基于双通道 R-FCN 的图像篡改检测模型

田秀霞[1)] 李华强[1)] 张 琴[2)] 周傲英[3),4)]

[1)](上海电力大学计算机科学与技术学院 上海 200090)

[2)](国网兰州供电公司互联网部信息通信安全实验室 兰州 730050)

[3)](华东师范大学数据科学与工程研究院 上海 200062)

[4)](华东师范大学软件学院上海市高可信计算重点实验室 上海 200062)

摘 要 随着大数据时代的到来和图像编辑软件的发展,恶意篡改图片的数量出现井喷式增长,为了确保图像的真实性,众多学者基于深度学习和图像处理技术提出了多种图像篡改检测算法.然而,当前提出的绝大多数方法在面对大量图片的情况下,篡改检测速率较低且小面积篡改区域检测效果较差.为了有效解决这些问题,本文首次将基于区域的全卷积网络(Region-based Fully Convolutional Networks,R-FCN)引入双通道篡改检测网络,通过彩色图像通道提取图像的表层特征,使用隐写分析通道挖掘图像内部的统计特征,并利用双线性池化层将两个通道的信息融合,构建了一种面向实际应用场景的图像篡改检测模型.其中,利用 R-FCN 中位置敏感得分图提高图像篡改检测效率,使用双线性插值算法提高小面积篡改区域的检测率.通过在国际主流的标准图像篡改数据集上进行实验,有效地验证了该模型的图像篡改检测速率相比当前最新模型提高 2.25 倍,检测精度提升 1.13%到 3.21%,本文提出的模型是一种更加高效而精准的图像篡改检测模型.

关键词 图像篡改检测;深度学习;双通道网络;基于区域的全卷积网络;双线性插值

中图法分类号 TP751;TP183 **DOI 号** 10.11897/SP.J.1016.2021.00370

Dual-Channel R-FCN Model for Image Forgery Detection

TIAN Xiu-Xia[1)] LI Hua-Qiang[1)] ZHANG Qin[2)] ZHOU Ao-Ying[3),4)]

[1)](*College of Computer Science and Technology, Shanghai University of Electric Power, Shanghai* 200090)

[2)](*State Grid Lanzhou Power Supply Company, Ministry of Internet, Information and Communication Security Laboratory, Lanzhou* 730050)

[3)](*College of Data Science and Engineering, East China Normal University, Shanghai* 200062)

[4)](*Shanghai Key Laboratory of Trustworthy Computing, Software Engineering Institute, East China Normal University, Shanghai* 200062)

Abstract With the explosive growth of malicious tampering images, many scholars have proposed multiple image forgery detection algorithms based on deep learning and image processing technologies. Although these algorithms have achieved good results, most of them have strong limitations in practical application. In order to solve this problem, we proposed a dual-channel forgery detection model empowered by Region-based Full Convolution Network (R-FCN), which was inspired by the two-stream network. The model included two parts: RGB channel and steganalysis channel. The design of dual channel enables the model to capture more features in the image and obtain a better detection effect. First of all, the model utilized the properties of each channel to extract the image's features. The RGB channel captured the boundary, color, texture and other surface features of the image, and analyzed the tampering artifacts which were left by image forgery. Steganalysis

收稿日期:2019-08-09;在线发布日期:2020-02-15.本课题得到国家自然科学基金(面上项目;重点项目)(61772327,61532021)、国网甘肃省电力公司电力科学研究院横向项目(H2019-275)资助.**田秀霞**,博士,教授,中国计算机学会(CCF)会员,主要研究领域为数字图像篡改检测、数据库安全、隐私保护(大数据和云计算)、安全机器学习、面向电力用户的安全计算.E-mail:xxtian@shiep.edu.cn.**李华强**,硕士研究生,主要研究方向为数字图像篡改检测、深度学习、目标检测.**张 琴**,硕士,工程师,主要研究方向为信息安全、信息网络、通信技术.**周傲英**,博士,教授,博士生导师,长江学者特聘教授,国家杰出青年科学基金入选者,中国计算机学会(CCF)会员,主要研究领域为数据库、数据管理、数字化转型、教育科技(EduTech)和物流科技(LogTech)等数据驱动的应用等.

channel used the Spatial Rich Model (SRM) filter layer to extract the residual noise of the image, and analyzed the inconsistency between the real area and the tampering area. Then, the model used the Region Proposal Network (RPN) to obtain the corresponding Region of Interest (ROI) location information from the feature maps, and combined the position-sensitive ROI pooling operations to get the score maps. Finally, the model used the bilinear pooling layer to fuse the information of the two channels, and processed the relevant features to obtain the corresponding category information and location information, so as to located the tampering area. On the one hand, the proposed model uses the design of the position-sensitive score map in R-FCN, which increases the number of shared computing network layers by changing the location of the ROI pooling layer, and improves the detection efficiency. On the other hand, bilinear interpolation is used to adjusting the output size of the feature map in the feature extraction network, which alleviates the weak expression ability of model features caused by convolution operation in the feature extraction process, and improves the detection accuracy of the small tampering area. Since there was not enough data in the standard dataset to train the neural network, we pre-trained our model on the synthetic dataset. We compared our model to four state-of-the-art models on three benchmark datasets, NIST, CASIA2.0 and Columbia. The comparison models were mainly divided into two categories: one traditional image forgery detection algorithm (CFA1) and three deep learning image forgery detection algorithms (Tam-D, J-Conv-LSTM and RGB-N). We have conducted a number of experiments to verify the performance of our model. The experimental results show that the dual-channel structure and bilinear pooling layer of the model improve the detection accuracy. In order to explore the superior performance of the model, we evaluate the model with three evaluation indexes: Average precision (AP), $F1$-score and Frames Per Second (Fps). The evaluation results show that the image tamper detection rate of this model is 2.25 times higher than the current latest model, and the detection accuracy is increased by 1.13% to 3.21%, verifying our proposed image forgery detection model is more efficient and accurate.

Keywords　image forgery detection; deep learning; dual-channel network; region-based full-convolution network; bilinear interpolation

1　引　言

随着多媒体、互联网和存储技术的快速发展，数字图像的传播与存储成本变得越来越低，它已经深入应用到人类的生活、工作等各个方面. 然而，一系列图像处理和编辑软件的出现，让图像篡改的门槛大大降低，无需专业图像处理知识也能对图像进行编辑处理，这对图像的真实性、完整性构成了极大的威胁. 2007 年 10 月 12 日，陕西省林业厅公布了一系列“野生华南虎”照片，但专家与网友对照片的真实性和完整性提出了质疑. 经调查发现，证实照片上的华南虎出自一张年画. 这仅是众多图像伪造问题中的一个例子，如果图片伪造被大面积地滥用，将会给国内甚至全世界舆论的导向造成极大的负面影响[1].

图像篡改手段分为三种：复制移动、拼接和移除[2]. 复制移动是指在图像中复制部分区域，然后将该区域粘贴到图像中与复制区域无重叠的部分，实现图像的篡改；拼接是从一幅图像中复制部分区域，然后将该区域粘贴到另一幅图像中；移除是对图像中的像素进行修改，实现目标区域的消除. 图像篡改检测技术可分为主动篡改检测技术和被动篡改检测技术(盲检测)两大类. 主动篡改检测技术主要以数字签名技术和数字水印技术为代表，需要在图像建立时主动对图像进行预处理，如对图像的哈希(Hash)串进行计算或在图像中嵌入特征信息，这些特征信息是具有特定意义的水印或图像等，检验时会对嵌入信息的完整性进行验证，从而判断图像是否发生过篡改；被动篡改检测技术则不需要事先为图像添加任何特征信息，仅凭借图像本身的统计信息或物理特性便可达到篡改检测目的[3]. 由于主动

篡改检测技术的使用具有很强的局限性，数字水印的抗攻击性也有待加强，所以被动篡改检测技术逐渐成为研究的热点.

