智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

尺度可变有约束图像拼接检测与定位及其对抗优化

刘亚奇, 蔡强, 石磊, 张一凡, 吕斌斌, 夏超, 许盛伟

引用本文:
刘亚奇, 蔡强, 石磊, 等. 尺度可变有约束图像拼接检测与定位及其对抗优化[J]. 智能系统学报, 2024, 19(6): 1479-1491.
LIU Yaqi, CAI Qiang, SHI Lei, et al. Scalable constrained image splicing detection and localization with adversarial optimizing[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(6): 1479-1491.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202307011

DOI: 10.11992/tis.202307011
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20240911.1952.003

尺度可变有约束图像拼接检测与定位及其对抗优化

刘亚奇[1], 蔡强[2], 石磊[3], 张一凡[1], 吕斌斌[1], 夏超[1], 许盛伟[1]
（1. 北京电子科技学院, 北京 100070; 2. 北京工商大学 食品安全大数据技术北京市重点实验室, 北京 100048; 3. 中国传媒大学 媒体融合与传播国家重点实验室, 北京 100024）

摘　要: 针对有约束图像拼接检测与定位这一图像取证任务, 提出一种尺度可变的检测与定位方法及其对抗优化架构。在所提有约束图像拼接检测与定位网络中, 设计基于高效通道注意力增强的尺度可变关联性计算模型, 采用高效通道注意力增强对通道特征进行校准, 利用强关联因子截断操作实现处理任意尺度大小的图像; 设计一种空间注意力增强的空洞空间金字塔池化对多尺度信息进行挖掘, 以及基于深度可分离卷积和残差结构的定位结果重构网络。此外, 提出一种块级对抗学习机制对预训练的尺度可变有约束图像拼接检测与定位网络进行优化。在公开数据集上的大量实验证明所提方法的有效性。

关键词: 有约束图像拼接检测与定位; 尺度可变; 关联性计算; 对抗学习; 图像取证; 空洞卷积; 金字塔池化; 深度可分离卷积
中图分类号: TP391.4　　**文献标志码:** A　　**文章编号:** 1673−4785(2024)06−1479−13

中文引用格式: 刘亚奇, 蔡强, 石磊, 等. 尺度可变有约束图像拼接检测与定位及其对抗优化 [J]. 智能系统学报, 2024, 19(6): 1479–1491.
英文引用格式: LIU Yaqi, CAI Qiang, SHI Lei, et al. Scalable constrained image splicing detection and localization with adversarial optimizing[J]. CAAI transactions on intelligent systems, 2024, 19(6): 1479–1491.

Scalable constrained image splicing detection and localization with adversarial optimizing

LIU Yaqi[1], CAI Qiang[2], SHI Lei[3], ZHANG Yifan[1], LYU Binbin[1], XIA Chao[1], XU Shengwei[1]
(1. Beijing Electronic Science and Technology Institute, Beijing 100070, China; 2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China; 3. State Key Laboratory of Media Convergence and Communiaction, Communication University of China, Beijing 100024, China)

Abstract: A scalable detection and localization method, along with its adversarial optimization architecture, is proposed for the image forensics task of constrained image splicing detection and localization (CISDL). In the CISDL network, a novel scalable correlation computation module is constructed using high-efficiency channel attention enhancement blocks. The channel features are then calibrated using high-efficiency channel attention enhancement. Images of arbitrary sizes are processed using truncation operations on closely correlated factors, and a mask reconstruction network is designed based on depthwise separable convolution and residual connections. Finally, a patch-level adversarial learning strategy is proposed to optimize the pretrained model. Extensive experiments on publicly available datasets demonstrate the effectiveness of the proposed method

Keywords: constrained image splicing detection and localization; scalable; correlation computation; adversarial learning; image forensics ; atrous convolution; pyramid pooling; depthwise separable convolution

随着数字图像编辑技术的进步, 对数字图像进行编辑越来越方便, 效果也越来越逼真[1], 而编辑后的图像如果被恶意篡改者所利用, 用于假消息的传播和作伪证, 会给网络空间安全、司法公

收稿日期: 2023−07−12.　**网络出版日期:** 2024−09−12.
基金项目: 中央高校基本科研业务费资金项目(3282023016); 国家自然科学基金项目(62102010, 62002003); 北京工商大学食品安全大数据技术北京市重点实验室开放课题(BTBD-2022KF02).
通信作者: 蔡强. E-mail: caiq@btbu.edu.cn.

