文章编号：2095-2163(2023)11-0014-08 中图分类号：TP391.41 文献标志码：A

面向图像篡改检测的双流卷积注意力网络

孙 冉，张玉金，张立军，郭 静
（上海工程技术大学 电子电气工程学院，上海 201620）

摘 要：拼接和复制-粘贴是最常见的两种图像篡改手段，伪造区域的定位是图像取证领域最具挑战性的科学问题。针对该问题，提出了一种双流卷积注意力网络，以检测出可疑图像的伪造区域。双流卷积注意力网络分别考虑不同通道间像素的重要性和同一通道不同位置像素的重要性可以学习更丰富的特征，以提高检测准确度。第一支流为 RGB 流，从 RGB 图片中提取边缘异常、颜色反差等特征；另一支流为噪声流，捕捉真实区域和伪造区域之间的不一致噪声信息。双流网络提取到的特征信息在双线性池化层进行特征融合，在 *softmax* 层输出篡改检测结果。 实验结果表明，本文方法在公共数据集上表现优于现有方法，并且对 JPEG 压缩具有较好的鲁棒性。

关键词：图像篡改；注意力机制；双流网络；双线性池化；图像篡改检测

Image forgery detection based on two-stream cascaded attention network

SUN Ran, ZHANG Yujin, ZHANG Lijun, GUO Jing
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Copy-move and splicing are the two most common image tampering methods. The location of the forged area is the most challenging problem in the field of image forensics. The paper proposes a two-stream cascaded attention network to detect the forgery area of a given tampered image. By considering the importance of pixels between different channels and the importance of pixels at different positions in the same channel, two-stream cascaded attention network can learn more features to improve detection accuracy. One of the two streams is an RGB stream, which extracts features such as edge abnormalities and color contrast from the RGB image; the other branch is the noise stream, which captures the inconsistent noise information between the real area and the fake area. The feature information extracted by the two-stream network is fused in the bilinear pooling layer, and the tampering detection result is output in the *softmax* layer. Experimental results demonstrate that the proposed method performs better on common datasets and is robust to JPEG compression.

Key words: image tamper; attention mechanism; two-stream network; bilinear pooling; image tampering detection

0 引 言

随着图像编辑技术的发展，图像篡改成为了低成本的操作，不同的人群篡改图片的目的不同，但都会使图像内容的真实性得不到保障。已有的研究工作表明，图像篡改类型主要包括：复制-粘贴篡改[1]、拼接篡改[2]和修复篡改[3]。其中，复制-粘贴篡改是指在同一幅图像上，把某一部分区域复制后粘贴到该图像的另一个位置，从而达到以假乱真的目的；拼接篡改是指将一幅图像的某个部分复制下来粘贴到其他图像中以合成一幅伪造图像；修复篡改是指基于图像原有信息还原缺失部分或移除原图某一区域。目前，主流的图像篡改检测方法可以分为主动检测和被动检测（盲检测）[4]，二者的主要区别在于是否在图像中预先嵌入附加信息，如数字水印等。

图像拼接使用的源图像一般来自两幅或多幅不同图片，人们在对图像进行篡改时，往往只关注 RGB 域的逼真程度，而忽略图像噪声域的统计特性变化。图像噪声是指存在于图像数据中的干扰信

基金项目：上海市自然科学基金项目（17ZR1411900）；上海市信息安全综合管理技术研究重点实验室项目（AGK2015006）。

作者简介：孙 冉（1996-），男，硕士研究生，主要研究方向：图像处理；张立军（1974-），男，博士，讲师，主要研究方向：图像处理、计算机视觉；
郭 静（1996-），女，硕士研究生，主要研究方向：图像处理、计算机视觉。
通讯作者：张玉金（1982-），男，博士，副教授，硕士生导师，主要研究方向：多媒体取证、图像处理、模式识别。Email：yjzhang@ sues.edu.cn

收稿日期：2022-11-09

息，图像成像过程中，CCD 和 CMOS 传感器采集数据时一般会受到传感器材料属性、工作环境和电路结构等影响而引入各种噪声[5]。由于拼接篡改使用的图像通常来源于不同成像设备，而这些设备的噪声分布往往具有一定的差异，因此，噪声的不一致性对图像拼接篡改的分析与鉴定具有较好的辅助作用。

