山西大学学报(自然科学版)49(2):209—219,2026
Journal of Shanxi University(Nat. Sci. Ed.)

DOI:10.13451/j.sxu.ns.2025093

一种基于多尺度内容感知的图像篡改定位方法

张雷[1*],王宝丽[1],陆晓栋[1],闫成梁[2],常敏慧[1]
(1.运城学院 数学与信息技术学院,山西 运城 044000;
2.太原师范学院 计算机科学与技术学院,山西 晋中 030619)

摘 要:随着图像编辑技术的快速发展,图像篡改定位面临日益复杂的挑战。现有方法在捕捉篡改区域的多尺度上下文信息方面表现不足,且在多尺度特征融合过程中采用固定采样核,难以关注特征的局部变化,导致定位精度受限。针对这些问题,本文提出一种基于多尺度内容感知的图像篡改定位方法。首先,设计了基于多尺度内容感知的特征融合模块,能够为特征图中的每个位置动态生成自适应的采样核,使模型在粗粒度上定位篡改区域的大致范围,同时在细粒度上识别篡改边缘特征。其次,采用深度可分离卷积解码器替代传统的多层感知机进行预测,进一步提升检测准确性。最后,提出一种结合二元交叉熵损失和Dice损失的联合损失函数,有效增强了模型的鲁棒性和泛化能力。在多个公开数据集上的跨数据集实验结果表明,提出的方法在CASIAv1、Defacto-12k、Coverage和Columbia数据集上,Pixel-level $F1$ score分别达到了69.4%、25.4%、37.8%和83.2%。相较于主流的MVSS-Net++和最新的IML-ViT,平均提升了10.2%和2.1%,显著提高了图像篡改定位的精度。

关键词:特征融合;深度可分离卷积;Vision Transformer;联合损失
中图分类号:TP391 **文献标志码**:A **文章编号**:0253-2395(2026)02-0209-11

A Multi-scale Content-aware Localization Method for Image Manipulation

ZHANG Lei[1*], WANG Baoli[1], LU Xiaodong[1], YAN Chengliang[2], CHANG Minhui[1]
(1. *School of Mathematics and Information Technology, Yuncheng University, Yuncheng* 044000*, China;*
2. *School of Computer Science and Technology, Taiyuan Normal University, Jinzhong* 030619*, China*)

Abstract: With the rapid advancement of image editing technologies, image manipulation localization is facing increasingly complex challenges. Existing methods exhibit limitations in capturing multi-scale contextual information of manipulated regions, and they often rely on fixed sampling kernels during the multi-scale feature fusion process, which makes it difficult to focus on local variations of features, thereby restricting the localization accuracy. To address these issues, we propose an image manipulation localization method based on multi-scale content-aware. Firstly, a feature fusion module based on multi-scale content-aware is designed, which can dynamically generate adaptive sampling kernels for each position in the feature map, enabling the model to locate the approximate range of manipulated regions at a coarse-grained level and identify manipulated edge features at a fine-grained level. Secondly, a depthwise separable convolutional decoder is employed to replace the traditional multilayer perceptron for prediction, further enhancing detection accuracy. Finally, a joint loss function combining binary cross-entropy loss and Dice loss is proposed, effectively improving the model's robustness and generalization capabilities. Cross-dataset experimental results on multiple public datasets demonstrate that, in terms of Pixel-level $F1$ score, the proposed method achieves 69.4%, 25.4%, 37.8%, and 83.2% on the CASIAv1, Defacto-12k, Coverage, and Columbia datasets, respectively. Compared to the mainstream MVSS-Net++ and the latest IML-ViT, it

收稿日期:2025-07-20;**修回日期**:2025-09-10
基金项目:国家自然科学基金(61703363);山西省基础研究计划项目(202403021221206);数据挖掘与工业智能应用科研创新团队资助项目(YCXYTD-202402);运城学院科研项目(YQ-2020021)
* **通信作者**:张雷(1980—),男,山西临猗人,博士,副教授,研究方向为计算机视觉。E-mail:zhanglei@ycu.edu.cn
引文格式:张雷,王宝丽,陆晓栋,等.一种基于多尺度内容感知的图像篡改定位方法[J].山西大学学报(自然科学版),2026,49(2):209-219. DOI:10.13451/j.sxu.ns.2025093.

achieves average improvements of 10.2% and 2.1%, respectively, significantly enhancing the accuracy of image tampering localization.

Key words: feature fusion; depthwise separable convolution; Vision Transformer; joint loss

0 引言

随着智能移动设备的快速普及,图像编辑变得越来越容易,严重威胁了图像的真实性和可信度[1]。篡改图像可能被用于虚假新闻传播、司法证据伪造、知识产权侵犯等领域,对经济社会发展造成严重影响[2-3]。作为图像取证领域的一个重要分支,图像篡改定位旨在识别和定位图像中被篡改的区域,近来受到越来越多研究者的关注[4-8]。

现有的图像篡改定位方法可以归纳为两类:基于传统特征提取的方法和基于深度学习的方法[2,9-10]。基于传统特征提取的方法主要根据图像的低级特征和统计特性进行篡改定位,主要包括基于压缩痕迹的方法、基于重采样的方法、基于噪声一致性的方法、基于光照和阴影的方法等。其中,基于压缩痕迹的方法通过分析JPEG压缩痕迹、双重压缩痕迹等检测篡改区域。基于重采样的方法通过缩放、旋转等重采样操作计算的周期性特征定位篡改区域。基于噪声一致性的方法通过检测图像中噪声分布的不一致性定位篡改区域。基于光照和阴影的方法通过分析图像中光照和阴影的不一致性识别篡改区域。这类方法的优点是可解释性强,但在复杂场景下的定位精度和鲁棒性有限。

近年来,基于深度学习强大的特征提取能力,越来越多的研究者将其应用于图像取证领域,逐渐形成了基于深度学习的图像篡改定位方法。这类方法通过深度神经网络自动提取图像中的篡改特征,显著提高了篡改定位的精度。基于深度学习的方法可以归纳为两类:基于卷积神经网络的方法[11-13]和基于Transformer的方法[14-15]。基于卷积神经网络的方法利用卷积提取图像的局部特征,并通过上采样或反卷积操作实现像素级定位。基于Transformer的方法通过引入注意力机制、融合不同尺度的特征加强模型对篡改区域的关注[16-17]。