正等带来挑战[2-3]。数字图像取证技术通过挖掘图像的不一致性判定图像是否经过了人为篡改，受到了学术界和产业界越来越多的关注[4-5]。经典数字图像取证通过挖掘单幅待检测图像的低阶统计不一致[6-7]或高阶语义不一致[8]，对图像或图像中区域的真伪进行判定，存在抗压缩、抗噪声能力差和运算效率低等问题[9]。由于异常信息仅从单幅图像获得，且十分微弱难以被捕捉，经典的数字图像取证问题本身就十分棘手[10-11]。

同时，用户进行数字图像取证分析，不仅关注一幅图像是否被篡改，而且关注该图像由谁篡改、如何篡改、又进一步合成了哪些篡改图像[12]。有约束图像拼接检测与定位可以对输入的 2 幅待分析图像进行像素级别的比较，给出篡改区域和篡改来源区域的定位结果[13-14]，是进行篡改来源图像溯源、图像来源分析的重要技术基础和中间步骤[15]。深度匹配与验证网络（deep matching and validation network, DMVN）是首个专门针对该任务设计的端到端深度学习网络，可以给出 2 个输入图像的关联度以及篡改和篡改来源区域，但是该网络对小的篡改区域检测和定位能力不强，对物体轮廓定位不准确，鲁棒性不强[12]。Ye 等[16]提出改进特征金字塔深度匹配与定位网络（feature pyramid deep matching and localization network, FPLN），使用特征金字塔结构，但效果提升还不够明显。Liu 等[13]引入空洞卷积操作对深度匹配特征的空间信息进行重构、利用跳跃结构对层次化特征进行挖掘、采用空洞空间金字塔池化挖掘信息，提出了 DMAC 网络，并设计了一种混合对抗学习策略，相对 DMVN 取得了显著进步。AttentionDM 是在 DMAC 基础上引入编码器解码器结构、注意力模块、标准化模块，效果进一步提升[14]。然而，DMVN、FPLN、DMAC 和 AttentionDM 均只能处理固定尺寸大小的图像，对高分辨率图像的处理能力不足，对于图像中较小或是经过复杂变换的篡改区域检测能力较弱。

为了解决上述问题，本研究提出一套尺度可变的有约束图像拼接检测与定位网络及其块级对抗优化机制，主要分为以下 3 个部分。

1) 提出一套尺度可变的有约束图像拼接检测与定位及其对抗学习优化架构。尺度可变主要是指所提网络架构可以处理任意尺度的图像。

2) 设计尺度可变的有约束图像拼接检测与定位网络，主要包括设计实现了结合标准化操作和高效通道注意力增强的尺度可变的关联性计算模型、空间注意力增强空洞空间金字塔池化模块以及基于深度可分离卷积和残差结构的定位结果重构网络。

3) 设计块级对抗学习优化网络及块级对抗优化策略对定位结果进行优化。块级对抗学习优化网络可以驱使尺度可变的有约束图像拼接检测与定位网络生成的定位结果，与真实定位结果在特征分布上更加接近。在对抗环境下，生成判别器更难区分的定位结果。

1　相关工作

有约束图像拼接检测与定位对输入的可疑篡改图像及其篡改来源图像进行像素级比较，分析 2 幅图像区域之间的关联性，从而找到可疑篡改图像中是否有拼接区域来自于相应的可疑篡改来源图像，并且给出篡改区域和篡改来源区域的具体位置和轮廓。相比传统图像取证方法，有约束图像拼接检测与定位不仅能够找到篡改区域，同时能够找到篡改来源区域，具有更强的可解释性和说服力。同时有约束图像拼接检测与定位能够利用图像的视觉相似性特征，相比传统方法的不一致性特征更容易提取、捕捉。有约束图像拼接检测与定位的应用需要具备可溯源的图像数据集，对图像真实性和拼接图像的派生关系进行系统分析，在其中起到“承上启下”的技术支撑作用。

图像来源分析[15]主要可以分为 2 个步骤：来源图像过滤、来源图像构建。来源图像过滤的目标是在图像数据集上检索得到一组与待分析图像存在关联的图像，然后来源图构建进一步分析这些图像与待分析图像之间的合成与被合成的关系，构建一个来源图。来源图像过滤与图像检索技术[17]类似，不同之处在于来源图像过滤明确需要检索到篡改来源图像，检索精度要求高。目前来源图像过滤的研究仍然处于“实验室”阶段，主要以关键点为单位进行检索[18]，相比实际应用场景仍有距离。在来源图构建方面[15]，构建来源图需要进行来源图像过滤，其过滤过程既要确保较高的召回率，又要有一定的准确率。由此，可以先进行具有较高召回率的来源图像过滤，获得一批可疑篡改来源图像。然后再进行有约束图像拼接检测与定位，进行更进一步的筛选，提高候选图像的精度。从而进行后续的来源图构建，分析图像之间的派生关系。

