计算机科学
COMPUTER SCIENCE

基于Transformer的多任务图像拼接篡改检测算法

张婧媛，王宏霞，何沛松

引用本文

张婧媛，王宏霞，何沛松.基于Transformer的多任务图像拼接篡改检测算法[J]．计算机科学, 2023, 50(1): 114-122.

ZHANG Jingyuan, WANG Hongxia, HE Peisong. Multitask Transformer-based Network for Image Splicing Manipulation Detection [J]. Computer Science, 2023, 50(1): 114-122.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.211100269

基于 Transformer 的多任务图像拼接篡改检测算法

张婧媛　王宏霞　何沛松
四川大学网络空间安全学院　成都 610065
(jyzhang.z@foxmail.com)

摘　要　现有基于深度学习的图像拼接篡改检测方法大多依赖卷积操作的局部计算过程,感受野有限。此外,现有方法大多仅将篡改区域定位用于指导检测模型训练,难以学习更加丰富的篡改痕迹特征。针对上述局限性,提出了基于 Transformer 的多任务图像拼接篡改检测网络(Multitask Transformer-based Network,MT-Net),利用 Transformer 中的自注意力机制在特征提取过程获取图像像素之间的相关性,自适应地为各像素提供不同的关注度,提升检测网络对篡改痕迹的表征能力。此外,MT-Net 同时考虑多个子任务从局部细化和整体感知两个方面共同引导网络学习,包括篡改区域定位、篡改边缘定位和篡改比例预测,并根据子任务特点设计了对应的损失函数来指导网络进行优化。实验结果表明,相比现有算法,所提算法在 CASIA V2.0,Columbia 和 IDM2020 这 3 个公开数据集上均取得了更好的检测准确性,F1 值分别达到了 0.808,0.913 和 0.675。可视化检测结果图表明,所提算法在定位拼接篡改区域时也有较好的表现。

关键词:数字图像取证;图像拼接检测;Transformer;自注意力机制;多任务网络
中图法分类号　TP391

Multitask Transformer-based Network for Image Splicing Manipulation Detection

ZHANG Jingyuan,WANG Hongxia and HE Peisong
School of Cyber Science and Engineering,Sichuan University,Chengdu 610065,China

Abstract　Most of existing deep learning-based methods for image splicing forgery detection use convolutional layer for forensics feature extraction. However, convolution kernel conducts the local computation process with the limited reception field. Moreover, existing methods mainly apply the location of tampering regions to guide the detection model to train, and it is difficult to learn richer tamper trace features. To overcome above-mentioned limitations, a multitask transformer-based network(MT-Net) is proposed for image splicing detection and localization. The self-attention mechanism of Transformer is leveraged in encoder to learn the pixel correlation, which is able to provide different attention levels for pixels and makes the detection network pay more attention to tampering traces. Meanwhile, MT-Net considers three subtasks simultaneously to guide the detection network expose tampering traces from both local and global information, including tampered edge detection, tampered area detection and the prediction of the tampered area's proportion. Finally, three specific loss functions for their corresponding subtask are designed to better optimize the detection network in the training phase. In experiments, the proposed method(MT-Net) achieves better detection results compared with other state-of-the-art methods on three public available datasets, including CASIA v2.0, Columbia and IDM2020, where F1 scores are 0.808, 0.913 and 0.675 respectively. The visualization results also demonstrate that the proposed method has the better capability of localizing the splicing regions.

Keywords　Digital image forensics, Image splicing detection, Transformer, Self-attention mechanism, Multitask network

1　引言

近年来,社交网络的兴起使得图像的获取途径变得十分广泛。在互联网中,用户可以便利地下载或上传图像。另一方面,随着图像编辑技术的普及和成熟,用户无须学习相关专业知识就能够轻易地完成对图像内容的修改[1]。图像编辑后,处理痕迹不明显且图像内容逼真度高,难以通过肉眼识别图像真伪。如果不法分子对原始图片进行恶意篡改,并将

到稿日期:2021-12-29　返修日期:2022-03-01
基金项目:四川省科技计划(2022YFG0320);国家自然科学基金(61902263,61972269);中央高校基本科研业务费专项资金(YJ201881,2020SCU12066);中国博士后科学基金(2020M673276)
This work was supported by the Science and Technology Program of Sichuan Province(2022YFG0320), National Natural Science Foundation of China(61902263,61972269), Fundamental Research Funds for Central Universities of Ministry of Education of China(YJ201881,2020SCU12066) and China Postdoctoral Science Foundation(2020M673276).
通信作者:何沛松(gokeyhps@scu.edu.cn)

篡改图片通过互联网传播，不仅可能扭曲事实，甚至会破坏网络空间安全的稳定。因此，提出针对篡改图像的取证算法具有重要的研究意义和应用价值，可为鉴定图像完整性与真实性提供技术支撑。

图像拼接篡改（简称拼接篡改）是最常见的图像篡改方式之一，其通常使用两张或多张图像合成一张新的篡改图片。图像拼接篡改检测（Image Splicing Detection，ISD）需定位图像中的篡改区域。目前，现有 ISD 方法大致可分为两类：基于手工特征的方法和基于深度学习的方法。基于手工特征的方法对图像采集过程的信号处理过程进行建模，如设备噪声[2]、光响应不均匀性（Photo Response Nonuniformity，PRNU）[3]、色彩滤波阵列（Color Filter Array，CFA）[4]、相机响应函数（Camera Response Function，CRF）[5]等。此类方法针对性较强，需要手工设计检测特征，具有一定的局限性。随着深度学习技术在多个领域的成功应用，多媒体安全领域的学者也将其应用于图像篡改检测。基于深度学习的检测方法能够自适应地学习图像篡改特征，对图像进行像素级的篡改区域定位，其检测效率较手工特征方法有所提升。目前基于深度学习的方法大多采用卷积神经网络作为特征提取器，如 DF-Net[6]，RRU-Net[7]，ManTra-Net[8]等。然而，卷积操作仅对图像局部区域进行运算，难以表征远距离像素之间的依赖关系。由于卷积操作感受野有限，当待测图像具有较多篡改区域时，检测效果不佳。此外，现有方法大多仅定位篡改区域，网络学习的篡改痕迹信息也比较有限。一些方法尝试在网络中融合多种类型的特征（如色彩[9]、重采样[10]和噪声[11]等），以得到篡改痕迹的丰富表征。但目前的融合方式较为直接，如加性融合（特征图像进行像素级相加或堆叠）。但在实际取证场景中，网络提取的特征之间存在着冗余信息，并且各类特征对篡改定位的重要性不尽相同。因此，对特征不加区分的融合方式并不利于充分利用不同特征之间的互补性。