虽然现有的基于深度学习的方法显著提高了篡改定位的精度,但仍然存在一些不足之处。图像的多尺度特征对于篡改区域的定位非常重要,不同尺度的特征之间往往存在一定的互补性和交互性。小尺度/高分辨率特征包含丰富的纹理和高频细节等信息,可以为篡改区域的定位提供精确的线索。而大尺度/低分辨率特征具有更大的感受野,能够捕捉全局上下文信息和高级语义信息,有助于发现篡改区域与周围环境在语义或物理规则上的不一致性。两者结合,模型既能关注到篡改的细微痕迹(小尺度),又能理解这些痕迹在整体场景中是否合理(大尺度)。现有方法往往未能充分利用篡改区域的多尺度上下文信息,没有充分挖掘不同尺度特征之间的交互性,使浅层特征和深层特征融合不够充分。另外,现有方法在进行特征上采样时经常采用反卷积,其在整个图像上使用相同的采样核,无法关注特征的局部变化,导致定位精度受限。

针对以上问题,本文提出一种基于多尺度内容感知的图像篡改定位方法。首先,设计了基于多尺度内容感知的特征融合模块,该模块通过引入内容感知采样机制,为特征图中每个位置动态生成自适应的采样核,使融合特征能够有效捕捉篡改区域在不同尺度下的上下文信息和局部细节特征。具体而言,该模块通过特征金字塔网络(Feature Pyramid Network,FPN)提取不同尺度的特征,再利用内容感知机制将不同尺度的特征进行融合,使模型能够在粗粒度上定位篡改区域的大致范围,同时在细粒度上精确识别篡改边缘和纹理细节,从而提升定位的准确性。其次,使用深度可分离卷积解码器(Depth Separable Convolutional Decoder)代替传统的多层感知机进行预测。通过引入深度卷积和逐点卷积,能够更好地捕捉局部特征的空间相关性,同时保留多层感知机的非线性映射能力,提升检测的准确性。最后,设计了一种结合二元交叉熵损失(Binary Cross-Entropy

Loss,BCE)和Dice损失(Dice Loss)的联合损失函数。BCE损失用于衡量预测结果与真实标签之间的差异,而Dice损失则通过计算预测区域与真实区域的重叠度,进一步优化模型的定位精度。通过联合这两种损失函数,增强模型的鲁棒性和泛化能力。

1 相关工作

图像篡改定位方法旨在检测和定位图像中的篡改区域,以应对日益增长的图像伪造问题。它通过分析图像的数字指纹、统计特征或深度学习特征,能够有效识别图像中存在的异常区域,为司法取证、新闻真实性核查、学术不端检测等领域提供重要的技术支撑。从特征提取的角度出发,现有的基于深度学习的方法可以归纳为2类:基于卷积神经网络的方法和基于Transformer的方法。

1.1 基于卷积神经网络的方法

Zhou等[18]开创性地将深度学习应用于图像篡改检测领域,提出一种双流网络RGB-N,其中RGB流从输入图像中提取特征以发现篡改伪影特征,噪声流通过噪声特征发现真实区域和篡改区域之间的噪声不一致性。Wu等[11]提出一种端到端的图像篡改检测和定位网络ManTra-Net,主要由图像操作痕迹特征提取器和局部异常检测网络构成,该网络引入异常特征学习机制,通过分析图像的局部异常特征来识别篡改区域。Hu等[12]提出了一种空间金字塔注意力网络SPAN,主要由特征提取器模块、金字塔空间注意力传播模块和决策模块构成,通过在不同尺度上提取图像特征,捕捉不同大小的篡改区域,增强网络对多尺度篡改痕迹的感知能力。另外,通过注意力模块增强网络对篡改区域的关注,抑制无关背景信息的干扰,提高定位的准确性。Yang等[13]提出了一种由粗到细的CR-CNN网络,该网络主要由可学习的操纵特征提取器、粗略操纵检测、精细操纵检测构成。可学习的操纵特征提取器直接从数据中学习各种内容操纵的统一的特征表示,粗略操纵检测由注意区域建议网络和预测模块构成,用来进行粗略的操纵区域定位,精细操纵检测进一步细化局部特征,并且在像素级别上执行操作分割。

1.2 基于Transformer的方法

Transformer是一种自然语言处理领域的深度学习模型架构,通过自注意力机制捕捉输入序列中元素之间的全局依赖关系,近年来越来越多的研究者将其应用于图像篡改检测与定位方法中。Hao等[14]提出了一种基于密集自注意力机制的图像篡改定位网络TransForensics,该网络主要由密集自注意力编码器模块和密集校正模块构成。密集自注意力编码器模块建模全局上下文以及不同尺度上局部块之间的所有成对交互,密集校正模块主要修正不同分支的输出。Wang等[15]提出了一种端到端多模态的图像篡改定位网络ObjectFormer,该网络主要由高频特征提取模块、对象编码器和块解码器模块构成,将图像中的对象作为基本处理单元,增强对篡改区域的感知能力。Ma等[16]提出了一种基于Vision Transformer(ViT)的图像篡改定位网络IML-ViT,该网络主要由图像填充、ViT编码器、多尺度特征融合、边缘监督等模块构成,增强了模型对篡改区域的定位精度。Zeng等[17]提出了一种基于查询的Transformer架构MGQFormer,通过掩码引导的方式,实现对图像篡改区域的定位。

2 本文方法

2.1 网络架构

本文方法以Vision Transformer[19]为骨干网络,其整体网络架构由4部分构成:Vision Transformer编码器、多尺度内容感知的特征融合模块、深度可分离卷积解码器、联合损失,具体网络架构如图1所示,其中H和W分别表示图像的高度和宽度。

首先,将待检测图像分割成固定大小的图像块,通过线性投影将每个图像块映射为嵌入向量(Patch Embedded),并为每个图像块进行位置编码(Positional Encoding),将这些嵌入向量输入到由多层Vision Transformer编码器中进行特征提取。然后,提取的特征被输入到多尺度内容感知的特征融合模块,将不同尺度的特征进行融合,充分挖掘不同尺度特征之间的交互性,使融合特征能够有效感知上下文信息和