2012 年，以 Alex-Net[6] 为代表的卷积神经网络（Convolutional Neural Network，CNN）在特征提取方面表现优异，随后一些学者开始使用深度学习技术来解决图像篡改检测问题。Yuan 等学者[7] 首次将卷积神经网络用于数字图像篡改检测，该方法从 RGB 彩色图像自动学习特征层次表示，并采用特征融合技术得到最终判别特征。Johnson 等学者[8] 提出了全卷积网络并应用于语义分割任务，实现了像素级别的分类。Salloum 等学者[9] 对此网络结构稍作修改，提出一种基于边缘强化的多任务图像被动取证框架用于像素级别的篡改区域分割，该算法采用 VGG16 网络提取图像篡改特征，并利用篡改区域掩码对篡改区域进行修正。Bondi 等学者[10] 结合图像成像设备属性的特点，提出利用相机指纹进行图像篡改检测和定位，该算法采用神经网络从图像块中提取相机模型特征，对拼接篡改具有良好的检测效果，但不适用于复制-粘贴的篡改类型。Bappy 等学者[11] 采用了一个混合的 CNN-LSTM 模型来捕捉篡改区域和非篡改区域之间的区分特征，LSTM（Long Short Term Memory）[12] 是长短期记忆模型，能够记录图像上下文信息，并将 LSTM 和 CNN 中的卷积层相结合来理解篡改区域和相邻非篡改区域共享边界上像素间的空间结构差异。Zhou 等学者[13] 基于 Faster R-CNN 网络[14] 提出一种双流网络，并对其进行端到端的训练，以检测可疑的篡改区域。

在上述双流网络中，RGB 流能够有效地反映图像篡改特性，噪声流则能更好地体现不同设备源图像进行拼接后的差异，故 RGB 流和噪声流对于图像篡改检测具有一定的互补性，但由于 Faster R-CNN 最优性能的限制，该网络仍存在提升空间。因此，本文在前人工作基础上改进了卷积注意力机制（Convolutional Block Attention Module，CBAM）[15] 加入到特征提取网络，并在 RPN 模块引入 Soft-NMS 算法[16]，构建了一种面向图像篡改检测的双流卷积注意力网络。改进的卷积注意力机制可有效抑制图片中冗余信息，达到对有效信息的专注检测，Soft-NMS 算法可以有效地降低漏检概率。本文所提的双流网络可以学习更丰富的图像特征，以提高图像篡改检测准确度。

1　网络总体框架

本文所提双流卷积注意力网络的整体流程如图 1 所示。RGB 流将原图输入网络中，通过加入改进卷积注意力机制的特征提取网络从 RGB 图像中提取特征，捕捉 RGB 域中的边缘异常、颜色反差等篡改痕迹；噪声流首先利用 SRM 模型[17] 提取噪声信息，再通过特征提取网络分析图像真实区域和被篡改区域噪声间的不一致性；最后，将 2 个支流中提取到的特征信息在双线性池化层[18] 融合得到最终的特征图，送入最后的全连接层进行分类和位置精修。

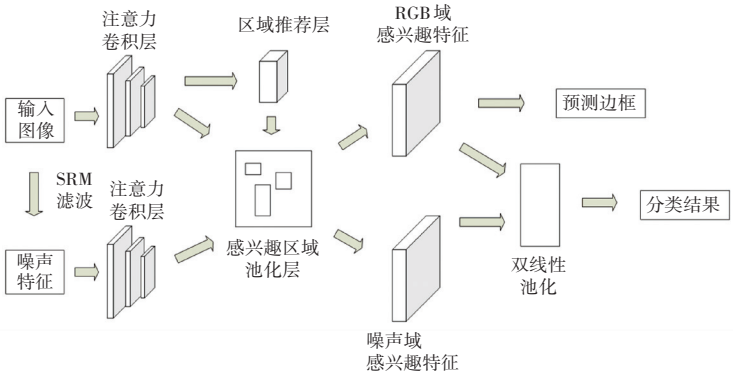


图 1　网络整体框架

Fig. 1　The framework of the network

1.1　改进的卷积注意力模块

注意力机制是提升网络性能的一种方式，在传统的卷积池化过程中，默认特征图的每个通道的重要性是相同的，而实际并非如此，SE block[19] 即是为了解决该问题而研发的。一个 SE 模块分为压缩（Squeeze）和激发（Excitation）两个步骤，通过对前一

个卷积层输出的特征图进行全局平均池化操作得到 $1*1*C$ 的压缩特征量,再经过 2 个全连接层,先对特征压缩量进行降维、再升维,增加了更多的非线性处理,更好地拟合通道之间复杂的相关性。最后与原始的特征图进行矩阵的对应元素相乘得到不同通道权重的特征图。

CBAM 是轻量级的卷积注意力模型,是对 SE block 的一种改进,由通道注意力机制和空间注意力机制级联而成,CBAM 对特征图进行操作,使提取到的特征更加精炼。其中,通道注意力和 SE block 类似,只是多了一个并行的全局最大池化的操作,研究认为不同的池化意味着提取到的高层次特征更丰富。图 2 展示了通道注意力的过程。