为了解决上述局限性，本文提出了基于 Transformer 的多任务拼接篡改检测网络（MT-Net），该检测网络采用编解码器（Encoder-Decoder）的架构。具体来说，MT-Net 的编码器利用 Transformer 中的自注意力机制捕捉图像像素之间的内部相关性，其解码器则用于改进特征融合方式，进行自适应特征融合，以避免冗余信息影响网络的学习效果。此外，MT-Net 采用多任务模型，包括像素级（Pixel-wise）子任务和图像级（Picture-wise）子任务。像素级子任务关注像素之间的差异，分别定位拼接边缘和拼接区域；图像级子任务预测篡改区域占比值，为网络提供整体感知，并根据子任务的特点设计对应损失函数以指导网络学习。本文的主要贡献有以下几点：

（1）构建了基于 Transformer 的多任务拼接篡改检测网络，即 MT-Net，将 Transformer 引入 ISD 任务，Transformer 内部的自注意力机制可有效获取像素之间的相关性，为像素分配差异化的注意力，提升网络对篡改痕迹的关注程度。

（2）针对篡改区域定位问题，优化了多任务检测网络的组合模式，包括像素级和图像级子任务。像素级子任务定位篡改边缘和篡改区域，使网络关注像素细节差异。此外，提出图像级的篡改比例预测模块（Proportion Prediction Module，PPM），实现对篡改区域的整体感知。同时，根据子任务特点设计了不同的损失函数，共同指导网络学习篡改特征。

（3）实验结果表明，本文提出的 MT-Net 相比现有方法有着更好的检测定位效果，在 3 个公开数据集上都取得了良好效果。此外，MT-Net 是端到端的结构，无须额外进行前处理或后处理操作。

2 相关工作

2.1 手工特征的图像拼接篡改检测

手工特征的拼接篡改检测方法需要具备特定的先验知识，手工特征从图像中遗留的痕迹提取而得，这些痕迹大致可分为图像采集过程产生的痕迹和图像编码过程产生的痕迹。

图像采集使用相机镜头进行，自然光首先经过凸透镜和光学滤波器得到光信号，再经过 CFA 选择性地通过单一色彩分量（红、绿或蓝）送入图像传感器，接着进行去马赛克处理（Demosaicing Process），得到其他两种颜色分量。整个过程中会留下固有痕迹，当图像被恶意拼接篡改后，会导致痕迹不一致。因此，Dirik 等[12]基于 CFA，检测相机留下的 CFA 伪影。Ferrara 等[4]假设篡改过程消除了由去马赛克处理带来的痕迹，提出新特征来描述局部去马赛克处理痕迹。此外，相机型号的差异会导致图像采集过程留下特定设备噪声线索。Mahdian 等[2]利用局部噪声标准差检测篡改区域和非篡改区域的差异。

图像编码过程多采用有损的 JPEG 压缩，减少了存储空间的占用。JPEG 压缩过程首先将图像划分为 8×8 块，然后采用 DCT 变换并进行量化，最后经过游程编码和熵编码得到压缩后的图像。该过程留下的固有痕迹也可用于检测篡改图像。Lin 等[13]利用 DCT 系数中的双量化效应，即篡改图像的 DCT 系数直方图中出现周期性波峰波谷的现象，提出了一种自动定位被篡改区域的方法。此外，图像经过固定质量因子的 JPEG 压缩后，会放大篡改区域和非篡改区域的不一致，误差级别分析（Error Level Analysis，ELA）[14]便利用该思想进行篡改检测。

基于手工特征的方法需借助研究者的先验知识。符合假设情况时，检测精度和检测效果良好。此类算法针对性较强，对特定类型篡改图像的检测精度高。然而，上述方法存在普适性受限，容易受到后处理操作的干扰。

2.2 深度学习的图像拼接篡改检测

近年来，深度学习快速发展，在诸多领域获得了成功应用，多媒体取证领域的学者已尝试将深度学习用于拼接篡改检测任务。现有方法利用卷积神经网络（Convolutional Neural Network，CNN）进行特征提取，并结合其他算法辅助进行检测。具体地，Rao 等[15]将 CNN 的第一层初始化为 Spatial Rich Model（SRM）[16]中滤波器的参数，进而约束特征提取过程。测试时使用训练好的 CNN 提取特征后再利用 SVM 进行分类。SG-Net[17]首先训练 Semi-Global 网络得到大致的预测图像，然后使用全连接条件随机场细化预测结果，以提高输入图像与网络输出图像的一致性。Bappy 等[18]引入长短期记忆网络（Long Short Term Memory，LSTM）构建不同区域的关联性，提出了 CNN-LSTM-CNN 的网络架构。该架构首先使用 CNN 进行特征粗提取，然后将得到的特征图分块后送入 LSTM 中进行特征精炼，最后使用一个卷积网络得到篡改区域的预测图。Deep Fusion Net[6]（DF-Net）首先训练

一组基本网络学习拼接篡改的不同特征，从中选择若干层组成 DF-Net 再进行微调。

上述方法均需要结合 CNN 和相应的辅助算法，无法最大程度地利用 CNN 强大的特征表达能力。对此，有学者尝试构建采用端到端训练方式的 CNN 检测器。RRU-Net[7]受人类大脑回忆和强化回忆机制的启发，提出了环形残差结构，能够强化卷积神经网络的学习效率。ManTra-Net[8]则设计了一种 Z-score 特征来捕获篡改图像的局部异常。Zhou 等[11]提出双分支的 Faster R-CNN 来融合 RGB 色彩特征和噪声特征。类似地，CAT-Net[19]结合 RGB 和 DCT 变换域特征来精确跟踪视觉线索和压缩痕迹特征，共同指导网络进行篡改区域检测。基于深度学习的方法依赖多层的卷积运算，但卷积关注图像局部，难以获得对图像的整体感知和远距离像素之间的依赖关系。

此外，现有方法大多仅定位拼接区域，网络学习的特征表达更加关注局部之间的差异，一些方法尝试进行多任务检测，促使网络学习更丰富的篡改痕迹信息。MFCN[20]基于 FCN 进行实现，可同时定位篡改区域和篡改边界，但如果区域预测结果和边界预测结果存在较大差异，则需在两者间进行取舍。Kniaz 等[21]使用生成对抗网络(Generative Adversarial Network，GAN)模拟图像的编辑过程，再经过网络完成拼接区域定位、拼接边缘定位以及语义区域定位 3 个子任务，引导网络学习丰富的篡改特征。同样地，Bi 等[22]认为，编码器学习到的特征信息越多，重构图像质量就越好。因此，在拼接区域定位和拼接边缘定位的基础上增加了图像重构子任务，并引入小波池化对特征进行压缩和重构。然而，现有多任务方法的各子任务均为像素级，着重关注像素之间的差异，缺少对篡改区域的整体感知。因此，本文方法也尝试对这一局限性进行改进。

