能,因而检测精度有限.

1.2 Cascade R-CNN 模型

Cascade R-CNN 模型^[12]是一个多阶段目标检测结构,它由 3 个检测器级联组成.这 3 个检测器的交并比(intersection over union, IoU)阈值逐级增加,这样做可以提高样本质量.当样本经过检测器之后,不会有太多的样本被丢弃,从而保证了样本数量,避免出现过拟合问题. Cascade R-CNN 模型是基于 Faster R-CNN^[13]的扩展工作,对于 Faster R-CNN,在两阶段的目标检测任务中,第 1 阶段为区域提议子网 H_0 ,通常用来获取初步的目标建议框;第 2 阶段为感兴趣区域(region of interest, RoI)检测子网 H_1 ,将第 1 阶段产生的建议框输入到 H_1 中进行处理,每个建议框将获得一个分类得分 C_1 和一个边界回归 B_1 ,如图 2 所示.

图 3 所示 Cascade R-CNN 模型是级联检测结构,第 1 阶段的检测器得到的边界回归 B_1 将输入到第 2 阶段检测器的 H_2 . 以此类推,将 B_2 输入到 H_3 中,每一个检测器都以上一个检测器的边界回归作为输入进行训练.因此,每阶段的检测器都是分开按顺序依次进行训练的.但 Cascade R-CNN 模型仅使用卷积层的最后一层作为输出,在一定程度上影响了检测性能.

图 2 Faster R-CNN 模型结构

图 3 Cascade R-CNN 模型结构

2 FCR-CNN 模型

本文提出的 FCR-CNN 模型主要包含卷积层、RPN 层以及级联回归层 3 个模块.卷积层是多尺度特征金字塔结构,由 FPN 模型实现,也可称为 FPN 层.首先将任意尺寸的单尺度篡改图像输入到 FPN 层中提取多尺度篡改特征,并将其输入到 RPN 层,获得区域提议框和区域分类;将区域提议输入到级联检测器的第 1 阶段检测器中,得到分类 C_1 和边界回归 B_1 ,再将边界回归 B_1 输入到第 2 阶段的检测器中,得到分类 C_2 和边界回归 B_2 .以此类推,最终第 3 阶段的检测器将输出分类 C_3 和边界回归 B_3 .通过级联 3 阶段的检测器,每一阶段将对上一阶段的回归结果进行优化,达到提高检测精度的目的. FCR-CNN 模型的结构如图 4 所示.

图4 FCR-CNN 模型结构

2.1 卷积层

FCR-CNN模型使用多层金字塔结构的FPN模型^[11]作为特征提取器。在图像篡改检测任务中,被篡改的区域存在多尺度问题,而FPN模型在多尺度目标检测任务中有良好的表现。FPN模型利用深层卷积网络所固有的多层金字塔层次结构,通过横向连接的自上而下的结构,利用较深层的低分辨率高语义信息与较浅层的高分辨率低语义信息构建出各种尺度的高级语义特征。FPN模型的输入为任意尺寸的单比例篡改图像,经过完全卷积,输出为多尺度的篡改特征映射。FPN模型提取特征的过程及相应的后续步骤如图5所示。

图5 FPN 工作流程及相关后续步骤

2.2 RPN 层

RPN^[13]是一个全卷积网络,并可以对其进行端到端的训练。RPN可以接收任意大小的图像作为输入,通过在输入的特征图上滑动小窗口来输出高质量的、带有目标概率的矩形建议框。每个滑动窗口与输入特征图进行 3×3 卷积,然后将卷积后

的特征图映射到低维特征,该特征再经过2个并行的 1×1 卷积,分别得到边界框回归和用于表示该建议框与所框选目标之间契合度的分类得分。每个滑动窗口的位置会同时预测 k 个区域建议框,将其参数化得到锚点。每个锚点的形成是由不同尺寸的面积和宽高比这2个因素共同决定的,这些锚点可以覆盖不同形状的目标。RPN模型结构如图6所示。

图6 RPN 模型结构

FCR-CNN模型将FPN层输出的多尺度特征图输入到RPN中。在特征金字塔的每一级都增加卷积和2个并行的卷积,称之为头部。由于头部在特征金字塔层次的每一层上密集滑动,因此不需要在特征金字塔的每一层上设置多尺度的锚点,而是为每一层设置一个单一比例的锚点。在 $\{P_2, P_3, P_4, P_5, P_6\}$ 上,分别设置具有 $\{32^2, 64^2, 128^2, 256^2, 512^2\}$ 个像素的区域,每一层采用宽高比为 $\{1:2, 1:1, 2:1\}$ 的锚点,特征金字塔共可以得到15个锚点。