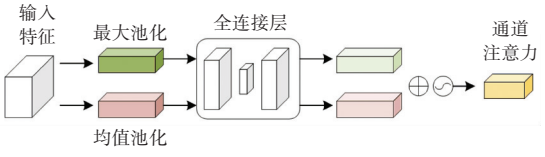


图 2 通道注意力

Fig. 2 Channel attention

空间注意力关注的是同一通道间不同位置像素的重要性,该模块的输入是上一个通道注意力的输出。图 3 为空间注意力过程。

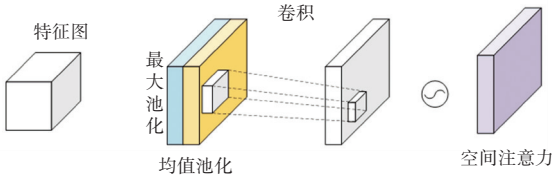


图 3 空间注意力

Fig. 3 Spatial attention

文献[20]中实验表明,SE block 中的 2 个全连接层中的降维操作会给通道注意力预测带来副作用,并且所捕获到通道之间的依存关系效率不高,研究提出一种有效的通道注意力机制(Efficient Channel Attention,ECA)模块,在不降维的情况下进行逐通道全局平均池化后,考虑每个通道及其 k 个近邻来捕获本地跨通道交互。受这种做法的启发,本文给出了改进的 CBAM 注意力模型(Improved CBAM, ICBAM),将 2 个全连接层换成大小为 k 的快速一维卷积生成权值,k 值的大小通过学习自适应确定,结构如图 4 所示。整个过程可以用公式(1)表示:

$$F' = Mc(F) \otimes F$$
$$F'' = Ms(F') \otimes F' \quad (1)$$

其中,F 为输入特征;Mc 为通道注意力特征;Ms 为空间注意力特征;"$\otimes$"表示逐项元素相乘。

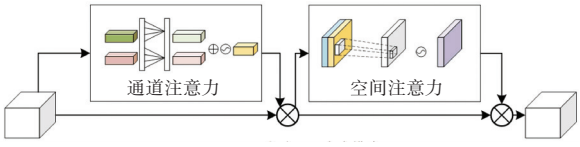


图 4 改进的卷积注意力模块

Fig. 4 Advanced CBAM block

在通道注意力模块,输入特征 F 经过并行的平均池化和最大池化得到 2 个通道描述子,分别通过卷积核大小为 k 的一维卷积计算权重,将得到的特征元素逐项求和,经由 *sigmoid* 函数得到权重系数 Mc,和输入特征 F 相乘得到新的特征。见式(2):

$$Mc(F) = \sigma(Ek(AvgPool(F)) + Ek(MaxPool(F))) \quad (2)$$

其中,σ 表示激活函数,E_k 表示一维卷积后的权重。

在空间注意力模块,输入是上一个通道注意力的输出,把带权重的通道特征送入 2 个大小为列通道维度的池化(最大池化和平均池化)得到 $H*W*2$ 大小的特征图,对该特征图进行卷积操作和 *sigmoid* 激活之后,和该模块带权重的输入对应元素相乘得到最后的结果。研究推得的计算公式为:

$$Ms(F) = \sigma(f^{7*7}([AvgPool(F), MaxPool(F)])) = \sigma(f^{7*7}([F_{avgS}; F_{maxS}])) \quad (3)$$

其中,σ 表示激活函数,f^{7*7} 表示 $7*7$ 卷积操作。

本文采用通道注意力机制在前、空间注意力机制在后的级联形式,将卷积注意力机制加到 ResNet[21]第一个卷积层和最后一个卷积层之后。ResBlock+ICBAM 结构如图 5 所示。

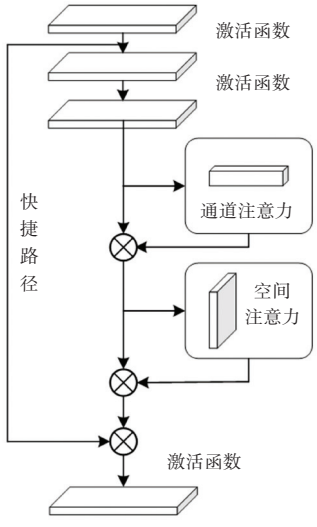


图 5 Resblock+ICBAM 结构

Fig. 5 Resblock with ICBAM

1.2 Faster R-CNN

Faster R-CNN 是一种两阶段目标检测算法，在目标检测领域取得优异成绩，该算法主要由 4 个部分组成：特征提取网络、区域推荐网络（Region Proposal Network，RPN）、RoI（Region of Interest）池化层、分类和回归。其中，特征提取网络提取图像的特征图送到 RPN，RPN 用于生成多个建议框，RoI 池化层综合特征图和 RPN 的建议框信息送入全连接层和 *softmax* 层进行分类，同时进行 bounding box 回归得到最终预测的目标位置。结构流程如图 6 所示。