目前，研究人员就被动篡改检测技术提出了多种算法，其代表性工作主要有文献[4-5]. Bianchi 等人[4]提出了一种概率模型，来估计 JPEG 图像中不同区域的离散余弦变换(Discrete Cosine Transform，DCT)系数和量化因子，进而判断出每个 DCT 块被篡改的概率. 当面对不同类型的篡改方式时，该模型的表现有好有坏，鲁棒性比较差. Rao 等人[5]率先将深度学习技术用于篡改检测中，模型使用空间富模型(Spatial Rich Models，SRM)作为预处理层，输出采用支持向量机(Support Vector Machine，SVM)进行二分类. 但该模型只能判断图像是否发生过篡改，无法做到对篡改区域的定位，在应用中存在很大的局限性.

本文利用双通道网络提取丰富的图像特征，并引入 R-FCN[6]提高模型的效率，针对现有图像篡改检测速率低和小面积篡改区域检测精度的不足的问题，提出了一种基于双通道 R-FCN 的图像篡改检测模型.

本文的创新点包括：

(1) 使用 R-FCN[6]对模型进行构建与优化. R-FCN 中移除了全连接层，只保留可学习的卷积层，这样可以进行端到端的训练，而且全卷积网络的设计提高了共享计算的网络层数，相比 Faster R-CNN[7]，模型的效率得到大幅的提升；

(2) 提出一种基于图像分辨率调整的小面积篡改区域检测技术. 为了提高图像篡改检测中小面积篡改区域的检测精度，本文采用双线性插值法，对特征提取网络中每一个堆叠卷积层输出的特征图尺寸进行调整，缓解在特征提取过程中造成的目标缺失问题；

(3) 相较于现有的图像篡改检测方法，本文方法在篡改图像中包含细微篡改痕迹时，可以更好地对篡改区域进行定位，检测效率的提升也使模型具有实用价值.

本文第 2 节对国内外在图像被动篡改领域的相关工作进行总结；第 3 节对本文用到的预备知识进行介绍；第 4 节给出本文模型结构，并对其进行详细地描述；第 5 节为实验的结果与对比分析；第 6 节总结全文并对未来的研究工作进行展望.

2 相关工作

图像被动篡改检测方法分为两类：基于图像统计信息的传统篡改检测算法和基于神经网络的深度学习篡改检测算法. 传统图像篡改检测算法是根据图像的特性设计人工特征，通过计算图像残差、邻域像素的相关系数、方差、直方图等特征，检测数字图像是否以某种方式发生了篡改；深度学习篡改检测算法首先对网络的输入数据进行相关的预处理，随后使用卷积神经网络(Convolutional Neural Networks，CNN)[8]对图像特征进行提取，最后进行篡改区域的定位与分类，实现端到端(end-to-end)的网络框架结构并以反向传播(back propagation)的方式对网络参数进行更迭. 传统篡改检测算法与深度学习篡改检测算法的框架对比如图 1 所示.

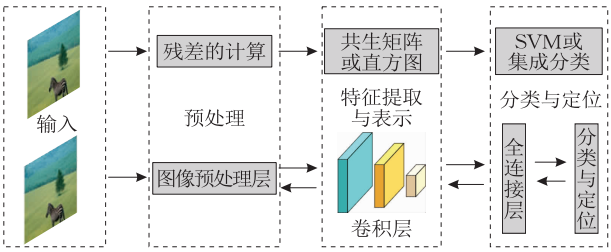


图 1 传统篡改检测算法框架(上)与
深度学习篡改检测算法框架(下)

2.1 传统篡改检测算法

为了在图像中检测篡改区域，研究者们使用多种方法进行尝试，目前比较热门的传统图像篡改检测方法可以分为：基于像素的图像篡改检测方法、基于成像设备的图像篡改检测方法和基于组合器的图像篡改检测方法.

基于像素的特性，2008 年 Mahdian 等人[9]基于内插信号及其导数包含特定可检测周期的特性，提出了一种能够找到重采样和插值痕迹的图像被动篡改检测方法，同时该方法在图像安全性和身份验证等领域都有着很强的应用价值. 2010 年 Stamm 等人[10]基于内在指纹的检测方法，通过全局图像和局部图像的对比度增强实现对图像的篡改检测，该方法可以很好地检测出经过对比度增强操作的篡改图像. 2014 年 Hashmi 等人[11]将加速鲁棒特征(Speeded Up Robust Features，SURF)变换和多种小波变换相结合来检测篡改图像，与一些算法不同，该算法是在整张图像上提取特征，而不是将图像划分为块区域提取特征.

基于成像设备的特性，2010 年 Hsu 等人[12]使用相机响应函数(Camera Response Function，CRF)实现图像篡改检测，使用局部平面辐照度点(Locally Planar Irradiance Points，LPIPs)的几何不变量对每

个自动分割区域进行 CRF 估计，计算基于 CRF 的交叉拟合和局部图像特征并将其馈送到统计分类器，最后推断图像的真实性. 由于单传感器数码相机是通过内部的色彩滤镜矩阵(Color Filter Array，CFA)插值得到彩色图像，而每个相机的 CFA 结构与算法不相同，其输出图像相邻像素间的线性关系也不相同. 2012 年 Ferrara 等人[13]利用色彩滤镜矩阵模式对检测图像进行采样后的重新插值，并将重新插值的图像与原图像进行比较，获得估计误差，最后根据两类像素不同的估计误差对图像进行篡改检测.

大多数传统方法仅针对某种特定篡改方式的图像进行检测，而图像篡改的方式是多种多样的，于是有研究者提出基于组合器的图像篡改检测方法. 2014 年 Gaborini 等人[14]提出了一种融合三个独立检测器的图像篡改检测方法，三个检测器分别为：基于光响应不均匀性(Photo Response Non-Uniformity，PRNU)的检测器、基于块匹配的检测器和图像来源检测器，这种组合在一定程度上提高了篡改检测的精度. 2017 年 Li 等人[15]提出了一种集合框架，该框架包含一个基于统计特征的检测器和一个复制移动方式的篡改检测器，最终通过阈值处理的方式检测图像的篡改区域. 基于组合器的图像篡改检测方法虽然提高了检测的精度，但由于该方法集成了多种检测器，所以提升了模型的复杂度.

2.2　深度学习篡改检测算法

基于数据驱动方式的深度学习方法在众多计算机视觉与图像处理任务上表现出优异的性能，因此很多研究者也将深度学习应用到图像篡改检测中，比较热门的几个方向有：通过改进网络结构提高检测精度、基于成像设备的相关性设计网络模型和利用隐写分析的特性获得更好的检测效果.

通过优化网络结构可以获得更好的检测性能，2016 年 Bayar 等人[16]提出了一种新的卷积层结构，来捕获图像篡改时图像中相邻像素关联性的变化，同时自适应地学习篡改特征，并最大程度地压缩图像内容对篡改检测的影响. 2017 年 Bappy 等人[17]提出了一种高置信度的网络模型，该模型将卷积神经网络和长短期记忆网络(Long Short-Term Memory，LSTM)融合，通过捕捉篡改区域与非篡改区域的边界差异，实现图像篡改检测.

同时，深度学习方法也应用在基于成像设备的相关性上. 2016 年 Baroffio 等人[18]首次提出利用卷积神经网络实现基于相机源的图像篡改检测，该模型可以直接从获取的图片中学习表征每个相机的特征，进一步判断图像是否发生过篡改. 2017 年 Bondi 等人[19]利用卷积神经网络，对不同型号相机所拍摄图片上遗留的特征痕迹进行检测，根据不同相机遗留噪声的不一致性，实现篡改区域的定位.

近年来，研究者们使用隐写分析方法对检测框架进行完善，通过分析图像相邻像素的局部噪声特征，可以更好地比对篡改区域与非篡改区域的不一致性. 2017 年 Cozzolino 等人[20]将卷积神经网络与 SRM 特征结合，实现图像篡改区域的定位；2018 年 Zhou 等人[21]提出一种基于 Faster R-CNN 架构的双流网络，该网络融合了 RGB 流与噪声流的特征，进一步提升了图像篡改的检测精度.