2.3 级联回归层

FCR-CNN模型使用Cascade R-CNN模型^[12]中的3阶段级联检测结构作为回归层,这3个检测结构的IoU阈值逐级增加。根据文献[12]可知,输入到检测器中的建议框的IoU值在经过边界回归之后有所提高,再将具有新的IoU值的建议框输入到下一个IoU值更高的检测器中进行训练,又会得到新的具有更高IoU值的建议框,测试时使用相同

的级联结构. 这样不仅可以改善 IoU 值, 获得质量更好的样本, 提高检测精度, 还能防止过拟合以及解决训练和测试阶段不同 IoU 值之间不匹配的问题. 本文所使用的 3 阶段检测器的 IoU 值分别为 {0.5, 0.6, 0.7}, 每个检测器都以上一个检测器的边界回归作为输入进行训练, 因此, 每阶段的检测器都是分开按顺序依次进行训练的. 级联检测过程如图 7 所示, 其中 bbx 表示边界回归框, cls 表示分类.

图 7 FCR-CNN 级联检测过程

2.4 损失函数

(1) 边界回归损失

边界回归可以进一步修正候选框, 使目标建议框更加接近真实目标区域, 从而降低边界框损失. 对于边界框 $b = (b_x, b_y, b_w, b_h)$, 使用回归函数 $f(x, b)$ 将边界框 b 回归到更加准确的目标框 g , 边界框损失定义为

$$R_{loc}[f] = \sum_{i=1}^N L_{loc}(f(x_i, b_i), g_i).$$

其中, L_{loc} 在 R-CNN^[14]中是 L_2 损失函数, 在 Faster R-CNN^[13]中为平滑的 L_1 损失函数.

文献[15-17]认为, 单一的回归过程不能保证准确定位; 相反, 对回归函数 f 迭代使用, 对边界框 b 进行后处理, 即 $f'(x, b) = f \circ f \circ \dots \circ f(x, b)$.

(2) 分类损失

分类器 $h(x)$ 将图像补丁 x 分配给 $M+1$ 个类之一; 其中类别 0 表示背景, 其他类表示要被检测的目标. $h(x)$ 是类的后验分布上的 $M+1$ 维估计, 即 $h_k(x) = p(y=k|x)$; 其中 y 是类标签. 在给定的训练集 (x_i, y_i) 上通过最小化分类损失

$$R_{cls}[h] = \sum_{i=1}^N L_{cls}(h(g_i), b_i)$$

来学习. 其中, L_{cls} 是分类交叉熵损失函数^[18].

(3) 总损失

在级联的每个阶段的检测器都包含一个分类器和一个优化 IoU 值的回归器. 边界框通常包含目标和背景, 通过 IoU 的值认定边界框是目标还是背景, 即当 IoU 值高于某一阈值 μ 时, 则认为该边界

框中包含的是目标; 否则, 认为背景. 假设类标签 y 是阈值 μ 的函数, 则有

$$y = \begin{cases} g_y, & \text{if } \text{IoU}(x, g) \geq \mu \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

其中, g_y 是地面真实目标 g 的类标签. IoU 阈值 μ 决定了检测器的质量.

在级联的检测器中, IoU 阈值 μ 满足 $\mu^t \geq \mu^{t-1}$, FCR-CNN 模型的总损失函数定义为

$$L(x^t, g) = L_{cls}(h_t(x^t), y^t) + \lambda[y^t \geq 1]L_{loc}(f_t(x^t, b^t), g).$$

其中, $b^t = f^{t-1}(x^{t-1}, b^{t-1})$; $\lambda=1$ 是权衡系数; $[\cdot]$ 是指示函数; y^t 是 x^t 的类标签, 由式(1)得到.

3 实验结果及其分析

本文在公开的标准图像篡改数据集 NC2016, CASIA 和 Columbia 上分别对 RGB-N^[8], RGB-E^[10], Cascade R-CNN, FPN 和 FCR-CNN 这 5 种模型开展了对比实验.

3.1 参数设置

在本文的实验过程中, 初始学习率设置为 0.0005, 当迭代次数达到 80000 和 120000 时减少 10 倍. RPN 层锚点设置如下: 为生成的 5 层特征金字塔结构 $\{P_2, P_3, P_4, P_5, P_6\}$ 的每层分别分配 $\{32^2, 64^2, 128^2, 256^2, 512^2\}$ 个像素的区域, 每层均采用 3 种宽高比 $\{1:2, 1:1, 2:1\}$, 每幅图像使用 256 个 RoI.