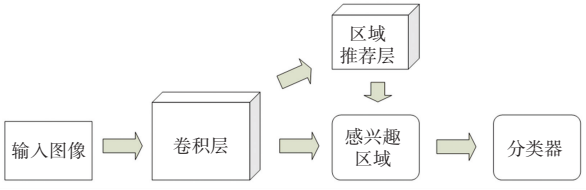


图 6　Faster R-CNN 流程

Fig. 6　The process of Faster R-CNN

1.3 RGB 流

RGB 流是一个基础 Faster R-CNN 网络，在特征提取模块，采用带卷积注意力机制的 ResNet 网络学习 RGB 图像中篡改的特征。RGB 流中的 RPN（region proposal network）模块用来推荐可能存在篡改的区域，这一层使用 *softmax* 层分类器判断建议框是正、还是负，RPN 模块的损失函数如下：

$$L_{RPN}(g_i, f_i) = \frac{1}{N_{cls}} \sum_i L_{cls}(g_i, g_i^*) + \lambda \frac{1}{N_{reg}} \sum_i g_i^* L_{reg}(f_i, f_i^*) \tag{4}$$

其中，g_i 表示候选框 i 可能被篡改的概率；g_i^* 表示候选框 i 为正样本标签；f_i 和 f_i^* 是候选框的四维标签；L_{cls} 表示 RPN 网络的交叉熵损失；L_{reg} 表示建议边框的 L_1 回归损失；N_{cls} 表示 RPN 网络中批量的大小；N_{reg} 表示建议边框的数量；λ 表示用于平衡 2 个损失的超参数，本文选取 $\lambda = 10$。

1.4 噪声流

RGB 流对篡改图像进行检测和定位精度和准确度有限，尤其是当篡改图像经过一些后处理操作、如滤波等，导致拼接区域的边缘不一致信息被隐藏，因此需要引入噪声流辅助检测和定位。

噪声流的设计是为了更关注噪声而不是图像的语义信息，富隐写分析模型（Steganalysis Rich Model，SRM）在图像隐写任务中表现优异，该模型主要从相邻像素中提取局部噪声。本文同样使用 SRM 模型来提取噪声输入到噪声流。在 SRM 的 30 个基础滤波器中，只使用 3 个滤波器也可以达到与 30 个滤波器近似的效果，另外的 27 个滤波器对噪声提取效果并没有明显的提升，因此本文采用 3 个滤波器，滤波器的权重如图 7 所示。

$$\frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 2 & -4 & 2 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \frac{1}{12}\begin{bmatrix} -1 & 2 & -2 & 2 & 1 \\ 2 & -6 & 8 & -6 & 2 \\ -2 & 8 & -12 & 8 & -1 \\ -2 & -6 & 8 & -6 & 2 \\ -1 & 2 & -2 & 2 & -1 \end{bmatrix}$$

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

图 7　SRM 滤波器

Fig. 7　SRM filter

本文将提取出来的噪声特征直接输入到噪声流，噪声流的网络也采用 Faster R-CNN，并且和 RGB 流共用 RoI 池化层的权重。

1.5 双线性池化

图像分别经过 RGB 流和噪声流的特征提取网络后，需要将 2 个特征图融合后再进行篡改的检测和定位操作。双线性池化主要用于特征融合，对于从同一个样本提取出来的特征 X 和特征 Y，将 2 个特征相乘得到矩阵 $\boldsymbol{b}$，对所有位置进行求和池化操作得到矩阵 $\boldsymbol{\xi}$，最后把矩阵 $\boldsymbol{\xi}$ 张成一个张量，记为双线性向量 $\boldsymbol{x}$，对 $\boldsymbol{x}$ 进行归一化操作之后，就得到融合后的特征。为了加速计算和节省内存，本文采用文献[22]提出的紧凑双线性池化。池化层之后的输出是：