综上，现有的图像篡改检测方法仍存在以下问题：(1) 大多数传统方法由于特征设计的专一性，在图像的篡改方式未知时，算法的鲁棒性比较差；(2) 传统的基于组合器的图像篡改检测方法虽然提高了模型的鲁棒性，但整体框架过于繁琐，不够精简；(3) 基于深度学习的篡改检测方法虽然在模型的鲁棒性和结构上都有了一定程度的优化，但与传统方法相比，检测的精度还需继续提升，尤其是面对小面积篡改区域的检测时；(4) 深度学习方法在检测速度上有一定的优化空间. 针对以上问题，本文从实际应用的角度出发，研究了高效率的图像篡改检测模型，并提高了小面积篡改区域的检测率.

3　预备知识

3.1　深度学习目标检测模型

目标检测可以通过对图像或视频的识别，将要检测的目标进行分类与定位，在视频监控、自动驾驶、人脸识别等多个领域都有应用. 近几年，因为计算机运算能力的指数级增长，机器学习、深度学习等人工智能技术迎来了新的春天，研究者们纷纷将其应用至目标检测领域[22].

基于深度学习的目标检测网络架构包含两个部分，第一部分是由卷积神经网络组成的特征提取部分，可以从输入图像中提取不同类型、不同层次的特征，并输出特征图；第二部分是目标检测模型的主体部分，它会利用第一部分输出的特征图，进行相关操作后实现目标的分类与定位.

根据模型候选框生成方法的不同，可将目标检测模型分为两类：第一类模型将问题分为两个阶段，首先生成区域建议(Regional Proposal)候选框，然

后根据提取的图像特征对候选区域进行分类与位置回归，其代表为 Fast R-CNN[23]、Faster R-CNN[7]、R-FCN[6]、Mask R-CNN[24]等；第二类模型则省去了区域建议的步骤，直接生成物体的类别概率和位置坐标值，通过单次检测便可得到检测结果，典型模型有 YOLO[25]、SSD[26]等. 两种模型的性能优缺点对比如表 1 所示. 考虑到 R-FCN 对小目标的检测效果要优于第二类模型，而且它在第一类模型中有速度上的优势，本文便将 R-FCN 用于图像篡改检测网络中.

表 1 模型优缺点对比

年份	模型	实时性	优点	缺点
2015	Fast R-CNN[23]	否	可以同时完成分类与定位，节省存储空间	候选框生成方法占用大部分时间
2015	Faster R-CNN[7]	是	提出 RPN 层生成候选框，实现端到端训练测试，提升了速度	模型较为复杂
2015	YOLO[25]	是	网络简单，速度较快	准确度低，小目标与多目标检测结果不好
2015	SSD[26]	是	网络简单，准确率高	对小目标检测结果不好
2016	R-FCN[6]	是	较 Faster R-CNN 大幅提升速度，精度也略微提升	模型较为复杂
2017	Mask R-CNN[24]	是	实例分割效果好，检测精度更高	模型复杂

3.2 R-FCN 目标检测模型

ResNet[27]为 R-FCN 的主干网络，它的基本组成单位是残差学习模块，其结构如图 2 所示. 残差学习模块不仅可以通过卷积层与非线性函数将原始输入映射到下一层，还允许原始的输入信息直接映射到后面的层，通过这种连接方式实现残差网络结构输入与输出的加叠，在减少计算量的同时，缓解了网络层数的增加造成的梯度消失现象.

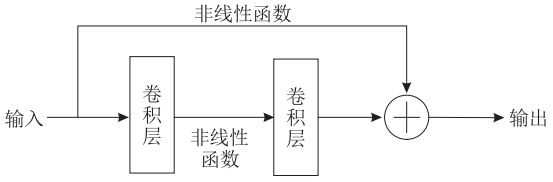


图 2 残差模块结构图

R-FCN[6]的网络结构，如图 3 所示. 首先，使用 ResNet 网络对图像进行特征提取，并生成特征图. 之后使用区域建议网络(Region Proposal Network，RPN)生成区域候选框，并对每个候选框进行背景的筛选与微调. 进行分类操作时，会在区域建议网络的基础上，通过卷积在整幅图像上为每类物体生成 $k\times k$(图 3 结构中 $k=3$)个位置敏感得分图，每个位置敏感得分图的通道数为 C(代表 $C-1$ 类的物体加 1 个背景). 对于一个大小为 $w\times h$ 的区域候选框，会将其划分为 $k\times k$ 个块，则每个块的大小为 $w\times h/k^2$，并对任意一个块 $bin(i,j)$ 执行位置敏感池化操作，其中 $0\leq i,j\leq k-1$，池化公式如下所示：

$$r_c(i,j|\Theta)=\sum_{(x,y)\in bin(i,j)}\frac{z_{i,j,c}(x+x_0,y+y_0|\Theta)}{n} \tag{1}$$

其中，$r_c(i,j|\Theta)$表示对应第 C 个类别块 $bin(i,j)$的池化响应，$z_{i,j,c}$表示块 $bin(i,j)$所对应的位置敏感得分图，(x_0,y_0)表示每个区域候选框左上角的坐标值，Θ 表示网络中所有可学习的参数. 随后，对 $k\times k$ 个块的池化响应 $r_c(i,j|\Theta)$进行均值池化，最后使用 Softmax 函数获得每个类别的概率. 位置回归的流程与分类相同，但每个位置敏感得分图的通道数由 C 个变为 4 个，分别表示区域候选框的 4 个位置参数.

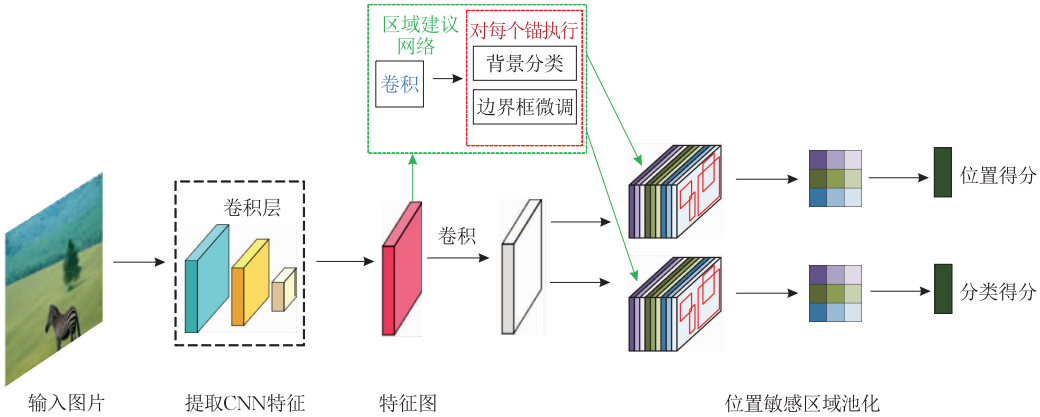


图 3 R-FCN 网络结构

3.3　空间富模型(SRM)

隐写术是一种防止别人发现秘密信息的技术，目前主要应用于对通信内容的保密. 它将特殊信息隐藏在文本、图像、视频等正常载体中，实现秘密信息的隐秘传播. 为了防止隐写术被不法分子利用，给社会带来巨大损失，研究者们在隐写分析领域上投入大量研究，对载体中是否含有秘密信息进行判断.

随着隐写术的不断发展，秘密信息的隐藏能力也在不断提高. 为了应对更高级的自适应隐写术，研究者们开始考虑更加复杂化、高维化的特征，其代表特征为基于对图像邻域复杂相关性进行建模的高阶统计量特征[28]. 由于该类特征中包含丰富的统计特性，因此该特征也被称为“富模型”特征. 代表模型 SRM[29] 的框架，如图 4 所示.