以卷积神经网络为主体构建有约束图像拼接检测与定位网络是该方向的主要技术思路，现有方法主要有 DMVN[12]、DMAC[13]、AttentionDM[14]、FPLN[16]。有约束图像拼接检测与定位需要进行

快速、大量的像素级匹配，传统基于关键点特征或者密集块特征的方法很难在效率上满足任务需求，而且传统方法在算法鲁棒性方面也较差[19]。为此，引入基于卷积神经网络的深度学习网络[20]，这些方法在效率方面具有明显优势，在鲁棒性、召回率方面也较高。但是当图像中篡改区域较小时，检测定位能力、对高分辨率图像的检测能力以及对强变换区域的匹配能力均有待提高。

2　研究方法

本研究所提方法流程如图 1 所示。网络采用空洞卷积改造 16 层的牛津大学视觉几何组网络（visual geometry group 16，VGG16）进行特征提取，获得 3 组包含丰富层次化信息的特征映射。设计尺度可变的关联性计算模型，将 3 组特征映射分别进行标准化操作和高效通道注意力增强，对注意力增强后的特征映射进行逐像素关联性计算，并对强关联因子进行截断操作，最后对 3 组截断后的关联映射进行合并和标准化处理，得到最终的关联映射，从而使得所提尺度可变的有约束图像拼接检测与定位网络可以处理任意尺度图像。设计空间注意力增强的空洞空间金字塔池化操作，对关联映射的多尺度信息进行挖掘；设计一种结合深度可分离卷积和残差结构的网络对定位结果进行重构。为了进一步优化定位结果的分布，本研究设计了块级对抗学习优化网络，网络可以处理任意尺度的输入，并对块级别的定位准确程度进行对抗优化，使其优化后的尺度可变得有约束图像拼接检测与定位网络的定位结果更接近实际值。

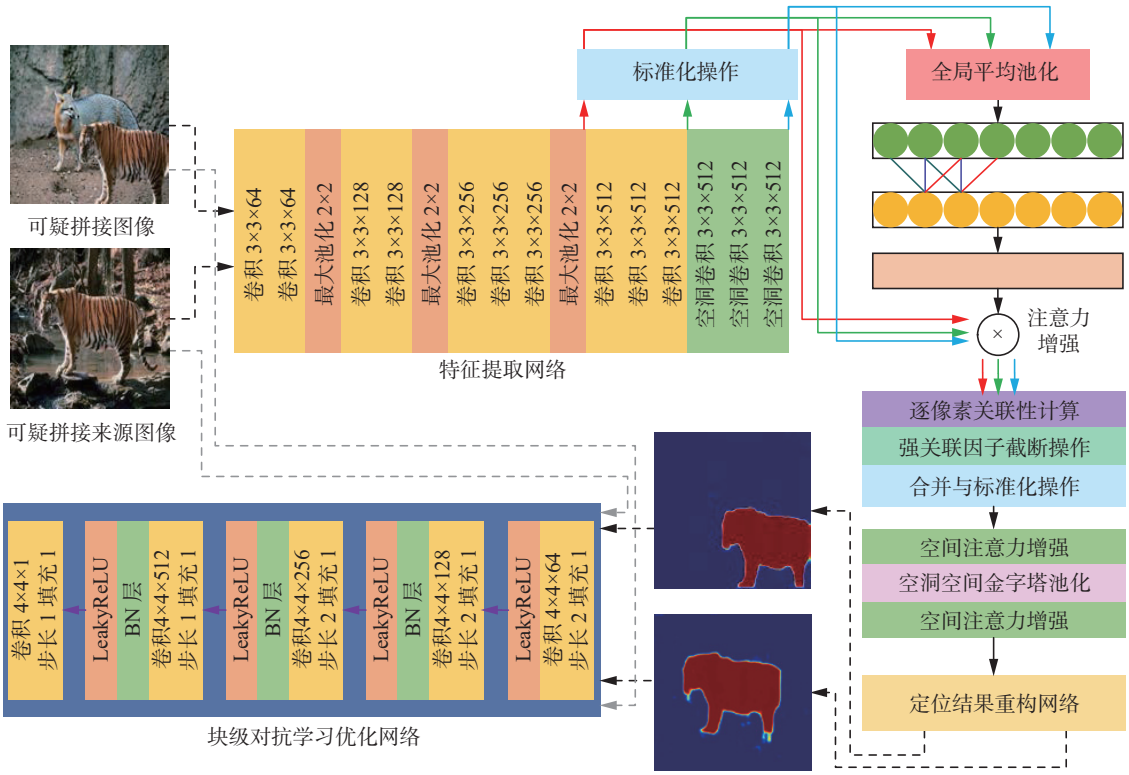

图 1　尺度可变的有约束图像拼接检测与定位及其块级对抗优化机制算法流程

Fig. 1　Schematic diagram of scalable constrained image splicing detection and localization with adversarial optimizing