3 图像拼接篡改检测算法

3.1 算法提出的动机

现有基于深度学习的拼接篡改检测方法大多使用卷积操作提取特征。为了提高运算效率，卷积核大小多为小尺寸，如 3×3 和 5×5 等。这导致特征提取过程类似一种对图像的局部滤波运算。若篡改区域面积较大，卷积操作在篡改区域内部运算时，将无法有效捕捉篡改区域和非篡改区域之间的差异。Transformer 是近年来在计算机视觉领域广泛应用的网络模块，其内部由自注意力机制组成，能够有效地对图像所有像素之间的相关性进行表达，并提供不同权重。因此，MT-Net 将采用 Transformer 构建编码器进行特征提取，有利于对篡改图像中不同区域的像素建立依赖关系，提升对篡改痕迹的感知能力。

另一方面，现有检测方法虽然考虑多种特征并进行融合，但方式单一。多特征结合的方式可为检测网络带来互补信息，但不同特征的来源之间往往存在差异，只进行简单融合无法有效挖掘多特征间的有效信息，进而影响网络学习效果。针对这一局限性，MT-Net 将引入自适应特征融合机制，避免冗余信息，提高有效特征的利用率。

此外，实际应用场景中，篡改图像的拼接区域往往具有明确的语义，如人物和动物等。若未充分考虑语义信息，将导致检测网络受到语义信息干扰，例如瓶子上的标签被篡改，但检测网络的预测结果为整个瓶子。针对该问题，MT-Net 提出了一种篡改区域比例预测模块，以增加对篡改区域的整体约束。

3.2 算法整体框架

本文提出的 MT-Net 结构如图 1 所示，在多任务学习框架下可同时预测拼接边缘、拼接区域和拼接区域占比值，共 3 个子任务。具体来讲，该检测网络包括编码器 E、边缘解码器 D_{edge}、区域解码器 D_{area} 和拼接区域比例预测模块 PPM 4 个部分，其中 D_{edge} 和 D_{area} 结构相同。首先，输入待检测图像 I，编码器对图像篡改特征进行学习得到多尺度特征 $\boldsymbol{F}_i$，并融合图像高频噪声特征 $\boldsymbol{N}_i$。其次，编码器的多尺度特征将在两个解码器中进行特征融合，并按照不同子任务解码特征得到边缘检测图 $\boldsymbol{M}_{\text{e}}$ 和区域检测图 $\boldsymbol{M}_{\text{a}}$，$D_{\text{area}}$ 将复用 D_{edge} 得到的结果。最后，D_{area} 的多尺度特征将送入 PPM 进行篡改区域占比预测并得到最终预测结果 P_p，为网络添加整体约束。

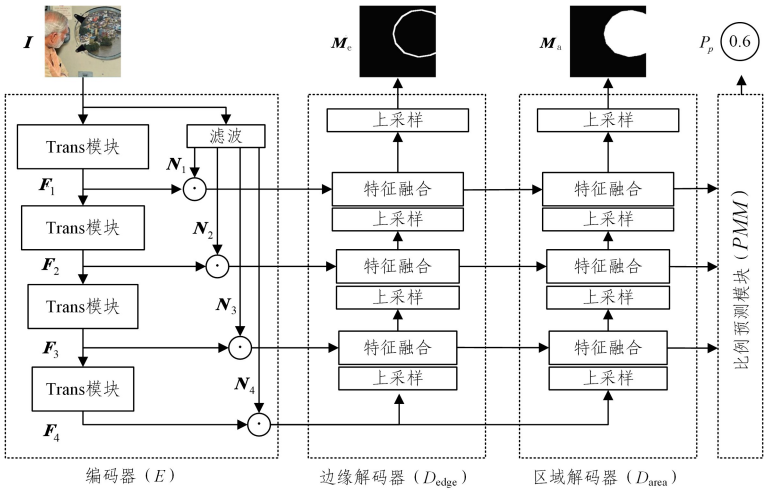

图 1 MT-Net 的结构

Fig. 1 Structure of MT-Net

3.3 基于 Transformer 的编码器

现有篡改检测网络大多基于卷积神经网络进行特征学习，但卷积操作是局部区域运算，即使堆叠多层也难以获得良好的全局感知能力。为弥补上述局限性，本文引入基于

Transformer 的编码器来提取图像特征。Transformer 自 2017 年由 Google 提出[23]以来，已成功应用于自然语言处理、图像分类等领域。由于 Transformer 内部结构使用自注意力机制实现，减少了该编码器对外部信息的依赖，并且增强了编码器对像素之间内部相关性的学习，能有效捕捉像素点之间的全局依赖关系。目前基于 Transformer 的网络大多需要对图像分块进行位置编码，为每个块嵌入其对应的绝对位置信息。该过程与图像大小密切相关，因此网络训练和测试时所使用的图像需尺寸一致。而在实际的篡改检测场景下，网络需检测不同尺寸的图像，对此本文基于 SegFormer[24]设计编码器结构，该编码器使用一个 3×3 的卷积操作为编码器提供位置信息，代替原始的绝对位置信息编码方式。通过上述改进，本文的 MT-Net 可测试不同尺寸的图像。此外，编码器内的分块处理机制使得编码器具有层次化的信息处理能力，可输出多个尺度的特征图，包括低维特征（如轮廓、纹理等）和高维特征（如语义），有助于细节特征的学习和像素级的定位。

具体地，编码器 E 由 4 个 Trans 模块（TransBlock）和 1 个滤波模块组成。其中，每个 Trans 模块结构相同，分别输出不同尺度大小的特征图 $\boldsymbol{F}_i \in \mathbb{R}^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}}$（$i=1,2,3,4$），其中 H 和 W 分别为输入图像 $\boldsymbol{I}$ 的长度和宽度。参照文献[16]，从滤波模块中选取 5 个固定参数的高频滤波器，具体参数如图 2 所示。该模块可用于提取高频噪声特征，得到的特征 N_i 将下采样至与对应 $\boldsymbol{F}_i$ 尺寸一致，即 $N_i \in \mathbb{R}^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}}$（$i=1,2,3,4$），并将 N_i 与 $\boldsymbol{F}_i$ 堆叠后送入解码器，为网络提供额外特征。Trans 模块内部结构如图 3 所示，包括重叠块嵌入模块（Overlap Patch Embedding，OPE）、N 个自注意力模块（Self-Attention，SA）和前反馈模块（Mix Feed Forward，MFF）。参考文献[24]中的设置，本文实验中 4 个 Trans 模块对应的 $N=[3,8,27,3]$。每个 Trans 模块可表示为：

$$\boldsymbol{F}_{i+1} = MFF(SA(OPE_{k,s,p}(\boldsymbol{F}_i))) \tag{1}$$

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

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 3 & -1 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

图 2 Filter 模块中滤波器参数

Fig. 2 Parameters of five filters in Filter module

$OPE(\cdot)$ 使用滑动窗口得到重叠块，k，s，p 分别表示块的大小、步长和填充大小。本文实验中，$k=7$，$s=4$，$p=3$。$SA(\cdot)$ 内部由自注意力机制实现，可表示为：

$$x_{\text{out}} = \boldsymbol{x}_{\text{in}} + \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_{\text{head}}}}\right)\boldsymbol{V} \tag{2}$$