局部特征变化。随后,将融合特征输入深度可分离卷积解码器,得到预测的篡改区域掩码图。最后,利用联合损失函数对预测掩码进行损失监督,优化预测掩码的质量。

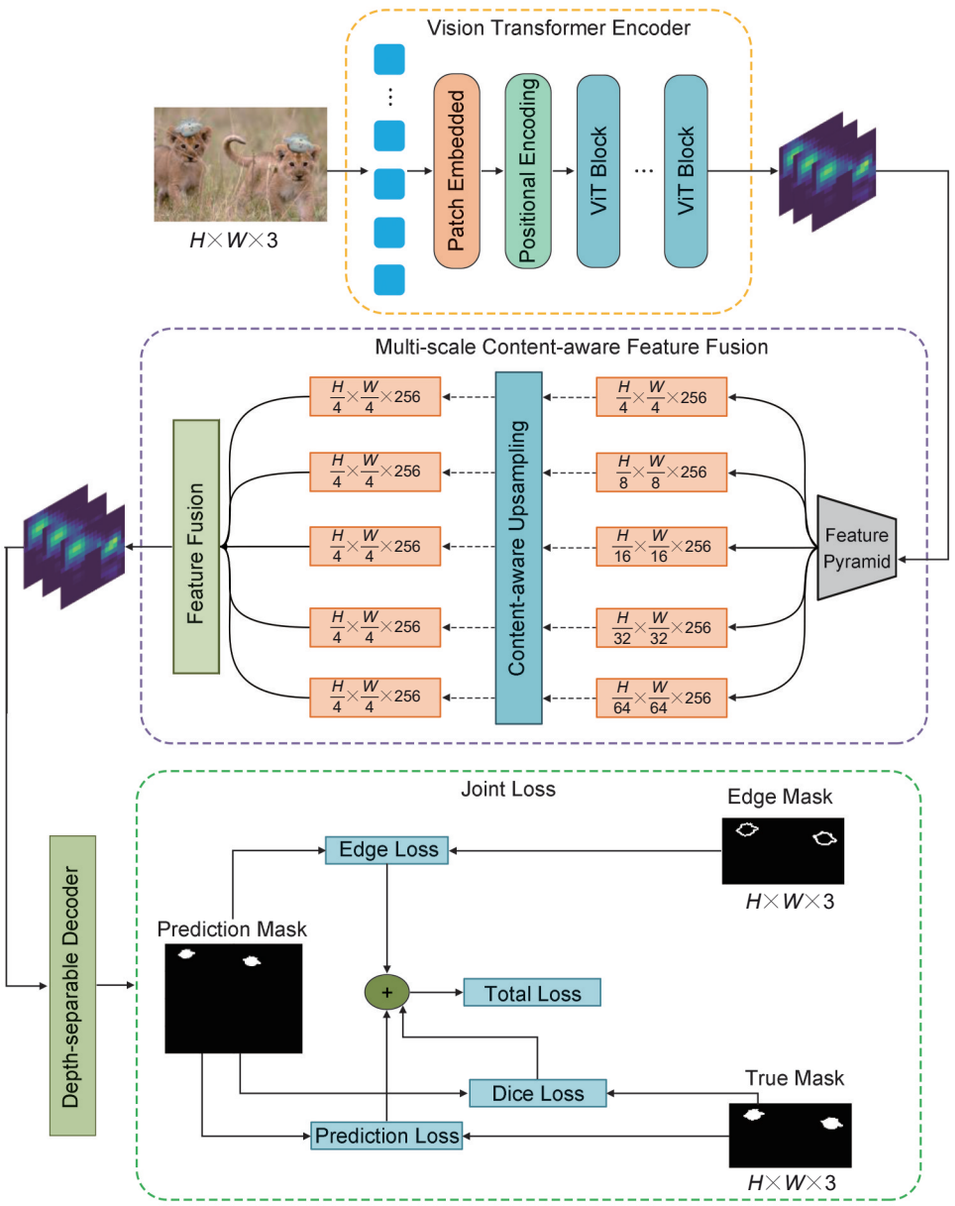

图1 本文方法的整体网络架构

Fig. 1 The overall network architecture of proposed method

2.2 Vision Transformer编码器

在本文中,Vision Transformer 编码器由 12 个相同的编码器层堆叠而成,其结构如图 2 所示[19]。每个编码器层首先对输入的特征进行层归一化(Layer Normalization),并通过多头自注意力机制(Multi-head Self-attention,MSA)计算输入特征序列中所有位置之间的全局依赖关系,生成注意力加权的特征表示,再进行残差连接;然后,对得到的特征进行层归一化,通过多层感知机(Multi-layer Perceptron,MLP)进一步提取非线性特征,最后进行残差连接。经过 12 个编码器层的堆叠,模型能够逐步捕捉从局部到全局的语义信息,最终输出包含全局上下文的高层次特征表示。

2.3 多尺度内容感知的特征融合

由于图像的篡改操作通常会在图像中引入多种尺度的不一致性,且会在篡改区域边界留下人为操作的痕迹,而单一尺度的特征提取

难以同时兼顾全局上下文信息和局部的细微变化。因此,近年来研究者更倾向于利用多尺度特征进行图像篡改定位。

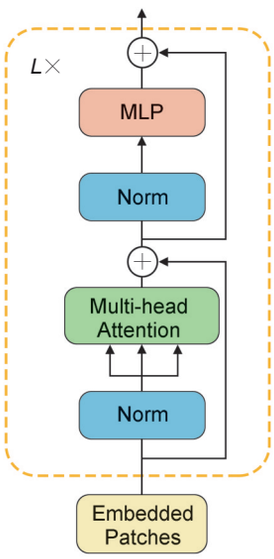

图2 Vision Transformer 编码器

Fig. 2 Vision Transformer encoder

在本文中,为了充分利用不同尺度的特征信息,有效捕捉篡改区域在不同尺度下的上下文信息和局部细节特征,受文献[16]的启发,我们设计了基于多尺度内容感知的特征融合模块。如图1所示,首先将编码器输出的特征通过特征金字塔网络,提取5个不同尺度上的特征$\{F_1^{1/4},F_2^{1/8},F_3^{1/16},F_4^{1/32},F_5^{1/64}\}$。然后,分别对每个尺度上的特征进行内容感知上采样,得到每个尺度上的上采样特征$\{F'^{1/4}_1,F'^{1/4}_2,F'^{1/4}_3,F'^{1/4}_4,F'^{1/4}_5\}$,即将所有尺度上的特征都上采样至原始特征的1/4。最后,将5个尺度上的特征进行通道拼接,再利用卷积进行特征融合。通过该过程,融合特征能够有效感知篡改区域的多尺度上下文信息和局部特征变化,使模型能够在粗粒度上定位篡改区域的大致范围,同时在细粒度上精确识别篡改边缘和纹理细节。其中内容感知上采样的网络架构如图3所示。