尺度可变有约束图像拼接检测与定位及其块级对抗优化方法主要包含 2 个部分：尺度可变的有约束图像拼接检测与定位网络和块级对抗学习优化网络。在尺度可变有约束图像拼接检测与定位网络中，主要包含 4 个部分：特征提取网络、尺度可变的关联性计算模型、空间注意力增强空洞空间金字塔池化模块、定位结果重构网络。

2.1　层次化特征提取与注意力增强特征重构

特征提取网络采用 VGG16 作为主干网络，并利用空洞卷积对最后一组卷积模块进行适配，去掉原 VGG16 中的 2 组最大池化操作，获得更大的特征映射，具体结构见图 1 中“特征提取网络”模块，可以获得 3 组尺度较大且尺度相同的特征映射：

$$\left\{\boldsymbol{F}_3^{(1)},\boldsymbol{F}_4^{(1)},\boldsymbol{F}_5^{(1)}\right\},\left\{\boldsymbol{F}_3^{(2)},\boldsymbol{F}_4^{(2)},\boldsymbol{F}_5^{(2)}\right\}=f_{\mathrm{VGG16\_AC}}\left(\boldsymbol{I}^{(1)},\boldsymbol{I}^{(2)}\right) \quad (1)$$

式中：$f_{\mathrm{VGG16_AC}}()$表示经过空洞卷积适配后的 VGG16 网络[13]；输入为可疑拼接图像和可疑拼接来源图像 $\boldsymbol{I}^{(1)}$ 和 $\boldsymbol{I}^{(2)}$，空间尺度大小为 $W\times H$（W 代表输入

图像的宽度，H 代表输入图像的高度）；输出为 3 层特征映射 $\{\boldsymbol{F}_3^{(1)},\boldsymbol{F}_4^{(1)},\boldsymbol{F}_5^{(1)}\}$ 和 $\{\boldsymbol{F}_3^{(2)},\boldsymbol{F}_4^{(2)},\boldsymbol{F}_5^{(2)}\}$，其空间尺度大小均为 $\frac{W}{8}\times\frac{H}{8}$，具体来讲特征映射 $\boldsymbol{F}_3^{(1)}$ 和 $\boldsymbol{F}_3^{(2)}$ 尺寸大小为 $256\times\frac{W}{8}\times\frac{H}{8}$，$\boldsymbol{F}_4^{(1)}$ 和 $\boldsymbol{F}_4^{(2)}$ 为 $512\times\frac{W}{8}\times\frac{H}{8}$，$\boldsymbol{F}_5^{(1)}$ 和 $\boldsymbol{F}_5^{(2)}$ 为 $512\times\frac{W}{8}\times\frac{H}{8}$。

得到 3 层特征映射后，将其输入到尺度可变关联性计算模型中进行关联映射 $(\boldsymbol{C}^{(1)},\boldsymbol{C}^{(2)})$ 的计算为

$$\boldsymbol{C}^{(1)},\boldsymbol{C}^{(2)}=f_{\mathrm{CORR}}\left(\left\{\boldsymbol{F}_3^{(1)},\boldsymbol{F}_4^{(1)},\boldsymbol{F}_5^{(1)}\right\},\left\{\boldsymbol{F}_3^{(2)},\boldsymbol{F}_4^{(2)},\boldsymbol{F}_5^{(2)}\right\}\right) \tag{2}$$

式中：$f_{\mathrm{CORR}}()$ 表示尺度可变关联映射计算，输出为计算得到的关联映射 $\boldsymbol{C}^{(1)}$、$\boldsymbol{C}^{(2)}$，其具体大小为 $c\times w\times h$（c 为通道数，w 为关联映射的宽度，h 为关联映射的高度），$w=\frac{W}{8}$，$h=\frac{H}{8}$。尺度可变关联性计算模型 $f_{\mathrm{CORR}}()$ 的具体计算过程在下面进行介绍。