3.2 预训练

当前公开的标准图像篡改数据集较小, 不足以训练深度网络. 为了解决这一问题, 使用 COCO 合成篡改数据集^[8]对模型进行预训练. 它是从 COCO 数据集^[19]的图像和标注自动创建而来的, 使用分割标注从 COCO 数据集中随机选择对象, 然后将其复制并粘贴到其他图像中. 在 COCO 合成篡改数据集中选择 90% 的图像制作成训练集, 其余为测试集, 其分配如表 1 所示.

表 1 训练集和测试集分配

数据集	图片数量
训练集	10650
测试集	1183

为方便后续实验对比, 本文在预训练 Cascade R-CNN 模型、FPN 模型以及 FCR-CNN 模型时, 使用的参数以及预训练数据集大小均相同. 表 2 所示

表2 不同模型在COCO合成篡改数据集上的mAP

模型	mAP
Cascade R-CNN ^[12]	0.8124
FPN ^[11]	0.8254
FCR-CNN	0.8582

为3种模型在COCO合成篡改数据集测试获得的均值平均精确率(mean average precision, mAP).

由表2可知, Cascade R-CNN模型与FPN模型在图像篡改检测任务中具有不错的检测精度,但相比较而言, FCR-CNN模型的表现更加优异. 考虑FPN模型的特征金字塔层次结构能够提取到更丰富的篡改特征, 而3阶段级联检测器通过逐级提高IoU值的方式获得更高质量的检测器, 将这2个模型结合起来组成的FCR-CNN模型能够得到更加准确的检测结果.

3.3 在标准数据集上的实验结果

目前公开的标准图像篡改数据集的数据量较小, 为解决这一问题, 考虑使用数据增强的方式扩大数据量, 使网络训练的模型更为准确, 从而提高检测精度. 本文使用与文献[10]相同的图像翻转方式对数据进行增强, 这种数据增强方式不会改变图像的分辨率和内容, 将数据量扩大为原有数据量的2倍, 从而提高模型的检测精度. 3个标准数据集数据分配情况如表3所示.

表3 3个标准数据集数据分配情况

数据集	NC2016	Columbia	CASIA
训练集	404		5123
测试集	160	180	921

本文分别在3个公开的标准数据集上对预训练模型进行微调, 使用 F_1 分数、精确度(precision)和召回率(recall)作为评价指标, 分别如表4、表5和表6所示.

表4的数据表明, 在3个公开的标准图像篡改数据集上, FCR-CNN模型的 F_1 分数高于其他4种模型. 这是由于FPN独有的特征金字塔层次结构

表4 不同模型在标准图像篡改数据集上的 F_1 分数

模型	NC2016	Columbia	CASIA
RGB-N ^[8]	0.9123	0.7467	0.5655
RGB-E ^[10]	0.9533	0.7514	0.5794
Cascade R-CNN ^[12]	0.9477	0.7403	0.6169
FPN ^[11]	0.9595	0.7599	0.6254
FCR-CNN	0.9783	0.8468	0.6629

表5 不同模型在标准图像篡改数据集上的精确度

模型	NC2016	Columbia	CASIA
RGB-N ^[8]	0.8830	0.7044	0.4996
RGB-E ^[10]	0.9503	0.7644	0.5158
Cascade R-CNN ^[12]	0.9333	0.8906	0.6728
FPN ^[11]	0.9565	0.8389	0.6982
FCR-CNN	0.9693	0.9216	0.7326

表6 不同模型在标准图像篡改数据集上的召回率

模型	NC2016	Columbia	CASIA
RGB-N ^[8]	0.9437	0.7944	0.6515
RGB-E ^[10]	0.9563	0.7389	0.6608
Cascade R-CNN ^[12]	0.9625	0.6333	0.5695
FPN ^[11]	0.9265	0.6844	0.5664
FCR-CNN	0.9875	0.7833	0.6110

能够提取到更丰富的篡改特征, 回归层使用的3阶段级联检测器通过逐级提高IoU值的方式获得更高质量的检测器. CASIA数据集中图像的背景更加复杂, 且图像包含了更多的后处理操作, 在该数据集上, 较FPN模型和Cascade R-CNN模型而言, FCR-CNN模型 F_1 分数分别提升了6.0%和7.5%.

3.4 篡改定位

篡改定位也是近年来图像篡改检测任务中的研究热点^[6,20-21]. 在实际应用中, 仅检测图像是否发生篡改并不能满足需求, 往往还需要检测出图像中哪里发生了篡改. 本文从2个公开的标准图像篡改数据集NC2016和CASIA中选取拼接图像和复制-移动伪造图像各2幅, 进行篡改区域定位, 验证FCR-CNN模型在定位篡改区域方面的表现.