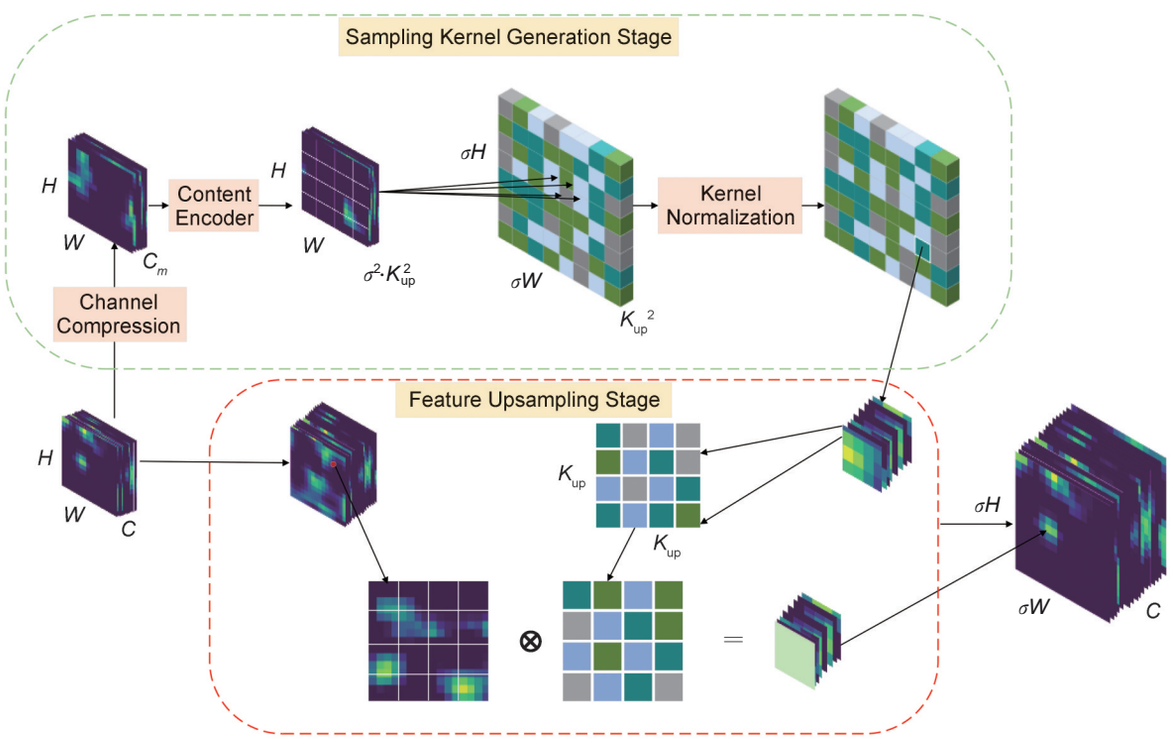

图3 内容感知上采样网络架构

Fig. 3 Upsampling network architecture of content-aware

如图3所示,内容感知上采样过程主要分为两个阶段[20]:采样核生成阶段和特征上采样阶段。在采样核生成阶段,根据特征的具体内容,为目标特征中每个位置动态生成自适应的采样核。具体过程如下:

对于输入的特征图$F\in R^{C\times H\times W}$,首先利用1×1卷积对其进行通道压缩,得到压缩后的特征图$F'\in R^{C_m\times H\times W}$,其中$C_m$为压缩后的通道

数。然后,利用内容编码器对其进行编码,得到$F''\in R^{(\sigma^2\times k_{up}^2)\times H\times W}$,其中内容编码器采用3×3卷积核,$\sigma$为上采样因子,$k_{up}$为上采样核大小。随后将$F''$进行像素重排,得到$F'''\in R^{k_{up}^2\times \sigma H\times \sigma W}$。最后,对$F'''$进行归一化,生成最终的采样核$F_{kernel}$。在$F_{kernel}$中,每个位置上包含$k_{up}^2$个权重,即每个位置上生成了一个$k_{up}\times k_{up}$采样核。

在特征上采样阶段,根据生成的采样核,对输入特征进行特征上采样。具体过程如下:

首先,对于输入的特征图$F\in R^{C\times H\times W}$中的每个位置$l=(i,j)$,提取以其为中心的$k_{up}\times k_{up}$邻域$N_l$。然后,在第一阶段生成的采样核$F_{kernel}$中的相应位置$l'=(i',j')$上,提取出对应的$k_{up}\times k_{up}$采样核$K_{l'}$。将$N_l$和$K_{l'}$对应位置相乘再求和,即得到目标特征图中该位置上的特征值。依次对每个位置进行相同的操作,即可得到输出的特征图$F_{out}\in R^{C\times \sigma H\times \sigma W}$。

2.4 深度可分离卷积解码器

传统多层感知机解码器计算复杂度高,且容易丢失图像的空间结构信息,导致图像篡改定位精度下降。而深度可分离卷积显著减少了计算量和参数量,同时保持了较高的特征提取能力。因此,在本文中我们使用了深度可分离卷积解码器进行篡改区域的预测。深度可分离卷积主要由深度卷积(Depthwise Convolution)和逐点卷积(Pointwise Convolution)两部分构成[21],其主要过程如图4所示。