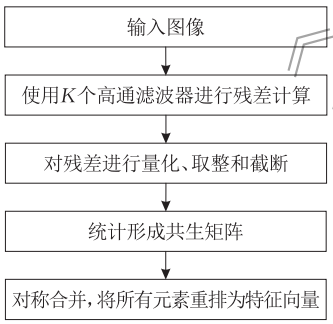


图 4　空间富模型框架

首先，使用 K 个高通滤波器对输入图像计算残差值. 因为与整幅图像相比，图像中秘密信息的嵌入所引起的扰动是很小的，而扰动相对于残差值却很强，所以需要使用高通滤波器对图像进行卷积获得残差值. 残差值计算的公式如下所示：

$$R_{i,j}=\varphi_{i,j}(\alpha_{i,j})-cX_{i,j} \tag{2}$$

其中，$R_{i,j}$ 表示图像 (i,j) 位置的残差值，$X_{i,j}$ 表示图像 (i,j) 位置的像素值，$\alpha_{i,j}$ 表示图像 (i,j) 位置邻域像素的集合，$\varphi_{i,j}$ 表示对图像 (i,j) 位置邻域像素进行卷积操作. 随后，对残差进行截断、量化和取整. 通过截断，可以降低特征维度，从而避免高残差值区域的冗余特征对模型性能造成过多的影响. 量化主要是针对截断操作造成的高残差值区域隐藏信息的缺失进行弥补，关于量化系数选择的公式如下所示：

$$q\in\begin{cases}\{c,1.5c,2c\} & \text{for } c>1\\ \{1,2\} & \text{for } c=1\end{cases} \tag{3}$$

其中，c 由滤波器的类型决定. 截断、量化和取整操作的公式如下：

$$R_{i,j}\leftarrow trunc_T\left(round\left(\frac{R_{i,j}}{q}\right)\right) \tag{4}$$

其中，q 表示对残差值进行量化的步长，$round(\cdot)$ 表示对残差值进行取整，$trunc_T(\cdot)$ 表示对残差进行截断. 最后，将截断量化后的残差值进行数值统计形成共生矩阵，对称合并后获得特征向量[29].

4　本文模型结构

为了同时应对篡改区域的分类与定位两个任务，本文采用了多任务处理的网络结构实现篡改的分类与边界框的回归，如图 5 所示. 其中，将 RGB 图像作为彩色图像通道(如图 5 上部分通道所示)的输入，并将经过 SRM 滤波器层处理的图像作为隐写分析通道(如图 5 下部分通道所示)的输入. 两种通道对图像进行处理后会分别获得不同的特征信息，之后通过双线性池化层(bilinear pooling layer)[31] 融合两个通道的空间特征信息. 最后，根据得分判断

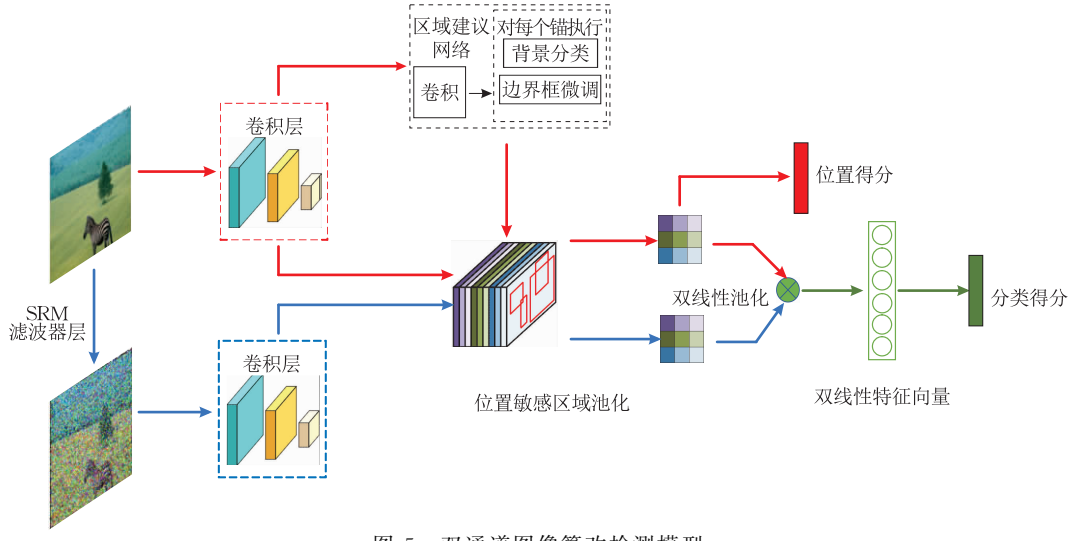


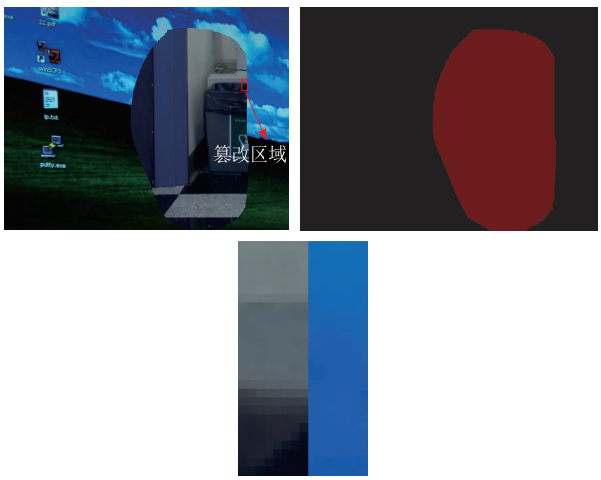
图 5　双通道图像篡改检测模型

区域是否发生过篡改，并对篡改区域进行标注. 在该模型中，区域建议网络层仅采用彩色图像特征作为输入，但位置敏感得分图采用彩色图像与隐写分析两种特征作为输入.

4.1 彩色图像通道

大多数篡改图像与真实图像相比，在视觉上有很大差异，其表现如图 6 所示. 在图 6(a)中，第 1 行第 1 幅图像中的篡改区域包含篡改区域与非篡改区域的相交边界，非篡改区域包含非篡改区域间的相交边界，第 1 行第 2 幅图像为篡改图像所对应的真值掩膜图像，第 2 行第 1 幅图像与第 2 幅图像分别为篡改区域与非篡改区域放大后的图像展示. 显然，包含篡改区域的边界在视觉上很不自然，边界较为平滑，而不包含篡改区域的边界在视觉上呈现锯齿状，边界较为尖锐，颗粒感明显. 在图 6(b)中，第 1 行第 1 幅图像中的篡改区域包含了篡改区域与非篡改区域的相交边界，第 1 行第 2 幅图像为篡改图像所对应的真值掩膜图像，第 2 行图像为篡改区域放大后的图像展示. 从图中可以看出，在由篡改区域向非篡改区域过渡时，色彩变化明显，边界两边区域的颜色对比较为强烈. 通过彩色图像通道，可以提取图像的表层特征，并辅助模型做出精准的判断.

(a) 边界差异



(b) 色彩差异

图 6 篡改图像在视觉上的表现

在使用 ResNet[27] 对图像特征进行提取时，随着网络逐渐加深，来自不同层次特征的空间表现会逐渐加强，但同时特征图的空间分辨率也在逐渐降低. 例如，ResNet 101 中 Conv1 卷积层输出的卷积特征图的图像尺寸为 112×112，而 Conv3_x 卷积层输出的卷积特征图尺寸为 28×28，是 Conv1 输出尺寸的 1/4. 分辨率降低使图像变得模糊，减少了图像携带的信息，导致模型的特征表达能力变弱，使模型提取到的特征数量减少，这对图像小区域篡改检测非常不利. 针对上述问题，本模型采用双线性插值方法，将特征提取网络中每一个堆叠卷积层输出的特征图调整为更大尺寸，来缓解检测中小目标信息缺失的问题. 在插值过程中，第 i 个位置的特征向量如下表示：

$$\boldsymbol{x}_i=\sum_k \alpha_{ik} o_k \tag{5}$$

其中，o_k 表示原始特征图，$\boldsymbol{x}_i$ 表示上采样后的特征图，插值权重 α_{ik} 取决于 i 和 k 两个相邻特征向量的位置. 此插值发生在空间域中，图 7 为插值过程示意图.