其中，$\boldsymbol{x}_{\text{in}}$，$\boldsymbol{x}_{\text{out}}$ 分别表示 $SA(\cdot)$ 的输入和输出；$\boldsymbol{Q}$，$\boldsymbol{K}$，$\boldsymbol{V}$ 是 $\boldsymbol{x}_{\text{in}}$ 的线性映射，其维度均为 $\mathbb{R}^{N\times C}$；$N=H\times W$；C 为 $\boldsymbol{x}_{\text{in}}$ 的通道数。$MFF(\cdot)$ 直接使用 3×3 的卷积为编码器提供位置信息，无需绝对位置编码，因此测试图像尺寸不必与训练图像保持一致。$MFF(\cdot)$ 可表示为：

$$\boldsymbol{y}_{\text{out}} = \boldsymbol{y}_{\text{in}} + MLP(GELU(Conv_{3\times3}(MLP(\boldsymbol{y}_{\text{in}})))) \tag{3}$$

其中，$\boldsymbol{y}_{\text{in}}$，$\boldsymbol{y}_{\text{out}}$ 分别表示 $MFF(\cdot)$ 的输入和输出；$MLP(\cdot)$ 表示多层感知机（Multilayer Perceptron，MLP）；$GELU(\cdot)$ 为高斯误差线性单元[25]；$Conv_{3\times3}(\cdot)$ 是大小为 3×3 的卷积（步长为 1），为重叠块提供位置信息。

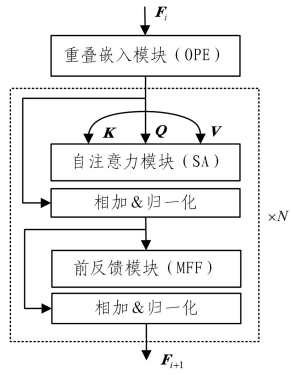

图 3 编码器中单个 Trans 模块的结构

Fig. 3 Structure of one Trans block in encoder

3.4 基于自适应特征融合的解码器

本文使用的解码器包括边缘解码器 D_{edge} 和区域解码器 D_{area}，分别用于检测篡改边缘和篡改区域。两个编码器的结构相同，解码器对编码器提取的特征进行融合并逐步上采样至与输入图像一致的大小，即可输出边缘检测图 $\boldsymbol{M}_{\text{e}}$ 和区域检测图 $\boldsymbol{M}_{\text{a}}$。目前大多数网络中的特征融合方式较为固定，多采用加性融合机制（如特征按元素相加或堆叠）。然而，对 ISD 任务而言，编码器提取的特征并非全部表征篡改痕迹，特征之间存在一定冗余信息，直接加性融合可能会引入额外的噪声，从而干扰网络学习。对此，本文使用自适应特征融合机制[26]学习重要特征，避免冗余信息的影响，其结构如图 4 所示，每个解码器包括 3 个自适应特征融合模块。相比简单的加性融合机制，该机制首先对两个特征相乘得到特征中相似的部分，再将上述结果特征通过残差连接的方式与原始特征相加，最后对特征进行精炼提取。该过程是对特征的一种自适应融合，能够有效地融合低维和高维特征。

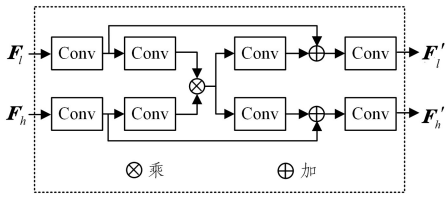

图 4 自适应特征融合模块

Fig. 4 Adaptive feature fusion module

该模块的输入为两个尺寸的特征，记为低维特征和高维特征，分别表示为 $\boldsymbol{F}_l$ 和 $\boldsymbol{F}_h$，则对应输出特征为 $\boldsymbol{F}_l'$ 和 $\boldsymbol{F}_h'$。由于编码器输出的 $\boldsymbol{F}_l$ 和 $\boldsymbol{F}_h$ 尺寸不同，因此在进行特征融合前需将 $\boldsymbol{F}_l$ 上采样至与 $\boldsymbol{F}_h$ 尺寸一致，即 $UP(\boldsymbol{F}_l)$，$UP(\cdot)$ 为上采样过程，本文使用双线性插值进行上采样。$\boldsymbol{F}_l'$ 和 $\boldsymbol{F}_h'$ 可表示为：

$$\boldsymbol{F}_h' = D_h[A_h(\boldsymbol{F}_h) + C_h[B_h(A_h(\boldsymbol{F}_h)) \times B_l(A_l(UP(\boldsymbol{F}_l))]] \tag{4}$$

$$\boldsymbol{F}_l' = D_l[A_l(\boldsymbol{F}_l) + C_l[B_h(A_h(\boldsymbol{F}_h)) \times B_l(A_l(UP(\boldsymbol{F}_l))]] \tag{5}$$

其中，A_l，B_l，C_l，D_l，A_h，B_h，C_h，D_h 为“Conv-BN-ReLU”的组合模块，分别表示 3×3 的卷积、批处理归一化和修正线性单元。下标 l 代表低维特征所经过的分支，h 代表高维特征所经过的分支。当前特征融合模块输出的 $\boldsymbol{F}_h'$ 将作为下一个特征融合模块的高维特征与来自编码器的低维特征进行融合。同时，D_{edge} 每层输出的 $\boldsymbol{F}_l'$ 将被送入 D_{area} 进行特征融合，而 D_{area} 每层输出的 $\boldsymbol{F}_l'$ 将被送入 PPM 进行篡改比例预测。

3.5 篡改比例预测模块

在 ISD 任务中，大多数方法进行像素级的定位（如预测篡改内容的边缘和内部区域等），网络更多地是关注细节上的局部内容不一致性，缺乏对篡改区域的整体感知。并且受数据影响，网络偏向预测带语义篡改区域。为更准确地定位篡改区域，本文提出了篡改比例预测模块（PPM），以增加网络的整体性约束。通过对解码器 D_{area} 输出的多尺度特征进行降维和融合，从图像全局对篡改区域的占比值进行预测，因此篡改比例预测的过程可看作一个图像级的子任务，增强网络关于图像篡改区域的整体感知能力。

PPM 模块的输入为多个不同尺度的特征图 $\boldsymbol{f}_i \in \mathbb{R}^{\frac{H}{2^{i+1}} \times \frac{W}{2^{i+1}}}$（$i=1,\cdots,4$），其结构如图 5 所示。