$$\boldsymbol{x} = \boldsymbol{f}_{RGB}^{\mathrm{T}} \boldsymbol{f}_N \tag{5}$$

其中，$\boldsymbol{f}_{RGB}$ 是 RGB 流的 RoI 特征，$\boldsymbol{f}_N$ 是噪声流的 RoI 特征。

1.6 Soft-NMS 算法

非极大值抑制算法[23]（Non-maximum suppression，NMS）是目标检测框架中的重要组成部分，主要用于去除冗余的建议框，找到最佳的目标检测位置。具体做法是将 RPN 推荐的建议框按照置信度得分排序，将得分最高的建议框作为候选框，删除与该框重叠面积比例大于设定阈值的其他建议框。为了解决在预设的重叠阈值之内篡改区域检测不到的问题，本文采用 Soft-NMS[16] 算法，该算法改良了传统 NMS 算法，对非最大得分的建议框检测分数进行衰减，降低了目标区域被漏检的概率。

传统的 NMS 的分数重置函数如下：

$$s_i = \begin{cases} s_i, & iou(M,b_i) < N_t \\ 0, & iou(M,b_i) \geqslant N_t \end{cases} \quad (6)$$

其中，s_i 表示置信度分数；M 表示当前得分最高的候选框；b_i 表示建议框；iou（Intersection over Union）表示交并比；N_t 表示 iou 阈值。

在 Soft-NMS 算法中，建议框 b_i 与候选框 M 重叠区域比例越大，出现漏检的可能性就越高，相应的分数衰减应该更严重，于是 Soft-NMS 中的分数衰减函数设计如下：

$$s_i = \begin{cases} s_i, & iou(M,b_i) < N_t \\ s_i(1 - iou(M,b_i)), & iou(M,b_i) \geqslant N_t \end{cases} \quad (7)$$

当 2 个建议框的 iou 大于设定的阈值时，s_i 的值就会相应减小，降低了因彻底移除而造成漏检的概率，从而达到检测精度的提升。

1.7 损失函数

图像经过特征提取网络的全连接层和 *softmax* 层之后得到了 RoI 区域，还需要对这些 RoI 区域做分类和边框回归。总的损失函数如下：

$$L_{total} = L_{RPN} + L_{tamper}(f_{RGB}, f_N) + L_{bbox}(f_{RGB}) \quad (8)$$

其中，L_{total} 表示总损失；L_{RPN} 表示 RPN 网络中的 RPN 损失；L_{tamper} 表示基于双线性池化特征的交叉熵分类损失；L_{bbox} 表示 bounding box 回归损失；f_{RGB} 和 f_N 是来自 RGB 和噪声流的 RoI 特征。

网络的训练是端到端的，输入的图像和提取的噪声特征的宽度调整为 600 像素。2 个支流 RoI 池化后的特征维度均为 7 * 7 * 1 024。双线性池化之后的特征尺寸为 16 384。训练过程中 RPN 推荐的 *batch size* 是 64，测试时设为 300。算法一共训练 110 000 次，初始学习率设置为 0.001，从第 40 000 步开始减小为 0.000 1，Soft-NMS 的阈值设为 0.2。

2 实验结果和分析

为了验证双流卷积注意力网络算法的有效性，本文在 CASIA[24-25]、COVER[26] 和 Columbia[27] 三个主流图像数据集上评估算法的性能。CASIA 数据集提供了多种物体的拼接和复制-粘贴操作，该数据集有 CASIA 1.0 和 CASIA 2.0 两个版本，其中 CASIA 1.0 包含 800 张真实图像和 921 张篡改图像，CASIA 2.0 包含 7 491 张真实图像和 5 123 张篡改图像。COVER 数据集是较小的复制-粘贴数据集，包含真实图像和篡改图像各 100 张。Columbia 数据集是未压缩的拼接数据集，包含 180 张拼接篡改图像，183 张真实图像。由于现有标准数据集的图片数量仍然较少，尚不能满足深度学习的训练过程，因此，本文在文献[13]合成的数据集进行预训练，Zhou 等学者在 COCO 数据集[28] 中复制图像内容后粘贴到其他图像上，复制的依据是图像的分割标注信息，真实图像和篡改图像各 42 000 张。