在得到关联映射 $\boldsymbol{C}^{(1)}$、$\boldsymbol{C}^{(2)}$ 后，下一步需要充分挖掘关联映射所提供的关联信息，从而重构定位结果。为了对关联映射 $\boldsymbol{C}^{(1)}$、$\boldsymbol{C}^{(2)}$ 特征之间的相关性进一步挖掘，这里采用空间注意力机制进行注意力增强[21]，关联映射中长距离的近似关联特征 $\boldsymbol{A}^{(k)}$ 为

$$\boldsymbol{A}^{(k)}=\lambda\sum_m\sum_n\beta^{(k)}(m,n)h\left(\boldsymbol{C}^{(k)}(n)\right)+\boldsymbol{C}^{(k)} \tag{3}$$

式中：λ 表示调节参数，初始化为 0，并参与后续训练调整为一个合适的值；$k\in\{1,2\}$；$h()$ 表示线性映射，$h(\boldsymbol{C}^{(k)}(m))=\boldsymbol{C}^{(k)}(m)\boldsymbol{W}_h+\boldsymbol{b}_h$；$\boldsymbol{C}^{(k)}(n)$ 为 c 维的向量，$\boldsymbol{W}_h\in\boldsymbol{R}^{c\times\frac{c}{8}}$，$\boldsymbol{b}_h\in\boldsymbol{R}^{\frac{c}{8}}$，$\beta^{(k)}(m,n)$ 计算为

$$\beta^{(k)}(m,n)=\frac{\exp\left(f(\boldsymbol{C}^{(k)}(m))^{\mathrm{T}}g(\boldsymbol{C}^{(k)}(n))\right)}{\sum\limits_n\exp\left(f\left(\boldsymbol{C}^{(k)}(m)\right)^{\mathrm{T}}g\left(\boldsymbol{C}^{(k)}(n)\right)\right)} \tag{4}$$

式中：$m,n\in[1,w\times h]$；$f()$ 和 $g()$ 为线性映射，具体可表示为 $f(\boldsymbol{C}^{(k)}(m))=\boldsymbol{C}^{(k)}(m)\boldsymbol{W}_f+\boldsymbol{b}_f$，$g(\boldsymbol{C}^{(k)}(n))=\boldsymbol{C}^{(k)}(n)\boldsymbol{W}_g+\boldsymbol{b}_g$，$\boldsymbol{W}_f,\boldsymbol{W}_g\in\boldsymbol{R}^{c\times\frac{c}{8}}$，$\boldsymbol{b}_f,\boldsymbol{b}_g\in\boldsymbol{R}^{\frac{c}{8}}$。为后续表达方便，将式 (3) 表示为 $\boldsymbol{A}^{(k)}=f_{\mathrm{SpaAtten}}(\boldsymbol{C}^{(k)})$。

注意力增强后的关联映射 $\boldsymbol{A}^{(1)},\boldsymbol{A}^{(2)}$ 中包含不同尺度大小的目标物体的关联信息，为了充分挖掘多尺度信息，这里构建空洞空间金字塔池化（atrous spatial pyramid pooling, ASPP）模块，ASPP 包含多尺度并行空洞卷积层，该层包含不同采样步长的空洞卷积操作，这样不同感受野的空洞卷积可以对不同尺度的篡改区域信息进行检测。ASPP 中空洞卷积的采样步长被分别设定为 {6,12,18}，同时并行包含一个 1×1 卷积和全局池化和上采样操作，这样的经过多尺度信息挖掘的特征映射再进一步输入到空间注意力模块 $f_{\mathrm{SpaAtten}}()$ 中进行增强，得到包含多尺度信息和长距离空间注意力增强信息的关联特征映射为

$$\ddot{\boldsymbol{A}}^{(k)}=f_{\mathrm{SpaAtten}}\left(f_{\mathrm{ASPP}}\left(\boldsymbol{A}^{(k)}\right)\right) \tag{5}$$

式中 $f_{\mathrm{ASPP}}()$ 表示 ASPP 模块。得到的关联特征映射 $\ddot{\boldsymbol{A}}^{(k)}$ 进一步输入到基于深度可分离卷积和残差结构构建的定位结果重构网络中，重构出准确的篡改区域，所提定位结果重构网络如图 2 所示。深度可分离卷积首先执行空间卷积，同时保持通道独立，然后进行深度卷积操作[22]。具体来讲，输入的关联特征映射 $\ddot{\boldsymbol{A}}^{(k)}$ 为 $P\times w\times h$ 的张量，张量 P 个通道中，每一个通道的特征映射均与一个 3×3 的卷积核进行空间卷积运算，从而得到 P 个 $w\times h$ 的特征映射，并重组为 $P\times w\times h$ 的张量，然后再与 $1\times1\times P\times Q$ 个卷积核进行常规的卷积运算，得到 $Q\times w\times h$ 的张量。深度可分离卷积在确保特征提取能力的同时，可以大幅减少存储空间。设深度可分离卷积操作为 $f_{\mathrm{dsc}}()$，对输入的特征映射 $\boldsymbol{X}$，要经历如下运算：

$$\boldsymbol{X}'=\mathrm{Upsample}(\mathrm{Relu}(\mathrm{BN}(f_{\mathrm{dsc}}(\mathrm{Relu}(\mathrm{BN}(f_{\mathrm{dsc}}(\boldsymbol{X}))))+\boldsymbol{X}))) \tag{6}$$