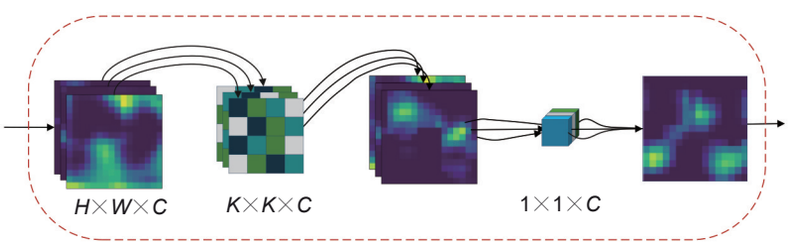

图4 深度可分离卷积解码器

Fig. 4 Depth-separable convolutional decoder

深度卷积对输入特征图的每个通道独立进行卷积操作,用于提取局部空间特征:

$$F_i'=F_i*K_i^{depth},i=1,2,\cdots,C, \tag{1}$$

其中$F\in R^{C\times H\times W}$为输入特征图,$F'\in R^{C\times H'\times W'}$为输出特征图,$K^{depth}$是卷积核。$F_i$为输入特征图的第$i$个通道,$F_i'$为输出特征图的第$i$个通道,$K_i^{depth}$为第$i$个卷积核。$C$个卷积核分别作用于输入特征图的对应通道。

逐点卷积使用1×1卷积核,在通道维度上进行特征组合,用于捕捉通道之间的相关性:

$$F^{out}=F'*K^{point}, \tag{2}$$

其中F^{out}为最终输出的特征图,F'为深度卷积的输出特征图,K^{point}为1×1卷积核。在本文中,K^{point}为1×1×C的卷积核,输出通道为1,用于生成篡改区域的预测掩码。

2.5 联合损失

在图像篡改定位方法中,损失函数的设计与优化对于模型的性能至关重要,直接影响模型的学习方向和收敛速度。在本文中,我们根据真实掩码、预测掩码、真实边缘、预测边缘等设计了联合损失。联合损失由预测损失(Prediction Loss)、Dice损失、边缘损失(Edge Loss)等构成。

预测损失从像素级别上衡量真实掩码和预测掩码之间的不一致性,即预测篡改区域与真实篡改区域之间的差异。由于二元交叉熵损失函数是一种广泛用于二分类任务的损失函数,用于衡量模型预测概率分布与真实标签分布之间的差异,因此,本文使用二元交叉熵损失函数计算预测损失,使模型能够更准确地预测每个像素是否属于篡改区域。其定义为:

$$L_{pred}=L_{BCE}(M_{pred},M_{gt}), \tag{3}$$

其中L_{pred}为预测损失,M_{pred}为预测掩码,M_{gt}为真实掩码。

在图像篡改定位任务中,篡改区域通常只占图像的很小一部分,导致正负样本极度不平衡。因此,本文使用Dice损失通过直接最大化预测区域与真实区域的重叠部分,能够更有效地引导模型学习篡改区域的特征,其定义为:

$$L_{\text{dice}}=1-\frac{2\sum(M_{\text{pred}}\cdot M_{gt})+\varepsilon}{\sum M_{\text{pred}}+\sum M_{gt}+\varepsilon},\tag{4}$$

其中L_{dice}为Dice损失,ε为一个平滑项,用于避免分母为零的情况。

由于篡改区域的边缘通常包含人为操作的痕迹,而这些痕迹信息能够帮助模型提高对篡改区域的识别,提高篡改定位的精度和鲁棒性。因此,本文使用边缘损失衡量真实掩码和预测掩码在篡改边界上的差异,其定义为:

$$L_{\text{edge}}=BCE(M_{\text{pred}},M_{\text{gt}},E_{\text{gt}}),\tag{5}$$

其中L_{edge}为边缘损失,E_{gt}为真实篡改区域边缘。

基于以上3个损失,联合损失定义为:

$$L=L_{\text{pred}}+L_{\text{dice}}^{2}+\lambda\times L_{\text{edge}},\tag{6}$$

其中λ为边缘损失的权重参数,本文设为20。

3 实验结果

3.1 实验设置

(1)数据集:为了验证本文方法的有效性,我们使用与目前主流方法相同的公开数据集进行实验,包括CASIAv1[22]、CASIAv2[22]、Coverage[23]、Columbia[24]、Defacto[25]。其次,为了更加科学准确地比较,我们采用了与文献[16]、[26-27]相同的跨数据集实验方法,在CASIAv2数据集训练模型,在其它数据集上进行测试。另外,由于Defacto数据集缺乏真实图像作为对比样本,我们参考MVSS-Net[26]的方法,从MS-COCO数据集[28]中随机选取了6 000张未修改的图像,并与Defacto数据集中的6 000张图像结合,构建了Defacto-12k数据集用于模型测试。

(2)实验环境:实验在一台配备NVIDIA GeForce RTX 3090 GPU的工作站上进行,采用PyTorch深度学习框架实现。训练过程中,我们使用Adam优化器,初始学习率设置为1×10^{-4},epoch为200。所有输入图像的分辨率均统一调整至256×256,对于较小的图像,采用填充方式扩展至该尺寸,而对于超过此阈值的图像,则将其长边缩放至256,并保持原始纵横比。另外,参考文献[16],我们采用在ImageNet-1k上的掩码自编码器(Masked Auto Encoder, MAE)预训练权重初始化ViT-B。

(3)评价指标:本文主要采用像素级$F1$分数(Pixel-level $F1$ Score)和曲线下面积(Area Under Curve, AUC)作为主要评价指标。在图像篡改定位任务中,它们是评估模型性能的两个核心指标,分别从不同角度衡量模型的定位精度与判别能力。

像素级$F1$分数是准确率(Precision,P)和召回率(Recall,R)的调和平均值,用于评估模型在像素级别对篡改区域的定位精度。其定义为:

$$F1=\frac{2P\times R}{P+R},\tag{7}$$

其中准确率定义为:

$$P=\frac{TP}{TP+FP},\tag{8}$$

召回率定义为:

$$R=\frac{TP}{TP+FN},\tag{9}$$

其中TP为真正例,FP为假正例,FN为假负例。AUC基于接收者操作特征曲线(ROC Curve)计算,用于评估模型在不同分类阈值下的整体判别能力。ROC曲线以真阳性率(True Positive Rate, TPR)为纵轴,假阳性率(False Positive Rate, FPR)为横轴,通过调整分类阈值生成。AUC值越接近1,表明模型在区分篡改区域与真实背景时的性能越好。