2.1 评价指标

本文使用 F_1 分数和 AUC 值来评估所提出的双流卷积注意力网络的性能。F_1 分数是将精确率（P）和召回率（R）结合的一种度量，精确率是指正确分类的正样本个数占分类器判定为正样本的样本个数的比例，见式(9)：

$$P = \frac{TP}{TP + FP} \quad (9)$$

召回率指分类正确的正样本个数占真正的正样本个数的比例，见式(10)：

$$R = \frac{TP}{TP + FN} \quad (10)$$

F_1 分数是精确率和召回率的调和平均值，见式(11)：

$$F_1 = 2\frac{P * R}{P + R} = \frac{2TP}{(2 * TP + FP + FN)} \quad (11)$$

其中，TP 为正确检测到的篡改像素数，FP 为错误检测到的篡改像素数，FN 为错误检测到的未篡改像素数。

F_1 分数越高，说明模型越稳健。AUC 值是 ROC 曲线下的面积值，AUC 值的大小反映模型泛化能力，AUC 值越大，模型泛化能力越强。

2.2 网络预训练

本文将合成数据集的 90%用来预训练，余下的用来测试。训练的过程是端到端的，特征提取网络分别对比使用了 CBAM - ResNet101 和改进的 CBAM-ResNet101。本文对比了文献[13]的预训练结果，见表 1，这里使用平均精度（Average Precision，*AP*）进行评估，结果表明精度有了明显提升。预训练之后，网络需要在公共数据集上做进一步训练，表 2 给出了训练集和测试集的划分。

表 1 合成数据集平均精度比较

Tab. 1 Comparison of the average accuracy of synthetic datasets

方法	*AP*
RGB Net	0.445
Noise Net	0.461
RGB-N	0.627
RGB-N+CBAM	0.685
Proposed	0.714

表 1 中，RGB Net 是一个单独的 RGB 网络，

Noise Net 是单独的噪声流网络，RGB-N 是双流网络，RGB-N+CBAM 为加入卷积注意力的算法，最后一行为是本文改进的算法。由表 1 中数据可知，单一的 RGB 流或噪声流提取的信息有限，双流网络综合 RGB 流和噪声流的特征信息后，平均精度有了明显提升。在双流网络中引入卷积注意力机制后，提取到的特征图包含更丰富的篡改特征信息，经过 Soft-NMS 算法降低漏检的概率后，平均精度有所提高。改进的注意力机制避免了降维带来的副作用，更有效地利用了不同通道间的依赖关系，进一步提升了网络的特征提取能力。

表 2　训练集和测试集的划分

Tab. 2　The division of training and testing sets

Step/Dataset	CASIA	Columbia	COVER
Training	5 123	-	75
Testing	921	180	25

2.3　结果对比

现有图像篡改取证方法分为传统算法和基于深度学习的算法，本文与以下方法进行对比分析。

(1) ELA[29]：识别图像中处于不同压缩因子的区域的算法。对于 JPEG 图像，整个图像应处于大致相同水平，如果某个区域压缩因子明显不同，则表示可能被篡改。

(2) CFA1[30]：基于 CFA 模型的评估算法。利用相邻像素来估算彩色滤波器阵列并推理出篡改区域。

(3) MFCN[9]：基于边缘强化的多任务图像被动取证框架。

(4) RGB-N[13]：融合噪声信息的双流神经网络算法。

(5) RGB+ELA[31]：基于双流 Faster R-CNN 的像素级图像拼接篡改定位算法。

本文采用 F_1 分数和 *AUC* 值对比上述 5 种算法，结果见表 3、表 4。表 3、表 4 的数据表明，基于深度学习的算法优于传统特征提取算法，原因是 ELA 和 CFA1 算法都只关注单一篡改特征，并且不能包含全部篡改信息。在深度学习算法中，本文所提算法表现优于 MFCN，在 CASIA 和 COVER 数据集表现优于 RGB-N。MFCN 性能较差的原因是采用小尺寸卷积核和上采样操作导致底层特征损失，因此对小区域篡改不敏感。RGB-N 采用大小不同的锚框(anchor)进行定位，较小区域的篡改也可以被检测到，本文在特征提取模块引入改进的 CBAM 注意力机制，并在预测时采用 Soft-NMS 降低漏检概率，检测结果在 3 个数据集上都有所提升。文献[31]通过将 SRM 滤波器替换为错误等级分析算法使提取到的噪声信息包含更多篡改信息，并添加一个预测分支做到了像素级分类。相比文献[31]，本文算法在 Columbia 数据集上略优，由于 CASIA 数据集拼接区域较为复杂，并且错误等级分析对篡改特征的提取效果优于 SRM，故本文算法性能略低于文献[31]算法。因为 COVER 数据集是复制粘贴数据集，所以来自噪声流提供的特征信息几乎失效，因此在该数据集表现较差。

表 3　3 个公共数据集 F_1 分数对比

Tab. 3　Comparison of F_1 scores from three public datasets

Method/Dataset	Columbia	COVER	CASIA
ELA[29]	0.470	0.222	0.214
CFA1[30]	0.467	0.190	0.207
MFCN[9]	0.612		0.541
RGB-N[13]	0.697	0.437	0.408
RGB+ELA[31]	0.745	-	0.665
RGB-N+CBAM	0.726	0.455	0.561
本文算法	0.763	0.480	0.633