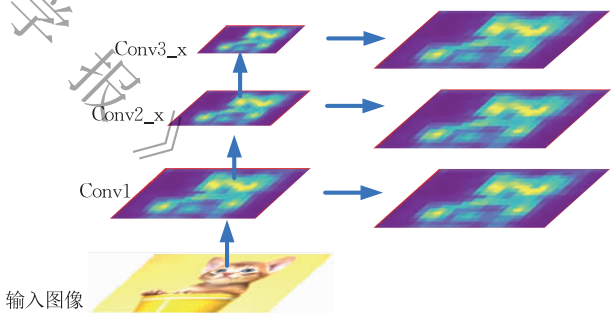


图 7 针对特征提取网络的双线性插值

在目标检测模型中，RPN 网络的作用是用来找出可能包含检测物体的区域，而本文的 RPN 网络是用来找出可能篡改区域. RPN 网络将特征提取网络输出的特征图作为输入，使用锚(anchors)机制生成区域候选框，对不包含物体的区域候选框进行去除，并对候选框的边界进行预调整，从而减少后续训练的时间. RPN 网络的损失函数公式如下：

$$L_R(g_i,t_i)=\frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(g_i,g_i^*)+\lambda\frac{1}{N_{\text{reg}}}\sum_i g_i^* L_{\text{reg}}(t_i,t_i^*) \tag{6}$$

其中，i 表示每个锚点，g_i 表示每个锚中包含篡改区域的概率预测，g_i^* 表示每个锚的真值标签，t_i 与 t_i^* 分别

表示每个锚边界框的四个描述值及其真值，用来表示锚的偏移量. $L_{cls}(\cdot)$表示两个目标(前景与背景)的交叉熵损失，N_{cls}表示小批量的大小. $L_{reg}(\cdot)$表示用于边界框回归的 smooth_{L_1} 损失函数，N_{reg}表示锚点位置的总数量，λ 表示平衡参数，用来平衡两种损失函数，本文将它设置为 10. smooth_{L_1} 损失函数公式如下：

$$\text{smooth}_{L_1}(x)=\begin{cases}0.5x^2, & |x|<1\\ |x|-0.5, & \text{其他}\end{cases} \tag{7}$$

模型除了在 RPN 网络中对候选框进行筛选，还使用非极大值抑制(Non Maximum Suppression，NMS)算法对同一目标重复覆盖的候选框进行消除，该算法即搜索局部最大值，抑制非极大值元素，保留最优候选框. 基于 RPN 网络，模型进行位置敏感区域池化操作，并通过彩色图像通道实现候选区域定位.

4.2　隐写分析通道

图像篡改的手段多种多样，同时篡改者也会对图像篡改痕迹进行掩饰，所以仅采用彩色图像通道的检测效果较差. 在关注图像表面痕迹的同时，更应该注意到图像内部的统计信息. 隐写分析通道可以很好地利用图像内部的噪声分布，将注意力放在篡改后被破坏的统计信息上，帮助模型提高检测精度.

与彩色图像通道不同，原始图像首先要经过 SRM 滤波器层进行预处理. 根据 Zhou 等人[21]的结论，使用 30 个滤波器在性能上并不会带来显著的提升，反而会大幅增加检测时间，为了保证模型的效率，本文同样保留三个滤波器，其权重如下所示：

$$\frac{1}{4}\begin{bmatrix}0&0&0&0&0\\0&-1&2&-1&0\\0&2&-4&2&0\\0&-1&2&-1&0\\0&0&0&0&0\end{bmatrix}$$

$$\frac{1}{12}\begin{bmatrix}-1&2&-2&2&-1\\2&-6&8&-6&2\\-2&8&-12&8&-2\\2&-6&8&-6&2\\-1&2&-2&2&-1\end{bmatrix} \tag{8}$$

$$\frac{1}{2}\begin{bmatrix}0&0&0&0&0\\0&0&0&0&0\\0&1&-2&1&0\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}$$

这三个滤波器是 SRM 滤波器层的构成，它的输出通道大小为 3.

经过 SRM 滤波器层预处理，得到的图像作为后面特征提取网络的输入，特征提取网络在结构上与彩色图像通道中的特征网络是相同的. 但是，隐写分析通道不包含区域建议网络结构，它使用彩色图像通道中生成的候选区域进行后续操作.

4.3　双线性池化层

双线性池化[31]最早被应用在细粒度图像分类上，它可以融合双通道的卷积神经网络，同时保留双通道的空间信息以提高检测精度. 本文使用压缩双线性池化层[32]将彩色图像通道与隐写分析通道的信息进行融合. 压缩双线性池化层的输出如下所示：

$$x=f_c^{\mathrm{T}}f_s \tag{9}$$

其中，f_c表示彩色图像通道的位置敏感图特征，f_s表示隐写分析通道的位置敏感图特征. 重组后的双线性特征向量会作为后续得分的依据，判断区域是否发生篡改.

使用交叉熵损失来评估篡改区域的分类，使用 smooth_{L_1} 损失函数来评估边界框的回归. 最后，模型总的损失函数如下所示：

$$L_t=L_R+L_C(f_c,f_s)+L_B(f_c) \tag{10}$$

其中，L_t表示模型总的损失，L_R表示 RPN 网络损失函数，L_C表示最后的交叉熵分类损失，它是由经过双线性池化层的双通道特征 f_c与 f_s共同决定的，L_B表示最后的边界框回归损失，它仅由来自彩色图像通道的特征 f_c决定.

5　实验结果及分析

5.1　参数设置及实验环境

为了保证输入图像经过 SRM 滤波器时获得更好的处理效果，本文将输入图像的短边调整为 600 像素. 在 RPN 网络层，将 4 个锚的尺寸分别设置为 8^2，16^2，32^2，64^2，并将长宽比分别设置为 1∶2，1∶1 以及 2∶1. 将 RPN 中用于判断正样本(可能为篡改区域)的 IOU 阈值设置为 0.7，负样本设置为 0.3. 非极大值抑制算法的阈值设置为 0.3. 实验运行环境如表 2 所示.

表 2　实验运行环境

类别	配置
电脑类型	台式电脑
显卡	Nvidia GeForce RTX 2080 TI
CPU	Intel Core i9-9900K
内存大小	32 GB
操作系统	Ubuntu 16.04.6 LTS
深度学习框架	Tensorflow
CUDA 版本	CUDA 10.1
cuDNN 版本	cuDNN v7.5.0
编程语言	Python 3.5.6

5.2 评估指标及数据集

5.2.1 评估指标

为了更好地对模型进行评估，本文采用平均精度 AP(Average precision)、$F1$ 分数($F1$-score)和检测速率 Fps(Frames per second)作为模型的评估指标. 其中，Fps 可以通过每秒钟检测器处理图片的张数评估模型在速度上的提升.

5.2.2 数据集

鉴于专业图像篡改数据集的规模较小，不能很好地满足深度神经网络的训练要求. 本模型使用专业的目标检测数据集，利用数据集中不同物体的标注信息，随机地将物体区域复制粘贴到其它图片中. 使用合成的数据集对模型进行预训练，并通过专业的图像篡改数据集对模型进行微调. 下面对本文用到的数据集进行介绍.

(1) 预训练数据集

PASCAL VOC[33]. 为图像识别与分类提供的一套标准数据集，共包含人、动物(如猫、狗等)、交通工具(如船、飞机等)、家具(如椅子、桌子、沙发等)在内的 20 个类别的物体. 本文利用 PASCAL VOC 2012 数据集中提供的标注信息，实现篡改数据集的创建，并成功生成 11725 张图片用于模型的训练与测试.

(2) 微调数据集

(a) NIST①. 该数据集中提供复制移动，拼接，移除三种篡改手段的图像，并提供真值掩膜图像用于模型的评估.

(b) CASIA2.0[34]. 该数据集中包含复制移动，拼接，移除三种篡改手段的图像，除了对篡改区域进行过精心选择外，还通过预处理对篡改痕迹进行掩盖，并提供真值掩膜图像用于模型的评估.

(c) Columbia②. 该数据集侧重于提供未压缩图像的拼接篡改图像，并提供真值掩膜图像用于模型的评估.

对 PASCAL VOC[33]、NIST、CASIA2.0[34] 和 Columbia 这 4 个数据集分别进行训练集与测试集的设计. 其中，将前三个数据集中 90% 的图片用作训练集，剩余的 10% 的图像用作测试集. 由于 Columbia 数据集中的图像数量较少，故将其全部设置为测试集. 每个数据集关于训练集与测试集的划分，如表 3 所示.