式 (6) 中主要包含了 2 次深度可分离卷积操作，一次残差连接，以及 BN 层、ReLU 层、上采样层。重复式 (6)3 次，得到最终的输出，其中，第 3 次重复需在上采样操作前增加一个 1×1 的卷积操作，将输出转换为一通道的定位结果图。

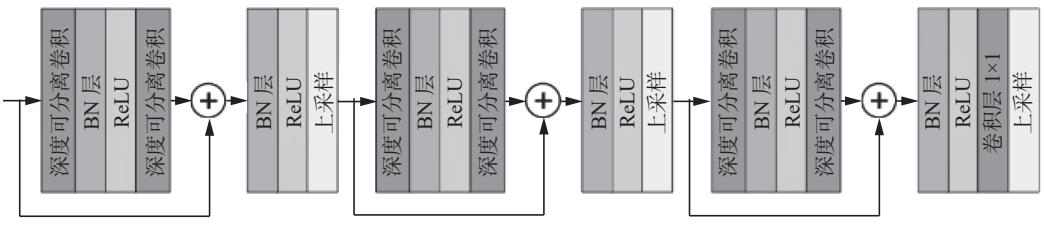

图 2　定位结果重构网络结构示意

Fig. 2　Architecture of the mask reconstruction network

2.2　尺度可变的关联性计算

本研究提出采用尺度可变关联性计算，打破 DMVN 所设计的强空间约束，直接采用点积运算再进行降序排序并进行强关联因子的截断操作，

从而使得模型具有更好的尺度灵活性，可以处理更大分辨率的图像。

以下主要介绍式 (2) 中，尺度可变关联映射计算 $f_{\mathrm{CORR}}()$ 的计算过程。设 $\boldsymbol{F}_l^{(k)}$ 为特征提取模块得到的图 k 的第 l 层特征，$k=1,2$，$l=3,4,5$。首先对输入特征进行 L2 标准化：

$$\bar{\boldsymbol{F}}_l^{(k)}(i,j)=\mathrm{L2\_Norm}\left(\boldsymbol{F}_l^{(k)}(i,j)\right)=\frac{\boldsymbol{F}_l^{(k)}(i,j)}{\left\|\boldsymbol{F}_l^{(k)}(i,j)\right\|_2} \tag{7}$$

通过 L2 标准化可以约束特征值的分布范围，获得标准化后的特征映射 $\bar{\boldsymbol{F}}_l^{(k)}$。这里进一步对标准化后的特征映射进行通道级的重构和增强，采用通道注意力模型进行通道加权。本研究采用了一种轻量级的高效通道注意力增强模型[23]，首先对输入的特征映射进行全局平均池化 $\boldsymbol{V}_l^{(k)}=f_{\mathrm{avg\_pool}}\left(\bar{\boldsymbol{F}}_l^{(k)}\right)$，输入的特征映射为 $C\times w\times h$ 的张量，输出为 C 维的一个向量，当 $l=3$ 时 $C=256$，$l=4,5$ 时 $C=512$，$w=\dfrac{W}{8}$，$h=\dfrac{H}{8}$。然后采用一个核尺度为 3 的一维卷积操作 $f_{\mathrm{conv1}d}$ 进行特征提取并进行归一化：

$$\boldsymbol{O}_l^{(k)}(t)=\frac{1}{1+\exp\left(-f_{\mathrm{conv1}d}\left(\boldsymbol{v}_l^{(k)}(t)\right)\right)} \tag{8}$$

经过注意力增强的特征映射计算为

$$\bar{\bar{\boldsymbol{F}}}_l^{(k)}=\boldsymbol{O}_l^{(k)}\times\bar{\boldsymbol{F}}_l^{(k)} \tag{9}$$

然后采用点积计算 2 组特征映射的关联性：

$$\boldsymbol{C}_l^{(k_1k_2)}(i_{12},j_{12},m_{12})=\bar{\bar{\boldsymbol{F}}}_l^{(k_1)}(i_1,j_1)^{\mathrm{T}}\cdot\bar{\bar{\boldsymbol{F}}}_l^{(k_2)}(i_2,j_2) \tag{10}$$