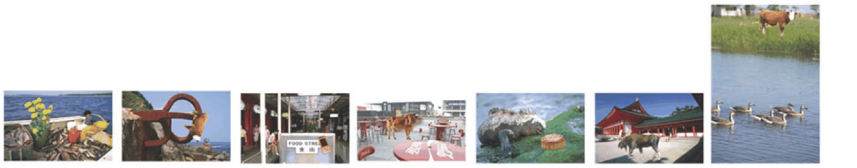
表 4　3 个公共数据集 *AUC* 值对比

Tab. 4　Comparison of *AUC* values for three public datasets

Method/Dataset	Columbia	COVER	CASIA
ELA[29]	0.581	0.583	0.613
CFA1[30]	0.720	0.485	0.522
MFCN[9]	-	-	-
RGB-N[13]	0.858	0.817	0.795
RGB+ELA[31]	-	-	-
RGB-N+CBAM	0.871	0.832	0.801
本文算法	0.905	0.856	0.818

2.4　检测结果分析

本文算法篡改检测定位效果如图 8 所示。图 8 中，(a) 表示拼接篡改图像，(b) 表示 ground-truth，(c) 表示文献[31]算法检测定位结果，(d) 表示 RGB-N+CBAM 定位结果，(e) 表示 RGB-N+ICBAM 定位结果。图像均来自于 CASIA 1.0 数据集。可视化结果显示，对于拼接边缘较为简单且篡改部分相对较小的区域如第 3 列和第 6 列，文献[31]所提算法和本文算法都能给出较为精确的定位结果，而对于拼接边缘较为复杂、且篡改部分相对较大的区域，文献[13]给出的可视化结果表现欠佳，也会存在未检测到的区域和检测错误的区域，本文算法则给出了篡改区域的矩形范围。改进后的注意力通过有效的通道注意力使提取到的篡改痕迹更加丰富，体现在可视化结果中表现为定位的矩形区域更加接近 Ground Truth。



(a) 拼接篡改图像

(b) ground truth

(c) 文献[13]算法

(d) RGB-N+CBAM

(e) 本文算法

图 8 拼接篡改定位可视化

Fig. 8 Visualization of image splicing detection

2.5 鲁棒性分析

为了验证本文算法的鲁棒性,在 CASIA1.0 数据库上利用质量因子 QF = 70 和 QF = 50 对图像进行 JPEG 压缩,表 5 给出了本文算法、文献[13]和文献[31]所提算法的 F_1 分数对比。结果显示,在 QF = 70 时,RGB-N 性能下降了 23.0%,文献[31]所提算法性能下降了 27.9%,在 QF = 50 时,RGB-N 性能下降了 26.3%,文献[31]所提算法性能下降了 31.7%,而本文所提算法通过在通道和空间维度对篡改痕迹进行更有效的特征提取,在 QF = 70 和 QF = 50 的情况下对比未压缩时分别下降 17.7%和 25.6%。从表 5 中可以进一步看出,除了在未压缩时 F_1 分数略低于文献[31]的算法,本文所提算法在 2 种不同的质量因子情况下性能均优于现有算法,说明本文算法能够更好地抵抗 JPEG 压缩攻击。

表 5 不同压缩因子下算法的 F_1 分数

Tab. 5 The F_1 score of the algorithm under different compression factors

Method	QF		
	100	70	50
RGB-N[13]	0.408	0.355	0.301
RGB+ELA[31]	0.665	0.479	0.453
本文算法	0.633	0.521	0.471

3 结束语

本文提出了一种双流卷积注意力网络对图像篡改区域进行检测和定位。首先,改进的卷积注意力机制能够抑制图片中无效信息,使提取到的特征更好地刻画伪造特性,双流网络加入噪声域信息可以学习更多丰富的特征;其次,通过引入 Soft-NMS 算法降低了伪造区域漏检的概率,提升了拼接篡改的检测精度。

实验结果表明，本文算法的检测性能优于一些现有算法，且对 JPEG 压缩也具有较好的鲁棒性。本文算法尚不能做到像素级定位，未来的工作将考虑改进当前网络，进一步精准定位篡改区域。

参考文献

[1] SADEGHI S, DADKHAH S, JALAB H A, et al. State of the art in passive digital image forgery detection: copy-move image forgery [J]. Pattern Analysis and Applications, 2017, 231: 284-295.

[2] HSU Y F, CHANG S F. Detecting image splicing using geometry invariants and camera characteristics consistency[C]//Proceedings of 2006 IEEE International Conference on Multimedia and Expo. Toronto, ON, Canada: IEEE, 2006.