表 3 训练集与测试集的划分

数据集	合成 PASCAL VOC	NIST	CASIA2.0	Columbia
训练集	10567	404	4611	—
测试集	1185	160	512	180

5.3 模型预训练

模型中用于 R-FCN 架构的 ResNet 101 是在 ImageNet[35] 上进行预训练的，并在 PASCAL VOC[33] 合成数据集上对双通道模型进行预训练. 为了增强模型的泛化能力，在训练过程中，对 PASCAL VOC 合成数据集中的图片进行了平移、旋转和缩放三种方式的数据增强. 训练中采用随机梯度下降法(Stochastic Gradient Descent，SGD)对网络参数进行更新，选取 Batchsize 为 64，初始学习率设为 0.001 迭代 40k 次，然后学习率改为 0.0001 迭代 70k 次. 如图 8 所示，对预训练过程中总损失函数的变化进行绘制，该图的横轴为模型的迭代次数，纵轴为损失函数的值，由于数据随机波动比较大，对数据进行拟合处理，可以清楚地观察到损失函数的变化情况.

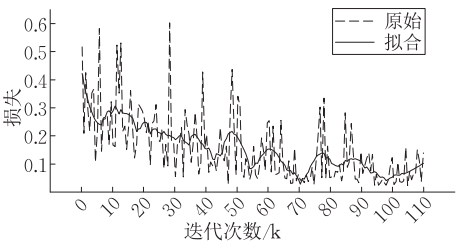


图 8 总损失函数的变化

为了分析不同通道与双线性池化层对模型精度与效率的影响，本文在相同的实验环境中，使用平均精度 AP 与速率对不同结 构的模型进行评估. 同时，为了保持对比实验的公平性，对比模型在数据集、参数设置与训练过程上，与本文模型保持一致，各模型的差异主要体现在模型中不同类型特征的输入和模型结构上，评估结果如表 4 所示. 其中，单彩色图像通道模型是仅使用 RGB 图像作为输入图像的单通道模型，为单个的 R-FCN[6] 结构；单隐写分析通道模型是仅使用经过 SRM 滤波器层处理的图片作为输入图像的单通道模型；组合模型由单彩色图像通道模型和单隐写分析通道模型构成，但并没有融合两个通道的特征，而是通过判别器对两个通道模型的输出结果进行筛选，选取最优检测区域；隐写分析 RPN 模型是在采用双通道模型的基础上，将隐写分析通道中特征提取网络输出的特征作为 RPN 网络的输入；双通道 RPN 模型是在采用双通

① Nist nimble 2016 datasets. https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation/

② Ng T T, Hsu J, Chang S F. Columbia image splicing detection evaluation dataset. http://www.ee.columbia.edu/ln/dvmm/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm

道模型的基础上，将彩色图像通道与隐写分析通道中特征提取网络输出的特征共同作为 RPN 网络的输入；本文模型是在采用双通道模型的基础上，将彩色图像通道输出的特征作为 RPN 网络的输入.

表 4　不同结构模型对比

模型	AP/%	*rate*/fps
单彩色图像通道	63.59	12
单隐写分析通道	65.74	11
组合模型	72.35	6
隐写分析 RPN	66.52	9
双通道 RPN	82.13	7
本文模型(彩色图像 RPN)	**82.96**	**9**

由表 4 可知，双通道网络在性能上要优于每个单通道网络，原因如下：(1) 在篡改过程中，图片上的篡改痕迹可能会遭到篡改者的处理，仅使用彩色图像通道捕获到的特征信息会大大减少，这给后续的分类与定位带来很大的难度；(2) 图像篡改时，内部统计信息变化的区域不能完美地反映篡改区域，因为篡改者对篡改痕迹进行掩饰时，图像中的非篡改区域也会受到影响，这时单使用隐写分析通道是不够的，同时也需要图像外部的残留痕迹辅助模型做出更加精准的判断. 通过检测速率的对比可以看出双通道网络与单通道网络相比，双通道网络的检测速度要慢于单通道网络，这是因为双通道网络增加了复杂度. 通过组合模型与本文模型的对比，可以看出双线性池化层的特征融合明显提高了模型的精度，这是因为它可以同时兼顾到两个通道的特征信息，增强了模型的一体性，而不是组合模型那样从两个通道中筛选出最优检测区域. 在检测速度上，组合模型比本文模型慢，主要原因是组合模型包含两个完整的通道结构，模型整体过于复杂，而本文模型的双线性池化层在一定程度上减轻了模型末端的复杂度，提升了模型的检测速度. 此外，后三行数据表明，单独使用彩色图像通道特征生成候选区域的模型会更有优势，它的效果甚至要好于使用双通道特征生成候选区域的模型.

5.4　实验结果分析

在 NIST、CASIA2.0[34] 和 Columbia 这 3 个标准数据集上进行微调训练后，进行了多角度的对比实验，实验中选取的对比模型有传统方法和深度学习方法. 其中，传统方法为 Ferrara 等人[13] 基于色彩滤镜矩阵图像篡改检测方法；深度学习方法包括 Bondi 等人[19] 基于成像设备相关性的方法、Bappy 等人[17] 基于改进网络结构的方法和 Zhou 等人[21] 基于双流网络的方法.

不同模型间平均精度与速率对比，如表 5 所示. 其中 2,3,4 列分别为模型在 3 个标准数据集上的精度得分，第 5 列表示模型的速率.

表 5　不同模型间平均精度与速率对比

模型	NIST/%	CASIA2.0[34]/%	Columbia/%	*rate*/fps
CFA1[13]	20.37	51.29	27.95	0.2
Tam-D[19]	—	72.30	79.12	1.7
J-Conv-LSTM[17]	81.93	78.42	78.61	2.0
RGB-N[21]	94.16	87.27	80.53	4.0
本文模型	**95.29**	**89.63**	**83.74**	**9.0**

不同模型间 *F*1 分数的柱状对比图，如图 9 所示. 其中横坐标代表不同的模型，纵坐标代表 *F*1 分数，三种颜色分别代表在三种不同的数据集上进行实验. 结合表 5 可以清楚地看出，基于深度学习的篡改检测方法要明显优于传统的篡改检测方法，如 CFA1[13]. 这是因为传统方法采用人工设计的特征，它更加专注于特定的篡改手段，面对多样的篡改图像，反而会限制它的性能. 本文的模型在三个数据集上都要优于 Tam-D[19] 模型. J-Conv-LSTM[17] 模型更多关注篡改区域的边缘信息，而本文模型考虑的篡改信息更加丰富，这使本文模型的性能要优于 J-Conv-LSTM[17]. 本文模型与 RGB-N[21] 模型在结构上相似，但存在两点明显差异：(1) RGB-N[21] 模型使用 Faster R-CNN[7] 作为主干网络，而本文模型使用 R-FCN[6]. R-FCN 利用位置敏感区域池化的思想，增加了共享参数的网络层数，降低了模型第二阶段的耗时，并去掉网络中的全连接层，使用全局平均池化的策略，进一步减少了模型第二阶段的耗时，提高了模型的检测速度；(2) 本文模型在特征提取部分，利用双线性插值法调整了特征图的分辨率，减少了生成特征图时缺失的信息，提高了小面积篡改区域的检测精度. 以上两点差异使本文模型在三个数据集上的表现更加出色.

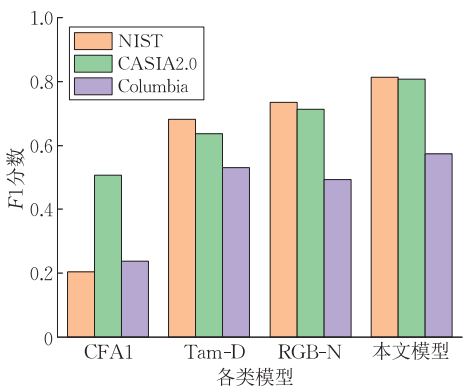


图 9　各类模型的 *F*1 分数

在 NIST 数据集中包含复制移动、拼接和移除三种篡改手段的图像. 为了探究模型在不同篡改方式上的表现,在 NIST 数据集上对每种篡改手段的检测结果进行了统计计算,结果如图 10 所示.