图8所示为拼接图像和复制-移动伪造图像示例, 其中包括真实图像、篡改图像及其二值篡改区域.

图9所示为RGB-N, RGB-E, Cascade R-CNN, FPN和FCR-CNN这5种模型对拼接图像分别进行篡改定位的结果. 第1行是有1处(船只)图像拼接情况下的定位检测结果; 第2行是有2处(自由女神像与大猩猩及其影子)图像拼接情况下的定位检测结果. 实验结果表明, 在发生1处拼接的情况下, 5种模型都能正确检测和定位出篡改区域. 但RGB-N模型未能完整地框选出船只的右侧部分; Cascade R-CNN模型未能完整框选出船只的左侧部分; FPN模型未能完整地框选出船只的右侧部分; 而RGB-N模型与本文的FCR-CNN模型能够完整框选出整个拼接区域.

在发生2处图像拼接的情况下, RGB-N模型只

图 8 拼接图像和复制-移动伪造图像示例

能完整地框选出大猩猩及其影子的部分, 而没有检测出自由女神像处的拼接; RGB-E 模型正确框选拼接的自由女神像和大猩猩, 但未完整框选出伪造的影子区域; Cascade R-CNN 模型和 FPN 模型能检测和定位出 2 处篡改, 但都未能完整地框选出自由女神像和大猩猩及其影子; 而 FCR-CNN 模型能够检测出 2 处拼接, 但只完整地框选出大猩猩, 未能完整地框选出大猩猩的影子和自由女神像的左侧部分。

图 10 所示为 RGB-N, RGB-E, Cascade R-CNN,

FPN 和 FCR-CNN 这 5 种模型对复制-移动伪造图像进行篡改定位的结果。第 1 行是有 1 处(铁塔)复制-移动伪造情况下的定位检测结果; 第 2 行是有 2 处(左上角的云与右上角的云)复制-移动伪造情况下的定位检测结果。实验结果表明, 在发生 1 处复制-移动伪造的情况下, 5 种模型都能检测和定位出篡改区域, 但 RGB-N 模型未能完整地框选出铁塔的上侧部分; Cascade R-CNN 模型虽完整地框选出了篡改区域, 但候选框过大; RGB-E, FPN 模型及 FCR-CNN 模型能够完整地框选出整个篡改区域。

图 9 5 种模型对拼接图像的篡改定位结果

图10 5种模型对复制-移动伪造图像的篡改定位结果

在发生2处复制-移动伪造的情况下,RGB-N模型和FPN模型只能完整地框选出左上角部分的云朵,未检测出右上角云朵处的篡改;RGB-E模型正确框选出图像左侧伪造云朵和右侧伪造云朵区域,但右侧候选框过大;Cascade R-CNN模型仅正确且完整地框选出左上角的云朵,错误地将真实人物区域也框选出来;而FCR-CNN模型能够完整地检测出2处云朵的篡改。

根据图9和图10的图像拼接和复制-移动伪造图像检测与定位结果,RGB-E模型和FCR-CNN模型的篡改定位在这5种模型中具有最优的检测性能。对于仅发生单个篡改的篡改图像,5种模型均能正确地检测出篡改区域,但相较于RGB-E模型和FCR-CNN模型,RGB-N模型、FPN模型和Cascade R-CNN模型框选的篡改区域不够完整。对于发生2处伪造的篡改图像,RGB-N模型仅能检测出其中一个篡改区域,FPN模型和Cascade R-CNN

模型对于拼接图像的检测结果要优于复制-移动伪造图像的检测结果,但检测结果都不是最理想的,RGB-E模型和FCR-CNN模型均能定位出2个篡改区域,但FCR-CNN模型在复制-移动篡改图像上的定位结果更加准确。

在发生2处拼接的拼接图像中,拼接边缘不明显,因此导致FCR-CNN模型检测时,2处拼接区域仅由一个框被框选出来。而在发生2处篡改的复制-移动伪造图像中,伪造边缘较明显,因此每个篡改区域均可以由FCR-CNN模型得到准确的预测框。

3.5 鲁棒性验证

为验证本文模型的鲁棒性,分别测试了5种模型在CASIA数据集上经历不同JPEG压缩(品质因子分别为85,70,55)攻击、高斯模糊(滤波器大小为 3×3 ,方差分别为0.5,1.0,1.5)和高斯白噪声攻击(均值为0,方差分别为5,10,15)的检测结果。其 F_1 分数如图11所示。