3.2 客观评价

为了客观地评估本文方法的有效性,我们进行了跨数据集测试实验。首先在CASIAv2数据集上训练模型,并保存了训练好的模型参数。随后,我们将该模型分别应用于CASIAv1、Defacto-12k、Coverage、Columbia等不同数据集上进行测试。这些数据集涵盖了不同来源、不同篡改类型、不同图像分辨率等特点,能够有效评估模型的泛化能力。我们记录了每种方法在不同数据集上的$F1$值,同时为了更直观地比较不同方法的整体性能,我们还计算了每种方法在四个数据集上的$F1$均值,结果如表1所示。

从表1可以看出,本文方法在CASIAv1、Defacto-12k、Coverage、Columbia等数据集上的$F1$均值为54.0%,相对于主流的MVSS-Net++,提高了10.2个百分点,相较于最新的

表1 不同方法在各数据集上的$F1$值对比

Table 1 Comparison of $F1$ score of different methods

方法	$F1$值/%				
	CASIAv1	Defacto-12k	Coverage	Columbia	均值
HP-FCN[29]	15.4	5.5	0.3	6.7	7.0
CR-CNN[13]	40.5	13.2	29.1	43.6	31.6
ManTra-Net[11]	15.5	15.5	28.6	36.4	24.0
RGB-N[18]	40.8	—	43.7	69.7	—
SPAN[12]	33.6	—	53.5	81.5	—
GSR-Net[27]	38.7	5.1	28.5	61.3	33.4
MVSS-Net[26]	45.2	13.7	45.3	63.8	42.0
MVSS-Net++[30]	51.3	9.5	48.2	66.0	43.8
IML-ViT[16]	65.8	15.6	42.5	83.6	51.9
本文方法	69.4	25.4	37.8	83.2	54.0

注:—表示该模型在原论文中未在该数据集上进行实验。

IML-ViT,提高了2.1个百分点。本文方法在CASIAv1和Defacto-12k两个数据集上获得了最高的$F1$分数。在Columbia数据集上,本文方法仅比IML-ViT低0.4个百分点。值得注意的是,本文方法在Coverage数据集上的$F1$值为37.8%,低于其他某些方法,原因在于Coverage数据集篡改类型相对单一,只包含复制移动篡改,且其中大部分图像的分辨率较低,在一定程度上会影响本文方法的定位精度,这也为我们后续的研究提供了方向。

3.3 主观评价

为了从主观视觉方面进行比较,我们将本文方法与其它方法在CASIAv1数据集上的预测掩码进行比较。通过对比预测掩码图像和真实掩码图像,可以直观地观察到不同方法在定位精度、边界清晰度、误检率和漏检率等方面的差异。部分结果如图5所示。

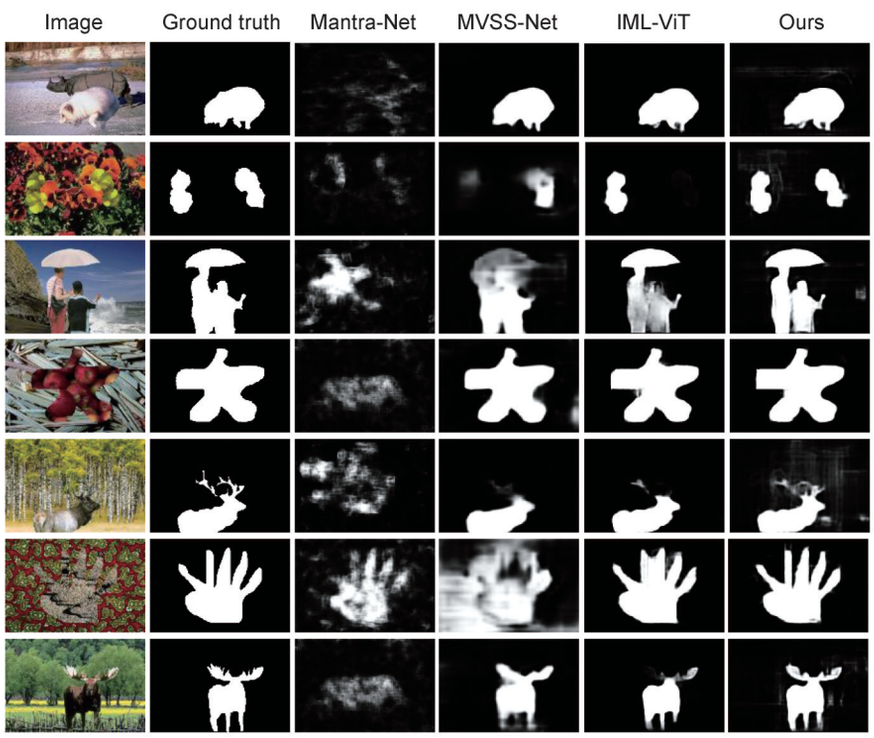

图5 本文方法与其它方法的预测掩码对比

Fig. 5 Comparison of predicted masks between the proposed method and other methods

从图5可以看出,相较于其他方法,本文方法对篡改区域的定位更加精准,预测的篡改区域边界更加清晰,与真实掩码的重合度更高。此外,本文方法在减少误检率方面也表现突出,误将真实区域中的像素识别为篡改区域像素的情况明显减少,这表明本文方法在区分真实区域与篡改区域的能力上更为准确。

3.4 消融实验

为了验证多尺度内容感知的特征融合模块、深度可分离卷积解码器、联合损失等的有效性,我们分别在各数据集上进行消融实验。实验方案设计为:

方案1:同时使用所有模块,验证模块在各数据集上的整体性能;

方案2:移除多尺度内容感知特征融合模块,将编码器输出的特征直接进行解码预测;

方案3:移除深度可分离卷积解码器,用传统的多层感知机解码器进行解码预测;

方案4:移除联合损失中的边缘损失和Dice损失,仅使用预测损失进行损失计算;

方案5:同时移除所有模块,验证基线模型在各数据集上的性能。

消融实验的结果如表2所示。

从表2可以看出,同时使用所有模块,模型在所有数据集上的$F1$均值为0.540,AUC均值为0.873。移除或替换掉其中一个模块,$F1$均值和AUC均值均有所下降,而同时移除所有模块,$F1$均值和AUC均值最低,验证了本文提出的各模块的有效性。其中,在方案3中,移除深度可分离卷积解码器,用传统的多层感知机解码器进行解码预测时,$F1$均值和AUC均值下降较多,分别下降了0.170和0.066,说明深度可分离卷积解码器能够很好地捕捉局部特征的空间相关性,同时保留多层感知机的非线性映射能力,提升了定位的准确性。