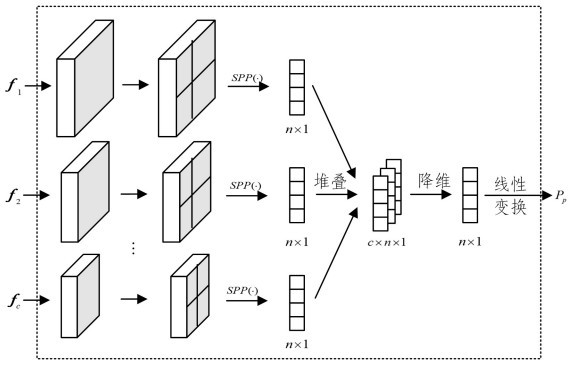

图 5 篡改比例预测模块 PPM

Fig. 5 Tamper proportion prediction module(PPM)

首先，各层特征图 $\boldsymbol{f}_i$ 经过 $SPP(\cdot)$ 生成 $\boldsymbol{f}_i' \in \mathbb{R}^{n_i \times 1}$（$n_i = 1^2, 2^2, 3^2, 4^2$），每个特征向量都是原始篡改特征在不同尺度下的表示。其中，$SPP(\cdot)$ 为空间金字塔池化（Spatial Pyramid Pooling，SPP）[27]。接着，对这些特征向量进行堆叠，得到多通道特征向量 $\boldsymbol{f} \in \mathbb{R}^{c \times n \times 1}$。然后，经过 1×1 的卷积对 $\boldsymbol{f}$ 进行降维，此时维度为 $\mathbb{R}^{1 \times n \times 1}$。最后输出预测结果的概率表示 P_p。该过程可表示为：

$$\boldsymbol{f}_i' = SPP(\boldsymbol{f}_i) \tag{6}$$

$$\boldsymbol{f} = Cat[\boldsymbol{f}_1', \boldsymbol{f}_2', \cdots, \boldsymbol{f}_i'] \tag{7}$$

$$P_p = \sigma(Conv_{1\times1}(\boldsymbol{f})) \tag{8}$$

其中，$Cat(\cdot)$ 对特征按通道进行堆叠；$Conv_{1\times1}(\cdot)$ 代表卷积核大小为 1×1；$\sigma(\cdot)$ 为 Sigmoid 激活函数，能将预测的输出值映射到[0,1]区间范围内作为拼接区域占比的概率表示。篡改区域占比的预测结果将与其对应真值进行交叉熵损失计算，以增强 MT-Net 对篡改区域整体性的约束，与本文的像素级子任务（D_{edge} 和 D_{area}）一同指导网络学习拼接篡改产生的异常痕迹。

3.6 损失函数

本文提出的检测网络的损失函数包括 3 个部分，分别是拼接边缘检测子任务中 D_{edge} 的损失 $\mathcal{L}_e$，拼接区域检测子任务中 D_{area} 的损失 $\mathcal{L}_a$ 和篡改比例预测（PPM）的损失 $\mathcal{L}_p$。其中，$\mathcal{L}_e$ 使用二分类交叉熵（Binary Cross Entropy，BCE）表示预测篡改边缘和实际篡改边缘之间的误差。$\mathcal{L}_a$ 在 BCE 的基础上结合交并比（Intersection over Union，IOU）来衡量预测篡改区域与实际篡改区域之间的误差。IOU 损失函数[28]使网络关注篡改区域而不是单一像素点，在样本类别不平衡状态下也有着良好的表现。$\mathcal{L}_p$ 表示预测篡改区域占比值与实际占比值之间的绝对值误差。3 个损失函数的定义如下：

$$\begin{cases} \mathcal{L}_e = -\sum_{i=1}^{H}\sum_{j=1}^{W}[G_e \log(\boldsymbol{M}_e) + (1-\boldsymbol{G}_e)\log(1-(\boldsymbol{M}_e))] \\ \mathcal{L}_a = -\sum_{i=1}^{H}\sum_{j=1}^{W}[G_a \log(\boldsymbol{M}_a) + (1-\boldsymbol{G}_a)\log(1-(\boldsymbol{M}_a))] + \\ \qquad \left[1 - \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{M}_a\boldsymbol{G}_a}{\sum_{i=1}^{H}\sum_{j=1}^{W}[\boldsymbol{M}_a + \boldsymbol{G}_a - \boldsymbol{M}_a\boldsymbol{G}_a]}\right] \\ \mathcal{L}_p = |G_p - P_p| \end{cases} \tag{9}$$

其中，$\boldsymbol{M}_e \in \mathbb{R}^{H\times W}$，$\boldsymbol{M}_a \in \mathbb{R}^{H\times W}$ 和 $P_p \in \mathbb{R}$ 分别代表预测的篡改边缘检测图、篡改区域检测图和篡改区域占比值，$\boldsymbol{G}_e \in \mathbb{R}^{H\times W}$，$\boldsymbol{G}_a \in \mathbb{R}^{H\times W}$ 和 $G_p \in \mathbb{R}$ 分别代表对应的实际篡改边缘图、实际篡改区域图和实际篡改区域占比值。$\mathcal{L}_e$ 和 $\mathcal{L}_a$ 用于像素级子任务，因此 (i,j) 代表图像像素坐标，H 和 W 分别对应图像的长和宽。总损失函数的定义如下：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_a + \lambda \cdot \mathcal{L}_e + \mu \cdot \mathcal{L}_p \tag{10}$$

其中，$\lambda(0 \leqslant \lambda \leqslant 1)$ 和 $\mu(0 \leqslant \mu \leqslant 1)$ 分别代表 D_{edge} 和 PPM 两个子任务的权重参数，具体取值将在后续实验中进行验证。考虑到 D_{area} 子任务对 ISD 任务具有较大重要性，其 $\mathcal{L}_a$ 的权重参数设置为 1。

4 实验结果及分析

4.1 实验设置

（1）实验数据。为验证本文方法在 ISD 任务上的有效性，本文使用常见开源 ISD 数据集 CASIA V2.0[29]，Columbia[30]，IDM2020[31] 作为实验数据集，分别包括 1 849，180 和 2 010 张拼接篡改图像。为了提高训练效率，3 个数据集图像分别裁剪（crop）为 256×384，256×256 和 256×256，裁剪后的图像按照 6∶2∶2 的比例随机划分为训练集、验证集和测试集。为模拟实际应用场景中数据的多样性，参照文献[7]中的方法，对训练集进行了数据增强，包括 5 种数据增强方式：随机上下或左右翻转、随机旋转（旋转角度范围为 30°～150°，间隔为 30°）、随机高斯噪声（方差范围为 0.002～0.01，间隔为 0.02）、随机质量因子的 JPEG 压缩（质量因子范围为 50～100，间隔为 10）、高斯模糊（高斯核大小为 3×3）。最终，CASIA V2.0 中训练、验证和测试图像数量分别为 6 654，369 和

371。Columbia 中训练、验证和测试图像数量分别为 17 280，399 和 384。IDM2020 中训练、验证和测试图像数量分别为 26 982，1578 和 1 452。