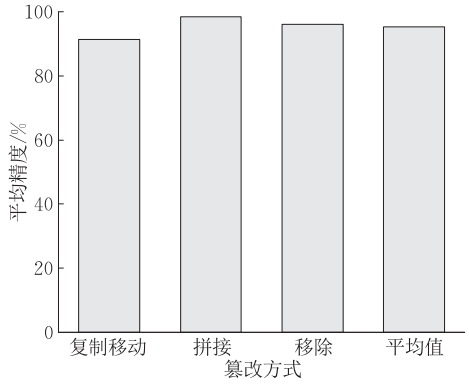


图 10 不同篡改方式下的平均精度对比

从图 10 可以看出,拼接的检测效果最好,因为在拼接过程产生的篡改区域往往与原图差别较大,边界和色彩差异比较强烈,方便网络的识别;移除过程中对背景区域的修复会对图像内部的统计特征造成一定影响,这会干扰它的检测性能;复制移动的检测效果不是很理想,因为复制的区域来自原图中,其对比度不会相差太大,而且复制区域与原图中相似的统计信息也会对隐写分析通道造成干扰. 今后,会使用相似性判别的方法,对复制移动的区域进行相似性校对,提升复制移动方式检测的准确率.

本文模型与当前先进模型 RGB-N[21] 的检测效果对比,如图 11 所示. 其中,第 1 行是对复制移动方式篡改图像的检测,图片取自 CASIA2.0 数据集;第 2 行是对拼接方式篡改图像的检测,图片取自 Columbia 数据集;第 3 行是对移除方式篡改图像的检测,图片取自 CASIA2.0 数据集;第 4 行是针对小区域篡改图像的检测,图片取自 COCO 数据集的合成图像,COCO 数据集是是由微软公司创建的,用于物体检测与分割的大型数据集;第 1 列到第 5 列分别为真实图像、篡改图像、真值掩膜图像、RGB-N 模型检测效果图和本文模型检测效果图.

由图 11 可以看出,由于复制移动方式下的篡改区域与原图区域相似度极高,导致 RGB-N 模型出现了混淆识别,而本文模型避免了这种情况的发生;在对拼接与移除方式的篡改图片进行检测时,本文模型能够对篡改区域进行较为精准的定位,而不会出现识别错误和不准确的情况;由于本文针对小区域检测进行了优化,所以能够很好地对图片中小面

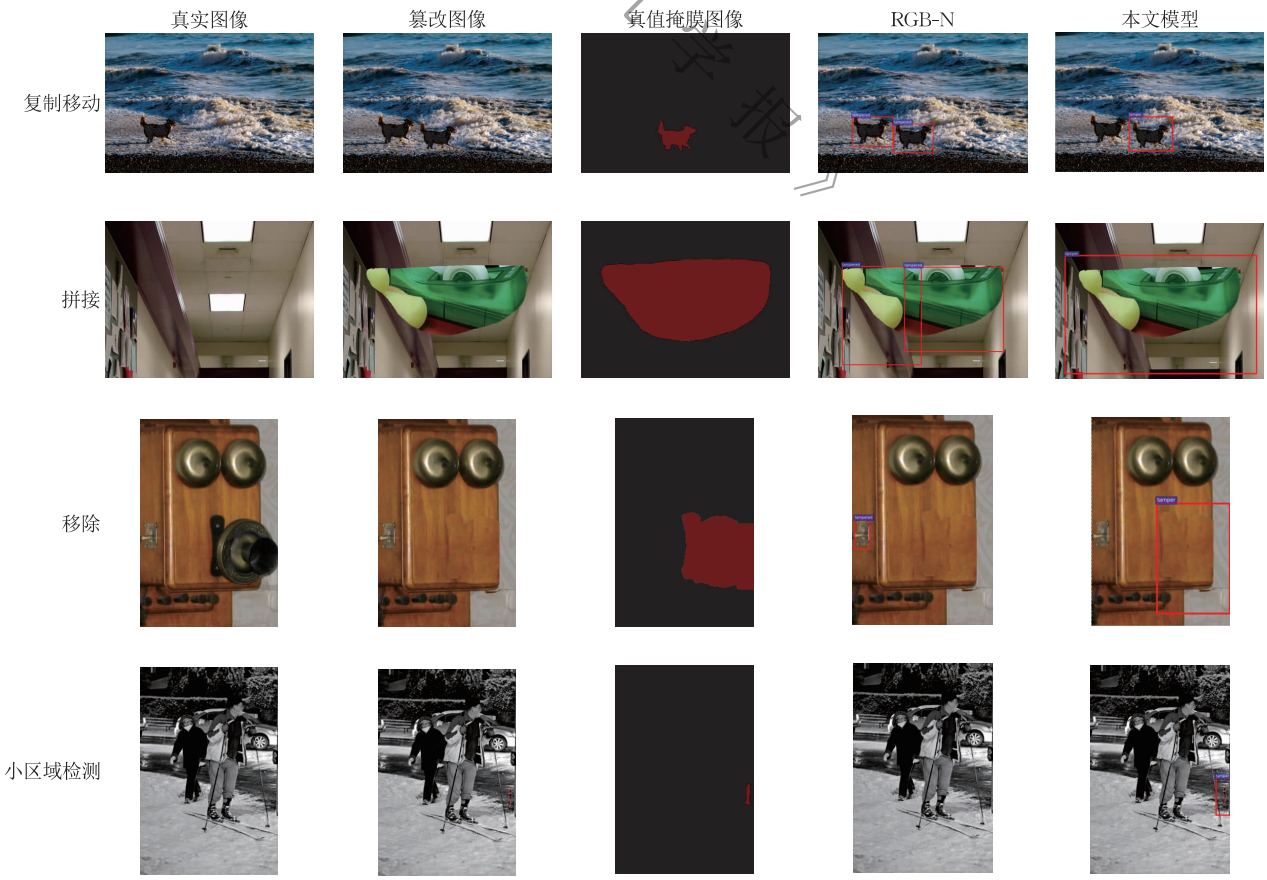


图 11 RGB-N[21] 与本文模型的检测效果比较

积篡改区域进行检测，而 RGB-N 模型却不能对其进行识别.

6　总结与展望

图像篡改检测是检测图像真实完整的重要方法，实现该技术在具体场景下的应用显得尤为重要. 如今，图片的数量呈指数级增长，如何实现篡改图像的快速检测成为当下的关键问题. 本文通过对现有图像篡改检测模型的分析，同时结合篡改检测对小面积篡改区域检测精度不高的问题. 首次提出了一种基于双通道 R-FCN 的图像篡改检测模型，并结合双线性插值法，在提升模型检测速度的前提下，提高了检测精度. 但是，复制移动方式下的篡改区域与原图有着相似颜色、纹理特征和统计特性，给篡改检测带来了许多困难. 针对上述问题，下一步计划是设计一种轻量级的图像篡改检测网络模型，并引入严格的相似性判别方案，实现模型检测精度的进一步提升.