3.5 鲁棒性评估

在本文中,我们在CASIAv1数据集上进行鲁棒性实验,采用了图像缩放(Resize)、高斯模糊(Gaussian Blur)、JPEG压缩(JPEG Compression)、对比度调整(Contrast)等4种方式对图像进行失真处理,实验结果如表3所示。

从表3可以看出,本文方法在4种不同的干扰攻击中均表现出较强的鲁棒性。对于所有失真类型,本文方法得到的$F1$均值为0.661,均高于SPAN、MVSS-Net、IML-ViT的$F1$均值0.341、0.437、0.605。本文方法得到的AUC均值为0.870,均高于SPAN、MVSS-Net、IML-ViT的AUC均值。

表2 消融实验结果

Table 2 Results of ablation experiments

方案	CASIAv1		Defacto-12k		Coverage		Columbia		均值	
	$F1$	AUC	$F1$	AUC	$F1$	AUC	$F1$	AUC	$F1$	AUC
1	0.694	0.925	0.254	0.803	0.378	0.856	0.832	0.909	0.540	0.873
2	0.631	0.907	0.234	0.750	0.347	0.834	0.733	0.899	0.486	0.848
3	0.211	0.778	0.213	0.777	0.403	0.816	0.641	0.855	0.370	0.807
4	0.543	0.878	0.213	0.729	0.438	0.820	0.667	0.831	0.465	0.815
5	0.188	0.687	0.152	0.709	0.246	0.719	0.276	0.748	0.216	0.716

表3 不同方法的鲁棒性实验结果

Table 3 Experiment results of robustness of different methods

干扰类型	参数设置	SPAN		MVSS-Net		IML-ViT		本文方法	
		$F1$	AUC	$F1$	AUC	$F1$	AUC	$F1$	AUC
无干扰		0.336	0.800	0.452	0.839	0.658	0.900	0.694	0.925
缩放	缩放因子=2	0.350	0.832	0.452	0.842	0.624	0.845	0.682	0.884
	缩放因子=0.1	0.337	0.803	0.452	0.824	0.631	0.872	0.692	0.892
压缩	质量指数=50	0.338	0.807	0.407	0.765	0.483	0.753	0.534	0.809
	质量指数=100	0.350	0.836	0.445	0.780	0.601	0.868	0.670	0.866
高斯模糊	核大小=3	0.349	0.831	0.436	0.795	0.624	0.890	0.686	0.879
	核大小=15	0.332	0.792	0.385	0.752	0.572	0.854	0.626	0.850
对比度	调整系数=0.3	0.338	0.805	0.452	0.822	0.633	0.875	0.674	0.857
	调整系数=0.7	0.341	0.812	0.452	0.831	0.612	0.854	0.687	0.872
均值		0.341	0.813	0.437	0.806	0.605	0.857	0.661	0.870

4 结语

本文提出一种基于多尺度内容感知的图像篡改定位方法。首先,设计了基于多尺度内容感知的特征融合模块,为不同尺度的特征动态生成自适应的采样核,使模型能够在粗粒度上定位篡改区域的大致范围,同时在细粒度上精确识别篡改边缘和纹理细节。其次,使用深度

可分离卷积解码器代替传统的多层感知机进行预测,更好地捕捉局部特征的空间相关性,同时保留多层感知机的非线性映射能力。最后,设计一种结合二元交叉熵损失和Dice损失的联合损失函数,更准确地计算预测区域与真实区域之间的差异,增强模型的鲁棒性和泛化能力。但是对于篡改类型单一、图像分辨率较低的数据集,本文方法的定位准确率还需进一步提升。在今后的研究中,尝试在特征融合过程中引入图像的频率信息,进一步提升模型的定位精度。

参考文献:

[1] JIN X, YU W, SHI W. Image Manipulation Localization *via* Dynamic Cross-modality Fusion and Progressive Integration[J]. *Neurocomputing*, 2024, **610**: 128607. DOI: 10.1016/j.neucom.2024.128607.

[2] 魏华建, 严彩萍, 李红. 基于集成多尺度注意力的图像篡改定位[J]. 计算机辅助设计与图形学学报, 2024, **36**(8): 1237-1245. DOI: 10.3724/SP.J.1089.2024.19954.
WEI H J, YAN C P, LI H. Image Tampering Localization Based on Integrated Multiscale Attention[J]. *J Comput Aided Des Comput Graph*, 2024, **36**(8): 1237-1245. DOI: 10.3724/SP.J.1089.2024.19954.

[3] 陈海鹏, 刘宏昕, 潘大力, 等. 基于边界不确定性学习的图像篡改定位方法[J/OL]. 吉林大学学报(工学版), 2025: 1-10. (2025-03-06) [2025-03-14]. https://link.cnki.net/doi/10.13229/j.cnki.jdxbgxb.20250014.
CHEN H P, LIU H X, PAN D L, *et al*. Image Manipulation Localization Method Based on Boundary Uncertainty Learning[J/OL]. *J Jilin Univ Eng Technol Ed*, 2025: 1-10. (2025-03-06)[2025-03-14]. https://link.cnki.net/doi/10.13229/j.cnki.jdxbgxb.20250014.

[4] BAI R Y. Image Manipulation Detection and Localization Using Multi-scale Contrastive Learning[J]. *Appl Soft Comput*, 2024, **163**: 111914. DOI: 10.1016/j.asoc.2024.111914.

[5] QU C F, ZHONG Y W, LIU C Y, *et al*. Towards Modern Image Manipulation Localization: A Large-scale Dataset and Novel Methods[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2024: 10781-10790. DOI: 10.1109/CVPR52733.2024.01025.

[6] GUO K, CAO G, LOU Z J, *et al*. A Lightweight and Effective Image Tampering Localization Network with Vision Mamba[J]. *IEEE Signal Process Lett*, 2025, **32**: 2179-2183. DOI: 10.1109/LSP.2025.3570240.

[7] PENG R X, TAN S Q, MO X B, *et al*. Active Adversarial Noise Suppression for Image Forgery Localization[EB/OL]. (2025-06-15) [2025-07-20]. https://arxiv. org/abs/2506.12871.