图11 5种模型在不同后处理技术下的 F_1 分数

图11的结果表明,当加入JPEG压缩攻击后,RGB-E模型的抗JPEG压缩性能更优,其他4种模型的 F_1 分数变化趋势与幅度基本一致。当加入高斯模糊攻击后,FCR-CNN模型较FPN模型与Cascade

R-CNN模型的 F_1 分数变化幅度更平缓一些,但仍不及RGB-E模型与RGB-N模型的抗高斯模糊性能。当加入高斯白噪声攻击后,FCR-CNN模型的抗高斯白噪声性能不及RGB-E,RGB-N及FPN模型。

4 结 语

本文提出了一种将目标检测网络 FPN 模型与 Cascade R-CNN 模型结合的图像篡改检测模型, 即 FCR-CNN 模型. 该模型通过使用多尺度特征金字塔结构的 FPN 提取多尺度的篡改特征, 由 RPN 输出篡改分类和区域建议框, 再将边界回归结果输入到 3 阶段 IoU 值逐级增大的级联检测器进行检测. 本文使用 COCO 合成篡改数据集对模型进行预训练, 在 3 个公开的标准数据集上对 FCR-CNN 模型进行了评估, FCR-CNN 模型在篡改定位以及鲁棒性验证中, 均有较好的表现. 但本文仅使用拼接图像和复制-移动伪造图像作为研究对象, 在未来的研究中, 计划将其他图像篡改类型作为重点, 以学习更丰富的篡改特征.

参考文献(References):

- [1] Cozzolino D, Poggi G, Verdoliva L. Splicebuster: a new blind image splicing detector[C] //Proceedings of the IEEE International Workshop on Information Forensics and Security. Los Alamitos: IEEE Computer Society Press, 2015: Article No.7368565
- [2] 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
- [3] Bianchi T, Piva A. Detection of nonaligned double JPEG compression based on integer periodicity maps[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(2): 842-848
- [4] Sun Peng, Yang Hongchen, Dai Xuejing, *et al.* An authentication method for splicing images with color temperature estimation[J]. Journal of Computer-Aided Design & Computer Graphics, 2012, 24(9): 1193-1196(in Chinese)
(孙鹏, 杨洪臣, 代雪晶, 等. 拼接篡改图像的色温估计取证方法[J]. 计算机辅助设计与图形学学报, 2012, 24(9): 1193-1196)
- [5] Bayar B, Stamm M C. A deep learning approach to universal image manipulation detection using a new convolutional layer[C] //Proceedings of the ACM Information Hiding and Multimedia Security Workshop. New York: ACM Press, 2016: 5-10
- [6] Bappy J H, Roy-Chowdhury A K, Bunk J, *et al.* Exploiting spatial structure for localizing manipulated image regions[C] // Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 4980-4989
- [7] Bartusiak E R, Yarlagadda S K, Güera D, *et al.* Splicing detection and localization in satellite imagery using conditional GANs[C] //Proceedings of the IEEE Conference on Multimedia Information Processing and Retrieval. Los Alamitos: IEEE Computer Society Press, 2019: 91-96
- [8] Zhou P, Han X T, Morariu V I, *et al.* Learning rich features for image manipulation detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 1053-1061
- [9] Boroumand M, Chen M, Fridrich J. Deep residual network for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1181-1193
- [10] Wei X Y, Wu Y R, Dong F M, *et al.* Developing an image manipulation detection algorithm based on edge detection and faster R-CNN[J]. Symmetry, 2019, 11(10): 1223
- [11] Lin T Y, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 2117-2125
- [12] Cai Z W, Vasconcelos N. Cascade R-CNN: delving into high quality object detection[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 6154-6162
- [13] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[C] // Proceedings of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99
- [14] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2014: 580-587
- [15] Gidaris S, Komodakis N. Object detection via a multi-region and semantic segmentation-aware CNN model[C] // Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 1134-1142
- [16] Gidaris S, Komodakis N. Attend refine repeat: active box proposal generation via in-out localization[C] //Proceedings of the British Machine Vision Conference. Guildford: BMVA Press, 2016: 90.1-90.13
- [17] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [18] Zagoruyko S, Lerer A, Lin T Y, *et al.* A multipath network for object detection[C] //Proceedings of the British Machine Vision Conference. Guildford: BMVA Press, 2016: 151-162
- [19] Lin T Y, Maire M, Belongie S, *et al.* Microsoft COCO: common objects in context[C] //Proceedings of the European Conference on Computer Vision. Heidelberg: Springer, 2014: 740-755
- [20] Mazaheri G, Mithun N C, Bappy J H, *et al.* A skip connection architecture for localization of image manipulations[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 119-129
- [21] 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