（2）实验环境。实验中使用的深度学习框架为 PyTorch1.6.0，Python 版本为 3.7，CUDA 版本为 10.1。所有实验均在 Intel i7-8700K CPU 和单块 RTX 2080 Ti 显卡上进行，训练时批（batch）大小为 32，训练迭代轮次为 50 个 epoch，共迭代约 14 000 次。初始化学习率为 0.05 并使用随机梯度下降（SGD）优化器，学习率随着训练轮次的增加逐渐下降。编码器的参数设置与 SegFormer[24] 中一致。

（3）评价指标。本文采用准确率（Accuracy，ACC）、召回率（Recall）、精确率（Precision）、F1 值和马修斯相关系数（Matthews Correlation Coefficient，MCC）[17] 作为评价指标来衡量方法的性能。所有评价指标均为像素级，使用篡改区域检测结果图 M_a 和实际篡改区域图 G_a（Ground Truth）进行计算。

$$ACC=\frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

$$Recall=\frac{TP}{TP+FN} \tag{12}$$

$$Precision=\frac{TP}{TP+FP} \tag{13}$$

$$F1=\frac{2\times Precision\times Recall}{Precision+Recall} \tag{14}$$

$$MCC=\frac{(TP\times TN-FP\times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{15}$$

其中，TP 表示检测正确的拼接区域像素数量，TN 表示检测正确的非篡改区域像素数量，FP 表示检测错误的非篡改区域像素数量，FN 表示检测错误的篡改区域像素数量。MCC 在样本不平衡的情况下也能够有效评价网络的分类效果。

4.2 消融实验

本实验将研究 MT-Net 中各模块对最终检测定位性能的影响。实验数据集采用 CASIA V2.0，实验结果如表 1 所列。其中，*edge_area* 包括拼接边缘定位和拼接区域定位两个子任务，*edge_area_filter* 表示在 *edge_area* 的基础上，编码器加入滤波模块，即添加高频噪声特征，*edge_area_ppm* 表示在 *edge_area* 的基础上，添加 *PPM* 模块。*edge_area_filter_ppm* 为本文提出的 MT-Net 的结构进行了多任务的拼接篡改检测。由实验结果可以看出，在网络中添加相关子任务时，整体检测性能有所提升，其中 F1 值和 MCC 值分别提高了约 2%和 3%。这说明本文提出的多个模块均对网络整体性能提升均有一定的贡献，并且实验结果也表明本文提出的多任务检测网络在 ISD 任务中能取得令人满意的检测效果。

表 1 在 CASIA V2.0 数据集中各模块对检测结果的影响

Table 1 Influence of each module on detection results in CASIA V2.0 dataset

Methods	*ACC*	*F1*	*MCC*
edge_area	0.9631	0.7819	0.7675
edge_area_filter	0.9656	0.8056	0.7931
edge_area_ppm	0.9654	0.8075	0.7921
edge_area_filter_ppm	**0.9668**	**0.8081**	**0.7997**

然后，本实验将通过调节损失函数中不同权重参数 λ 和 μ 的取值，研究不同子任务（D_{area}，D_{edge} 和 *PMM*）对算法性能的影响。本实验采用 CASIA V2.0 数据集，实验结果如表 2 所列。其中，λ 和 μ 分别代表 $\mathcal{L}_e$ 和 $\mathcal{L}_p$ 的权重参数，分别对应 D_{edge} 和 *PPM* 两个子任务，考虑拼接区域定位子任务的重要性，D_{area} 的权重参数均设置为 1。实验结果表明，当降低 $\mathcal{L}_e$（拼接边缘检测子任务）和 $\mathcal{L}_p$（拼接区域占比值）中任意一部分的权重时，检测网络的性能均有所下降。当 3 个损失函数（$\mathcal{L}_a$，$\mathcal{L}_e$ 和 $\mathcal{L}_p$）的权重参数均为 1 时，网络的检测性能达到最优，说明各子任务对检测网络均有贡献并且贡献程度相当，能够相互补充共同起到优化作用。

表 2 损失函数中不同比例系数的实验对比结果

Table 2 Comparison of detection results of different proportional coefficients in loss function

λ	μ	*ACC*	*F1*	*MCC*
0.5	0.5	0.9577	0.7914	0.7785
0.5	1	0.9629	0.7995	0.7858
1	0.5	0.9625	0.7958	0.7822
1	1	**0.9668**	**0.8081**	**0.7997**

接下来，本实验对特征提取部分采用的结构进行分析，验证基于 Transformer 的编码器所提取的检测特征是否更适用于篡改检测和定位。为此，本实验将采用其他编码器实现方式进行对比，结果如表 3 所列。其中，Res-based 表示检测网络中的编码器使用 ResNet-50[32]，而 Trans-based 表示本文使用的编码器结构，其中特征提取部分基于 SegFormer[24] 实现。为了公平比较，两者均输出 4 个不同尺度的特征图 $\boldsymbol{F}_i$（$i=1,2,3,4$）用于后续特征融合。由表 3 中的数据可以看出，本文提出的检测网络使用基于 Transformer 的编码器进行特征提取，在 3 个数据集上均取得了明显的性能提升，优于基于 ResNet-50 的编码器。上述结果说明 Transformer 中自注意力机制适用于 ISD 任务，能够学习像素之间的相关性，得到远距离像素之间的依赖关系，并为像素分配权重，使得网络可以更多地关注篡改痕迹。此外，实验结果也表明，基于 Transformer 的编码器能够有效地捕捉篡改痕迹，所提取的特征更有利于进行 ISD 任务。

表 3 基于不同编码器的检测结果对比

Table 3 Comparison of detection results under different encoder

Dataset	Methods	*ACC*	*F1*	*MCC*
CASIA V2.0	Res-based	0.9494	0.6966	0.6778
	Trans-based	**0.9668**	**0.8081**	**0.7997**
Columbia	Res-based	0.9722	0.9083	0.7652
	Trans-based	**0.9803**	**0.9131**	**0.7797**
IDM2020	Res-based	0.8697	0.5921	0.4870
	Trans-based	**0.8905**	**0.6756**	**0.6025**

如图 6 所示，实验还分析了检测网络包含不同子任务时，在 CASIA V2.0 验证集中检测准确率随迭代次数的变化趋势。具体地，$T(e\_a)$，$T(e\_a\_p)$，$T(e\_a\_f)$ 和 $T(e\_a\_f\_p)$ 表示不同的模块组成，并且编码器的特征提取基于 Transformer 实现。而在 $R(e\_a\_f\_p)$ 中，编码器的特征提取基于 ResNet-50 实现。其中，下标 e 代表 D_{edge} 子任务，a 代表 D_{area} 子任务，f 代表 *Filter* 模块，p 代表 *PPM* 子任务。由图 6 可知，检测