参　考　文　献

[1] Li Li-Chun, Zhang Xiao-Hu, Liu Xiao-Chun, et al. Research on camera measurement of "South China Tiger" photograph. Science & Technology Review, 2008, 26(1): 59-67(in Chinese)
(李立春，张小虎，刘晓春等. "华南虎"照片的摄像测量研究. 科技导报，2008，26(1)：59-67)

[2] Zhou Lin-Na, Wang Dong-Ming. Digital Image Forensics. Beijing: Beijing University of Posts and Telecommunications Press, 2008(in Chinese)
(周琳娜，王东明. 数字图像取证技术. 北京：北京邮电大学出版社，2008)

[3] Luo Wei-Qi, Huang Ji-Wu, Qiu Guo-Ping, et al. Robust region copy image tamper detection technology. Chinese Journal of Computers, 2007, 30(11): 1998-2007(in Chinese)
(骆伟祺，黄继武，丘国平. 鲁棒的区域复制图像篡改检测技术. 计算机学报，2007，30(11)：1998-2007)

[4] Bianchi T, Rosa A D, Piva A. Improved DCT coefficient analysis for forgery localization in JPEG images//Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Prague, Czech Republic, 2011: 2444-2447

[5] Rao Y, Ni J. A deep learning approach to detection of splicing and copy-move forgeries in images//Proceedings of the 2016 IEEE International Workshop on Information Forensics and Security (WIFS). Abu Dhabi, UAE, 2016: 1-6

[6] Dai J, Li Y, He K, et al. R-FCN: Object detection via region-based fully convolutional networks//Proceedings of the Neural Information Processing Systems (NIPS). Barcelona, Spain, 2016: 379-387

[7] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks//Proceedings of the Neural Information Processing Systems (NIPS). Montreal, Canada, 2015: 91-99

[8] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors. Nature, 1986, 323(6088): 533-536

[9] Mahdian B, Saic S. Blind authentication using periodic properties of interpolation. IEEE Transactions on Information Forensics and Security, 2008, 3(3): 529-538

[10] Stamm M C, Liu K J R. Forensic detection of image manipulation using statistical intrinsic fingerprints. IEEE Transactions on Information Forensics and Security, 2010, 5(3): 492-506

[11] Hashmi M F, Anand V, Keskar A G. A copy-move image forgery detection based on speeded up robust feature transform and wavelet transforms//Proceedings of the Computer and Communication Technology (ICCCT). Narvik, Norway, 2014: 147-152

[12] Hsu Y F, Chang S F. Camera response functions for image forensics: an automatic algorithm for splicing detection. IEEE Transactions on Information Forensics and Security, 2010, 5(4): 816-825

[13] Ferrara P, Bianchi T, De Rosa A, et al. Image forgery localization via fine-grained analysis of CFA artifacts. IEEE Transactions on Information Forensics & Security, 2012, 7(5): 1566-1577

[14] Gaborini L, Bestagini P, Milani S, et al. Multi-clue image tampering localization//Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS). Atlanta, USA, 2014: 125-130

[15] Li H, Luo W, Qiu X, et al. Image forgery localization via integrating tampering possibility maps. IEEE Transactions on Information Forensics and Security, 2017, 12(5): 1240-1252

[16] Bayar B, Stamm M C. A deep learning approach to universal image manipulation detection using a new convolutional layer//Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security. Vigo, Spain, 2016: 5-10

[17] Bappy M J H, Roy-Chowdhury A, Bunk J, et al. Exploiting spatial structure for localizing manipulated image regions//Proceedings of the International Conference on Computer Vision (ICCV). Venice, Italy, 2017: 4970-4979

[18] Baroffio L, Bondi L, Bestagini P, et al. Camera identification with deep convolutional networks. IEEE Signal Processing Letters, 2016, 24(3): 259-263

[19] Bondi L, Lameri S, Güera D, et al. Tampering detection and localization through clustering of camera-based CNN features //Proceedings of the Computer Vision and Pattern Recognition Workshops. (CVPRW). Honolulu, USA, 2017: 1855-1864

[20] Cozzolino D, Poggi G, Verdoliva L. Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection//Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security. Philadelphia, USA, 2017: 159-164

[21] Zhou P, Han X, Morariu V I, et al. Learning rich features for image manipulation detection//Proceedings of the Computer Vision & Pattern Recognition (CVPR). Salt Lake City, USA, 2018: 1053-1061

[22] Chen Chao, Qi Feng. Review on development of convolutional neural networks and its application in computer vision. Computer Science, 2019, 46(3): 63-73(in Chinese)
(陈超, 齐峰. 卷积神经网络的发展及其在计算机视觉领域中的应用综述. 计算机科学, 2019, 46(3): 63-73)

[23] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Oregon, USA, 2013: 580-587

[24] He K, Gkioxari G, Dollar P, et al. Mask R-CNN//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Venice, Italy, 2017: 2961-2969

[25] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, real-time object detection//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA, 2016: 779-788

[26] Liu W, Anguelov D, Erhan D, et al. SSD: Single shot multibox detector//Proceedings of the European Conference on Computer Vision (ECCV). Amsterdam, The Netherlands, 2016: 21-37

[27] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA, 2016: 770-778

[28] Yuan Ya-Fei, Lu Wei, Feng Bing-Wen, et al. Research and implementation of online steganographic blind analysis system based on multiple pre-training models. Journal of Network and Information Security, 2017, 3(5): 32-37(in Chinese)
(袁亚飞, 卢伟, 冯丙文等. 基于多预训练模型的在线隐写盲分析系统研究与实现. 网络与信息安全学报, 2017, 3(5): 32-37)

[29] Fridrich J, Kodovsky J. Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882

[30] Zhang Hao-Jie. Information Hiding Countermeasure Research Based on Characteristic Analysis of Residual Extraction Filter. Shenzhen University, Shenzhen, 2017(in Chinese)
(张浩杰. 基于残差提取滤波器特性分析的信息隐藏对抗研究. 深圳大学, 深圳, 2017)

[31] Lin T Y, Roychowdhury A, Maji S. Bilinear CNN models for fine-grained visual recognition//Proceedings of the IEEE International Conference on Computer Vision (ICCV). Santiago, Chile, 2015: 1449-1457

[32] Gao Y, Beijbom O, Zhang N, et al. Compact bilinear pooling //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, USA, 2016: 317-326

[33] Everingham M, Gool L V, Williams C K I, et al. The PASCAL visual object classes (VOC) challenge. International Journal of Computer Vision, 2010, 88(2): 303-338

[34] Dong J, Wang W, Tan T. CASIA Image tampering detection evaluation database//Proceedings of the Signal and Information Processing (ChinaSIP). Beijing, China, 2013: 422-426

[35] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database//Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Miami, USA, 2009: 248-255

TIAN Xiu-Xia, Ph. D., professor. Her main research interests include digital image forgery detection, database security, privacy preserving (big data and cloud computing), secure machine learning, security computing for the benefit of power users.

LI Hua-Qiang, M. S. candidate. his current research interests include digital image forgery detection, deep learning, target detection.

ZHANG Qin, M. S., engineer. Her main research interests include information security, information network and communication technology.

ZHOU Ao-Ying, Ph. D., professor, Ph. D. supervisor. His main research interests include database, data management, digital transformation, data-driven applications such as Educational Technology (EduTech) and Logistics Technology (LogTech).

Background

In the field of information security, image forgery detection is a highly practical technology, which has been widely applied in politics, economy, culture and other fields. Therefore, image forgery detection has always been a research hotspot. In recent years, the deep learning method based on data-driven approach has been applied in many computer visions and image processing tasks. Deep learning shows excellent performance, which promotes many researchers to apply it in image forgery detection, including the several popular directions: improve the detection accuracy by updating the network structure,

design network model based on the correlation of imaging equipment, and obtain better detection results through the characteristics of steganalysis. After analyzing the existing deep learning forgery detection algorithms, we found that the efficiency of most algorithms was low in image forgery detection when faced with a large number of images, and the detection accuracy of small area tampering region was poor. To solve the above problems, we proposed a dual-channel image forgery detection model based on Region-based Full Convolution Network (R-FCN).

First of all, the design of the dual-channel model allowed us to mined richer image features and made more accurate judgments. Secondly, the R-FCN model took good advantage of the full convolution network thanks to the design of position-sensitive score map, which greatly reduced the computation of the model and improved the model's detection efficiency. Thirdly, we used the bilinear interpolation method to adjust the size of the feature map in the feature extraction network, which reduced the loss of information and improved the detection accuracy of the small tampering area. We pre-trained the model with the synthetic dataset, and conducted experiments on several international mainstream datasets. By compared our model with the state-of-the-art works, we found that the proposed model has improved performance in multiple evaluation indicators. Experimental results showed the proposed image forgery detection model is more efficient and accurate.

This article was partially supported by the National Natural Science Foundation of China (General Program; Key Program) (Nos. 61772327, 61532021), the Electric Power Research Institute of State Grid Gansu Electric Power Company Enterprise Project (H2019-275). The research goals of these projects included privacy protection issues for data users and owners, user data integrity audits, access control mechanisms, efficient search techniques for encrypted outsourced databases, and some methods that can be used in the field of image forgery.