[3] 朱新山，钱永军，孙彪，等. 基于深度神经网络的图像修复取证算法[J]. 光学学报，2018, 38(11): 105-113.

[4] FRIDRICH A J, SOUKAL B D, LUKÁŠ A J. Detection of copy-move forgery in digital images [C]//Proceedings of Digital Forensic Research Workshop. Cleveland, OH: [s.n.], 2003: 55-61.

[5] LI Haodong, LUO Weiqi, QIU Xiaoqing, et al. Image forgery localization via integrating tampering possibility maps [J]. IEEE Transactions on Information Forensics and Security, 2017, 12 (5): 1240-1252.

[6] KRIZHEVSKYA, SUTSKEVER I, HINTON G. ImageNet classification with deep Convolutional Neural Networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[7] YUAN Rao, NI Jiangqun. A deep learning approach to detection of splicing and copy-move forgeries in images [C]// 2016 IEEE International Workshop on Information Forensics and Security (WIFS). UAE: IEEE, 2017: 1-6.

[8] JOHNSON J, KARPATHY A, LI F F. Fully Convolutional networks for semantic segmentation [C] // Computer Vision & Pattern Recognition. New York: IEEE, 2015: 3431-3440.

[9] SALLOUM R, REN R, KUO C C J, et al. Image splicing localization using a Multi-task Fully Convolutional Network (MFCN) [J]. Journal of Visual Communication & Image Representation, 2018, 51: 201-209.

[10] BONDI L, LAMERI S, GUERA D, et al. Tampering detection and localization through clustering of camera-based CNN features [C]// Computer Vision & Pattern Recognition Workshops. Honolulu, HI, USA: IEEE, 2017: 1855-1864.

[11] BAPPY M, ROY-CHOWDHURY A K, BUNK J, et al. Exploiting spatial structure for localizing manipulated image regions [C]//Proceedings of IEEE International Conference on Computer Vision. Venice, Italy: IEEE Computer Society, 2017: 4980-4989.

[12] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.

[13] ZHOU Peng, HAN Xintong, 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 (CVPR). Salt Lake City, USA: IEEE, 2018: 1053-1061.

[14] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

[15] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2018: 3-19.

[16] BODLA N, SINGHB, CHELLAPPA R, et al. Soft-NMS --Improving object detection with one line of code [J]. arXiv preprint arXiv: 1704.04503, 2017.

[17] FRIDRICH J, KODOVSKY J. 1 Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.

[18] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition [J]. arXiv preprint arXiv: 1504.07889v1, 2015.

[19] HU Jie, SHEN Li, AlBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42: 2011-2023.

[20] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1-12.

[21] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.

[22] GAO Yang, BEIJBOM O, ZHANG Ning, et al. Compact bilinear pooling[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 317-326.

[23] NEUBECK A, GOOL L. Efficient non-maximum suppression [C]// Proceedings of International Conference on Pattern Recognition. Hong Kong, China: IEEE Computer Society, 2006: 1-6.

[24] DONG J, WANG W, TAN T. Casia image tampering detection evaluation database 2010 [EB/OL]. [2010]. http://forensics.idealtest.org.2, 5.

[25] DONG Jing, WANG Wei, TAN Tieniu. Casia image tampering detection evaluation database [C]// Proceedings of 2013 IEEE China Summit and International Conference on Signal and Information Processing. Beijing, China: IEEE, 2013: 422-426.

[26] WEN Bihan, ZHU Ye, SUBRAMANIAN R, et al. COVERAGE-A novel database for copy-move forgery detection [C]// Proceedings of International Conference on Information Processing (ICIP). Phoenix, Arizona, USA: IEEE, 2016: 161-165.

[27] HSUY F, CHANG S F. Detecting image splicing using geometry invariants and camera characteristics consistency[C]//Proceedings of International Conference on Multimedia and Expo (ICME). Toronto, Canada: IEEE, 2006: 549-552.

[28] LINK T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[M]//FLEET D, PAJDLA T, SCHIELE B, et al. Computer Vision-ECCV 2014. ECCV 2014. Lecture Notes in Computer Science. Cham: Springer, 2014, 8693: 740-755.

[29] RAWETZ N, SOLUTIONS H F. A picture's worth [J]. Hacker Factor Solutions, 2007, 6(2): 2.

[30] FERRARA P, BIANCHI T, ROSA A D, et al. Image forgery localization via fine-grained analysis of CFA artifacts [J]. IEEE Transactions on Information Forensics & Security, 2012, 7(5): 1566-1577.

[31] 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.