网络的准确率在前4000次迭代中损失下降速度较快，并且在迭代8000次后变化趋于平稳，这说明整个检测网络能够有效地收敛。此外，可以看到本文基于多任务的检测网络对应的准确率，即 $T(e\_a\_f\_p)$ 相对较高。另一方面，基于 ResNet-50 实现的检测网络整体检测准确率明显低于基于 Transformer 实现的方法，并且在迭代6000次左右时检测准确率曲线出现了明显波动。这说明采用 Transformer 能够提取更加丰富的篡改特征，且在训练过程中更加稳定。

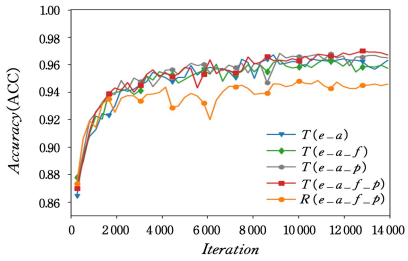

图6 CASIA V2.0 中不同任务下的检测准确率上升曲线

Fig. 6 Accuracy curve of different missions in CASIA V2.0 dataset

4.3 与其他方法对比

本文在3个数据集上分别与其他6种典型方法进行对比。其中，NOI[2]，CFA[4]和 ELA[14]为常用的手工特征检测方法，利用取证先验知识手工设计检测特征。RRU-Net[7]，CAT-Net[19]和 MWC-Net[22]均为基于深度学习的检测方法，使用卷积神经网络提取篡改特征，具体对比结果如表4所列。下文将简单介绍这些方法。

(1)NOI[2]：引入了一种检测噪声水平变化的分割方法，利用局部噪声标准差检测篡改区域和非篡改区域的差异。

(2)CFA[4]：利用颜色滤波阵列(CFA)伪影的色彩模式和插值模式作为相机特征进行检测。

(3)ELA[14]：对有损压缩(如 JPEG 压缩)图像中的压缩伪影进行分析检测篡改。

(4)RRU-Net[7]：在编码器和解码器中引入环形残差结构，放大拼接区域和非拼接区域之间的差异进行检测。

(5)CAT-Net[19]：提出双分支网络分别提取 RGB 色彩特征和 JPEG 压缩特征进行检测。两个分支的输入分别为原始图像和 Y(亮度)通道的 DCT 系数以及量化表。

(6)MWC-Net[22]：采用多任务网络，包括拼接区域定位、拼接边缘定位和原图重构3个子任务。引入小波池化(Wavelet-pooling)精炼特征。

表4 在 CASIA V2.0，Columbia 和 IDM2020 数据集上不同方法的检测结果

Table 4 Detection results of different methods in CASIA V2.0, Columbia and IDM2020 datasets

Methods	CASIA V2.0			Columbia			IDM2020		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
ELA	0.285	0.095	0.103	0.715	0.042	0.072	0.367	0.048	0.060
NOI	0.086	0.078	0.060	0.638	0.106	0.132	0.370	0.150	0.151
CFA	0.144	0.004	0.007	0.299	0.128	0.110	0.282	0.006	0.010
RRU-Net	0.621	0.532	0.573	0.893	0.610	0.725	0.490	0.415	0.449
CAT-Net	0.743	0.879	0.689	0.858	0.901	0.802	0.534	0.899	0.535
MWC-Net	**0.837**	0.831	**0.834**	0.855	0.803	0.828	—	—	—
Ours	0.798	**0.944**	0.808	**0.940**	**0.958**	**0.913**	**0.650**	**0.930**	**0.675**

注：“—”表示文献[22]未将 MWC-Net 在 IDM2020 数据集上进行实验

由表4可知，本文方法在大部分情况下能够取得更好的检测效果，尤其是在 Columbia 和 IDM2020 数据集上，Precision，Recall 和 F1 都表现出了更好的性能；在 CASIA V2.0 数据集上，本文方法的 Recall 值比 MWC-Net 提升了约11%；在 IDM2020 数据集上，本文方法的 F1 值达到0.675。这是由于本文方法是多任务架构，可从整体和细节共同关注篡改痕迹。同时，网络可自适应地学习检测特征，当网络收敛时可学习多元的层次化特征表达，对篡改痕迹的捕捉能力更强。此外，本文对比了基于深度学习的方法的时间复杂度，其结果如表5所列。实验结果表明，本文方法的单个迭代轮次(epoch)耗时较短，训练速度优于对比方法。测试时，本文方法的单张图片处理耗时约0.08 s，这是由于本文方法同时预测篡改边缘、篡改区域和篡改区域占比值，因此时间消耗略多于其他方法，但仍在可接受范围内。总的来说，本文方法所取得的性能提升主要得益于：1)Transformer 中的自注意力机制能够学习像素之间的相关性，从而为像素分配权重，使得网络能够更好地关注到篡改痕迹而不是图像的视觉特征；2)自适应的融合机制避免了冗余信息，使特征更加精炼；3)本文检测网络采用的多任务架构既能关注局部拼接边缘的痕迹线索，又能感知拼接区域在图像上的整体分布。与各子任务特点对应的损失函数也使得检测网络在训练过程中得到更好的优化，最终使整体收敛到最优结果。

表5 在 CASIA V2.0 数据集上不同方法的训练时间和测试时间

Table 5 Training time and testing time of different methods in CASIA V2.0 dataset

(单位：s)

Methods	Training time of single epoch	Testing time of single image
RRU-Net	421.57	**0.03**
CAT-Net	869.76	0.31
MWC-Net	465.63	0.05
Ours	**285.95**	0.08

图7为在3个数据集上各方法的可视化结果图。从每个数据集中随机选取一张示例图像，对于每种方法输出的结果图，白色部分表示预测的篡改区域像素，黑色部分表示预测的非篡改区域像素，GT 表示实际篡改区域的 Ground Truth 图像。换句话说，检测算法预测的篡改区域与 GT 越接近，说明算法检测的准确率越高。由图7可知，本文提出的 MT-Net 在定位拼接区域时在细节和整体上都有着更好的表现。