式中：$k_1\in\{1,2\}$；$k_2\in\{1,2\}$。

再对获得的关联映射进行降序排序，并采取前 T 个值的截断操作：

$$\tilde{\boldsymbol{C}}_l^{(k_1k_2)}(i_{12},j_{12},1:T)=\mathrm{Top\_T}\left(\mathrm{Sort}\left(\boldsymbol{C}_l^{(k_1k_2)}(i_{12},j_{12},:)\right)\right) \tag{11}$$

最终 2 个得到的特征映射分别为

$$\boldsymbol{C}_l^{(1)}=\mathrm{L2\_norm}\left(\mathrm{Concat}\left(\tilde{\boldsymbol{C}}_l^{(12)},\tilde{\boldsymbol{C}}_l^{(11)}\right)\right) \tag{12}$$

$$\boldsymbol{C}_l^{(2)}=\mathrm{L2\_norm}\left(\mathrm{Concat}\left(\tilde{\boldsymbol{C}}_l^{(21)},\tilde{\boldsymbol{C}}_l^{(22)}\right)\right) \tag{13}$$

式中：$\mathrm{Concat}(\cdot,\cdot)$ 表示将截断后的 2 个关联映射的 T 个通道进行联结，本研究 $T=32$。由于在式 (11) 中采用了与空间关系无关的排序和截断操作，且式 (10) 采用了点积运算，因此关联性计算可以处理不同大小的输入特征映射，可以做到尺度可变的关联性计算。结合 3 个层级层次化特征后的关联映射为 $\boldsymbol{C}^{(1)}=\left\{\boldsymbol{C}_3^{(1)},\boldsymbol{C}_4^{(1)},\boldsymbol{C}_5^{(1)}\right\}$，$\boldsymbol{C}^{(2)}=\left\{\boldsymbol{C}_3^{(2)},\boldsymbol{C}_4^{(2)},\boldsymbol{C}_5^{(2)}\right\}$。

2.3　块级对抗学习优化

尺度可变有约束图像拼接检测与定位网络采用交叉熵损失进行预训练，这种训练方式并未充分考虑输出结果的得分的分布，为此本研究提出采用块级对抗学习进行进一步的优化，使得输出的定位结果与真实定位结果分布更加接近。在数值上，采用变换后的真实定位结果 $\widetilde{\boldsymbol{G}}_t$（取值在 [0,0.1] 和 [0.9,1] 区间），使得生成的定位结果 $\boldsymbol{M}$ 得分更多分布在 [0,0.1] 和 [0.9,1] 上，避免落在模糊区间 [0.1,0.9]；通过将定位结果 $\boldsymbol{M}$ 与图像进行相乘，可以借助图像的纹理和边缘信息进行定位结果优化。

块级对抗优化可估计真实定位结果分布，采用对抗学习方式，促使所提网络能够产生与真实定位结果难以区分的定位结果。“块级”指对抗学习网络的输出是一个矩阵，而矩阵的每一个元素的值表示定位结果中不同块与真实定位结果的接近程度。由此，在块级对抗学习优化过程中，尺度可变的有约束图像拼接检测与定位网络的混合损耗函数为

$$L_{\mathrm{dm}}=L_{\mathrm{ce}}+\eta L_{\mathrm{adv}}^{G} \tag{14}$$

式中：L_{ce} 表示空间交叉熵损失，L_{adv}^{G} 表示检测与定位网络的对抗优化损失，η 为该对抗优化损失的权重。给定可疑拼接图像 $\boldsymbol{I}^{(1)}$ 和可疑拼接来源图像 $\boldsymbol{I}^{(2)}$，以及对应的真实定位结果 $\boldsymbol{G}_t^{(1)}$ 和 $\boldsymbol{G}_t^{(2)}$，通过尺度可变的有约束图像拼接检测与定位网络计算的定位结果为 $\boldsymbol{M}^{(1)}$ 和 $\boldsymbol{M}^{(2)}$，则空间交叉熵损失为

$$L_{\mathrm{ce}}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{k,h,w,s}\boldsymbol{G}_{t(n,h,w,s)}^{(k)}\log\left(\boldsymbol{M}_{(n,h,w,s)}^{(k)}\right) \tag{15}$$

式中：N 表示一个批次中样本的数量；h、w 表示定位结果的高、宽；s 表示类别数；本研究中为二分类问题即篡改和非篡改区域的判定，故 $s=0,1$；k 表示可疑拼接图像和可疑拼接来源图像的标签，$k=1,2$。