[8] CHEN Y, CHENG H, WANG H C, *et al*. EAN: Edge-aware Network for Image Manipulation Localization[J]. *IEEE Trans Circuits Syst Video Technol*, 2025, **35**(2): 1591-1601. DOI: 10.1109/TCSVT.2024.3473933.

[9] 姜燕燕. 基于边界引导的图像篡改定位算法研究[D]. 长春: 吉林大学, 2024. DOI: 10.27162/d. cnki. gjlin.2024.001758.
JIANG Y Y. Research on Image Tampering Localization Algorithm Based on Boundary Guidance[D] Changchun: Jilin University, 2024. DOI: 10.27162/d. cnki. gjlin.2024.001758.

[10] 蔺聪, 黄轲, 温雅敏, 等. 匹配对聚类的图像复制粘贴篡改检测[J]. 中国图象图形学报, 2024, **29**(12): 3595-3611. DOI:10. 11834/jig. 230454.
LIN C, HUANG K, WEN Y M, *et al*. Image Copy-move Forgery Detection Based on the Clustering of Matched Pairs[J]. *J Image Graph*, 2024, **29**(12): 3595-3611. DOI:10. 11834/jig. 230454.

[11] WU Y, ABDALMAGEED W, NATARAJAN P. ManTra-Net: Manipulation Tracing Network for Detection and Localization of Image Forgeries with Anomalous Features [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 9535-9544. DOI: 10.1109/CVPR.2019.00977.

[12] HU X F, ZHANG Z H, JIANG Z Y, *et al*. SPAN: Spatial Pyramid Attention Network for Image Manipulation Localization[M]//Computer Vision-ECCV 2020. Cham: Springer International Publishing, 2020: 312-328. DOI: 10.1007/978-3-030-58589-1_19.

[13] YANG C, LI H Z, LIN F T, *et al*. Constrained R-CNN: A General Image Manipulation Detection Model[C]// 2020 IEEE International Conference on Multimedia and Expo (ICME). New York: IEEE, 2020: 1-6.. DOI: 10.1109/icme46284.2020.9102825.

[14] HAO J, ZHANG Z X, YANG S C, *et al*. TransForensics: Image Forgery Localization with Dense Self-attention[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2021: 15035-15044. DOI: 10.1109/ICCV48922.2021.01478.

[15] WANG J K, WU Z X, CHEN J J, *et al*. ObjectFormer for Image Manipulation Detection and Localization[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2022: 2354-2363. DOI: 10.1109/CVPR52688.2022.00240.

[16] MA X C, DU B, JIANG Z H, *et al*. IML-ViT: Bench-

marking Image Manipulation Localization by Vision Transformer[C]// Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver: AAAI Press, 2024: 1−16. DOI:10.48550/arXiv.2307.14863.

[17] ZENG K L, CHENG R, TAN W M, *et al*. MGQFormer: Mask-Guided Query-based Transformer for Image Manipulation Localization[C]// Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver: AAAI Press, 2024, **38**(7): 6944−6952. DOI: 10.1609/aaai.v38i7.28520.

[18] ZHOU P, HAN X T, MORARIU V I, *et al*. Learning Rich Features for Image Manipulation Detection[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 1053−1061. DOI: 10.1109/CVPR.2018.00116.

[19] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. An Image is Worth 16X16 Words: Transformers for Image Recognition at Scale[EB/OL]. (2021−06−03) [2025−03−14]. https://arXiv. org/abs/2010.11929. DOI: 10.48550/arXiv.2010.11929.

[20] WANG J Q, CHEN K, XU R, *et al*. CARAFE: Content-aware ReAssembly of FEatures[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 3007−3016. DOI: 10.1109/ICCV.2019.00310.

[21] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017: 1800−1807. DOI: 10.1109/CVPR.2017.195.

[22] DONG J, WANG W, TAN T N. CASIA Image Tampering Detection Evaluation Database[C]//2013 IEEE China Summit and International Conference on Signal and Information Processing. New York: IEEE, 2013: 422−426. DOI: 10.1109/ChinaSIP.2013.6625374.

[23] WEN B H, ZHU Y, SUBRAMANIAN R, *et al*. COVERAGE: A Novel Database for Copy-move Forgery Detection[C]//2016 IEEE International Conference on Image Processing (ICIP). New York: IEEE, 2016: 161−165. DOI: 10.1109/ICIP.2016.7532339.

[24] HSU Y F, CHANG S F. Detecting Image Splicing Using Geometry Invariants and Camera Characteristics Consistency[C]//2006 IEEE International Conference on Multimedia and Expo. New York: IEEE, 2006: 549−552. DOI: 10.1109/ICME.2006.262447.

[25] MAHFOUDI G, TAJINI B, RETRAINT F, *et al*. DEFACTO: Image and Face Manipulation Dataset[C]// 2019 27th European Signal Processing Conference (EUSIPCO). New York: IEEE, 2019: 1−5.

[26] CHEN X R, DONG C B, JI J Q, *et al*. Image Manipulation Detection by Multi-view Multi-scale Supervision [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2021: 14165−14173. DOI: 10.1109/ICCV48922.2021.01392.

[27] ZHOU P, CHEN B C, HAN X T, *et al*. Generate, Segment and Refine: Towards Generic Manipulation Segmentation [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2020, **34**(7):13058−13065. DOI: 10.48550/arXiv.1811.09729.

[28] LIN T Y, MAIRE M, BELONGIE S, *et al*. Microsoft COCO: Common Objects in Context[M]. Heidelberg: Springer International Publishing, 2014, **8693**: 740−755.

[29] DONG C B, CHEN X R, HU R H, *et al*. MVSS-net: Multi-view Multi-scale Supervised Networks for Image Manipulation Detection[J]. *IEEE Trans Pattern Anal Mach Intell*, 2023, **45**(3): 3539−3553. DOI: 10.1109/tpami.2022.3180556.

[30] LI H D, HUANG J W. Localization of Deep Inpainting Using High-pass Fully Convolutional Network[C]// 2019 IEEE/CVF International Conference on Computer Vision (ICCV). New York: IEEE, 2019: 8300−8309. DOI: 10.1109/ICCV.2019.00839.