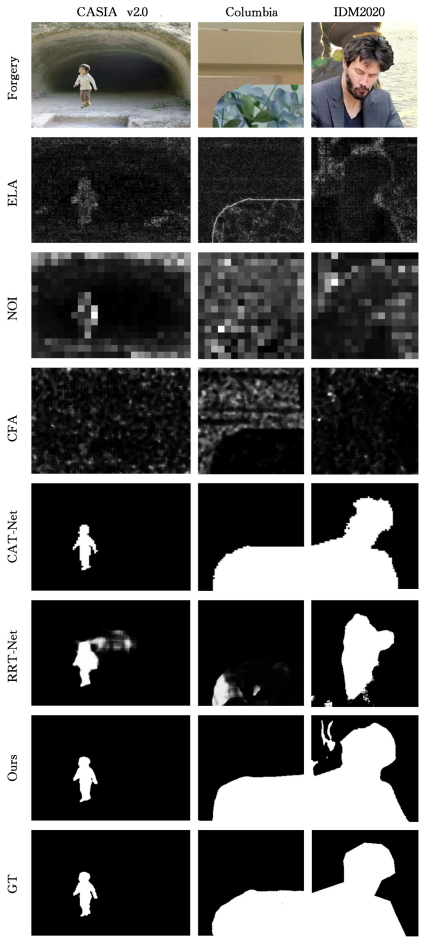

图 7 不同方法检测结果的可视化

Fig. 7 Visualization results of different methods

结束语 针对当前拼接篡改检测算法中存在的特征提取方式感受野有限(多采用卷积操作)和任务模式单一(多为像素级任务)等问题，本文提出了一种基于 Transformer 的多任务拼接篡改检测网络，即 MT-Net。MT-Net 利用 Transformer 中的自注意力机制学习像素间的相关性，为像素提供不同的关注度，使网络更加聚焦拼接痕迹信息。然后，引入自适应融合机制有效避免冗余信息带来的干扰。同时，采用多任务架构从全局和局部对拼接特征进行整体感知和细节学习，与子任务对应的损失函数也能指导网络优化得到较好的检测效果。实验结果表明，在 CASIA V2.0，Columbia 和 IDM2020 3 个典型的公开 ISD 数据集上，本文方法均表现出了良好的检测效果。本文提出的各功能模块均对性能提升有所贡献。在接下来的工作中，将对自注意力机制在篡改检测任务中的作用机理进行进一步研究，在优化网络、提高检测准确性的同时，探寻网络的鲁棒性。

参 考 文 献

[1] LIU Y，WANG H X，CHEN Y，et al. A passive forensic scheme for copy-move forgery based on superpixel segmentation and K-means clustering[J]. Multimedia Tools and Applications，2020，79(1/2)：477-500.

[2] MAHDIAN B，SAIC S. Using noise inconsistencies for blind image forensics[J]. Image and Vision Computing，2009，27(10)：1497-1503.

[3] HOU J U，LEE H K. Detection of Hue modification using photo response nonuniformity[J]. IEEE Transactions on Circuits and Systems for Video Technology，2017，27(8)：1826-1832.

[4] 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 and Security，2012，7(5)：1566-1577.

[5] CHEN C，MCCLOSKEY S，YU J. Image splicing detection via camera response function analysis[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2017：1876-1885.

[6] LIU B，PUN C M. Deep fusion network for splicing forgery localization[C]// European Conference on Computer Vision(ECCV). 2019：237-251.

[7] BI X L，WEI Y，XIAO B，et al. RRU-net：The ringed residual U-net for image splicing forgery detection[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR) Workshops. 2019：30-39.

[8] WU Y，ABDALMAGEED W，NATARAJAN P. Mantra-net：Manipulation tracing network for detection and localization of image forgeries with anomalous features[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2019：9543-9552.

[9] ZHOU J H，NI J Q，RAO Y. Block-based convolutional neural network for image forgery detection[C]// International Workshop on Digital Watermarking(IWDW). 2017：1-10.

[10] BAPPY J H，SIMONS C，NATARAJ L，et al. Hybrid LSTM and encoder-decoder architecture for detection of image forgeries[J]. IEEE Transactions on Image Processing，2019，28(7)：3286-3300.

[11] ZHOU P，HAN X T，MORARIU V I，et al. Learning rich features for image manipulation detection[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2018：1053-1061.

[12] DIRIK A E，MEMON N. Image tamper detection based on demosaicing artifact[C]// 16th IEEE International Conference on Image Processing(ICIP). 2009：1497-1500.

[13] LIN Z C，HE J F，TANG X O，et al. Fast，automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis[J]. Pattern Recognition，2019，42(11)：2492-2501.

[14] KRAWETZ N. A picture's worth：digital image analysis and forensics[EB/OL]. [2021-11-29]. http://hackerfactor.org/papers/bh-usa-07-krawetz-wp.pdf.

[15] RAO Y，NI J Q. 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). 2016：1-6.

[16] FRIDRICH J，KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEETransactions on Information Forensics

and Security, 2012, 7(3): 868-882.
[17] CUN X D, PUN C M. Image splicing localization via semi-global network and fully connected conditional random fields[C]// European Conference on Computer Vision(ECCV). 2019: 252-266.
[18] BAPPY J H, ROY-CHOWDHURY A K, BUNK J, et al. Exploiting Spatial Structure for Localizing Manipulated Image Regions [C]// IEEE International Conference on Computer Vision(ICCV). 2017: 4980-4989.
[19] KWON M J, YU I J, NAM S H, et al. CAT-Net: Compression artifact tracing network for detection and localization of image splicing[C]// IEEE Winter Conference on Applications of Computer Vision(WACV). 2021: 375-384.
[20] SALLOUM R, REN Y Z, JAY K C C. Image splicing localization using a multi-task fully convolutional network(MFCN)[J]. Journal of Visual Communication and Image Representation, 2018(51): 201-209.
[21] KNIAZ V V, KNYAZ V A, REMONDINO F. The point where reality meets fantasy: Mixed adversarial generators for image splice detection[C]// 35th Conference on Neural Information Processing Systems(NeurIPS). 2019: 215-226.
[22] BI X L, ZHANG Z P, LIU Y B, et al. Multi-Task wavelet corrected network for image splicing forgery detection and localization[C]// IEEE International Conference on Multimedia and Expo(ICME). 2021: 1-6.
[23] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.
[24] XIE E, WANG W H, YU Z D, et al. SegFormer: Simple and efficient design for semantic segmentation with Transformers[C]// Neural Information Processing Systems(NeurIPS). 2021: 1-18.
[25] HENDRYCKS D, GIMPEL K. Gaussianerror linear units(GELUs)[J]. arXiv: 1606. 08415, 2016.
[26] WEI J, WANG S H, HUANG Q M. F3Net: Fusion, feedback and focus for salient object detection[C]// 34th AAAI Conference on Artificial Intelligence. 2020: 12321-12328.
[27] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.
[28] RAHMAN M A, WANG Y. Optimizing intersection-over-union in deep neural networks for image segmentation[C]// International Symposium on Visual Computing(ISVC). 2016: 234-244.
[29] DONG J, WANG W, TAN T, CASIA image tampering detection evaluation database[C]// 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]// IEEE International Conference on Multimedia and Expo(ICME). 2006: 549-552.
[31] NOVOZAMSKY A, MAHDIAN B, SAIC S. Imd2020: A large-scale annotated dataset tailored for detecting manipulated images [C]// IEEE Winter Conference on Applications of Computer Vision(WACV) Workshops. 2020: 71-80.
[32] HE K M, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition(CVPR). 2016: 770-778.

ZHANG Jingyuan, born in 1996, postgraduate. Her main research interests include digital image forensics and deep learning.

HE Peisong, born in 1991, Ph. D, associate professor. His main research interests include multimedia security and deep learning.

(责任编辑:杨雪敏)