块级对抗学习优化网络的结构如图 1 所示，块级对抗学习优化网络作为判别器，本研究将该网络表示为 $\mathrm{Dis}(\cdot)$，其输入为生成的定位结果 $\boldsymbol{M}^{(k)}$，变换之后的真实定位结果为

$$\widetilde{\boldsymbol{G}}_t^{(k)}(x)=\begin{cases}\max\left(\boldsymbol{M}^{(k)}(x),\eta_1\right), & \boldsymbol{G}_t^{(k)}(x)=1\\ \min\left(\boldsymbol{M}^{(k)}(x),\eta_0\right), & \boldsymbol{G}_t^{(k)}(x)=0\end{cases} \tag{16}$$

式中：$\eta_1=0.9$，$\eta_0=0.1$，本研究用该公式模拟“好”定位结果。对抗学习过程中损失函数为

$$L_{\mathrm{adv}}^{G}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{k}\mathrm{avg}\left(\left(\mathrm{Dis}\left(\boldsymbol{M}_{(n)}^{(k)}\times\boldsymbol{I}_{(n)}^{(k)}\right)-1\right)^2\right) \tag{17}$$

$$\begin{aligned}L_{\mathrm{adv}}^{D}=-\frac{1}{N}\sum_{n=1}^{N}\sum_{k}\Bigg[&\mathrm{avg}\left(\mathrm{Dis}\left(\boldsymbol{M}_{(n)}^{(k)}\times\boldsymbol{I}_{(n)}^{(k)}\right)^2\right)+\\&\mathrm{avg}\left(\left(\mathrm{Dis}\left(\widetilde{\boldsymbol{G}}_{t(n)}^{(k)}\times\boldsymbol{I}_{(n)}^{(k)}\right)-1\right)^2\right)\Bigg]\end{aligned} \tag{18}$$

式中：L_{adv}^{G} 为混合损耗中的检测与定位网络的对抗优化损失；L_{adv}^{D} 为块级对抗学习优化网络 $\mathrm{Dis}(\cdot)$ 的

损失函数，2 个损失函数采用对抗优化的方式交替使用，均需要通过优化网络Dis(·)计算得出。优化网络Dis(·)的输出是一个矩阵，该矩阵中每一位置的值代表了该位置对应在 $\boldsymbol{M}_{(n)}^{(k)}\times\boldsymbol{I}_{(n)}^{(k)}$ 或 $\widetilde{\boldsymbol{G}}_{t_{(n)}}^{(k)}\times\boldsymbol{I}_{(n)}^{(k)}$ 中的感受野区域被判定为真实定位结果的概率，这里采用avg(·)来求得矩阵的平均值，×表示逐元素相乘。

预训练的尺度可变有约束图像拼接检测与定位网络进行块级对抗学习优化的算法流程如算法 1 所示。

算法 1　块级对抗学习优化算法

输入　利用式 (15) 预训练的有约束图像拼接检测与定位网络

输出　块级对抗学习优化后的有约束图像拼接检测与定位网络

1) **FOR** 优化训练迭代次数 **DO**

2) **FOR** K 步 **DO**

3)采样 N 对待分析的图像 $\left\{\boldsymbol{I}_{(n)}^{(1)},\boldsymbol{I}_{(n)}^{(2)}|n=1,2,\cdots,N\right\}$ 以及对应的 N 对真实定位结果 $\left\{\boldsymbol{G}_{t_{(n)}}^{(1)},\boldsymbol{G}_{t_{(n)}}^{(2)}|n=1,2,\cdots,N\right\}$

4) 有约束图像拼接检测与定位网络计算 $\left\{\boldsymbol{M}_{(n)}^{(1)},\boldsymbol{M}_{(n)}^{(2)}|n=1,2,\cdots,N\right\}$

5) 固定有约束图像拼接检测与定位网络的参数，通过计算式 (18) 的梯度 $\nabla L_{\text{adv}}^{D}$，更新块级对抗学习优化网络Dis(·)

6) **END FOR**

7)采样 N 对待分析的图像 $\left\{\boldsymbol{I}_{(n)}^{(1)},\boldsymbol{I}_{(n)}^{(2)}|n=1,2,\cdots,N\right\}$ 以及对应的 N 对真实定位结果 $\left\{\boldsymbol{G}_{t_{(n)}}^{(1)},\boldsymbol{G}_{t_{(n)}}^{(2)}|n=1,2,\cdots,N\right\}$

8) 有约束图像拼接检测与定位网络计算 $\left\{\boldsymbol{M}_{(n)}^{(1)},\boldsymbol{M}_{(n)}^{(2)}|n=1,2,\cdots,N\right\}$

9) 固定块级对抗学习优化网络Dis(·)的参数，通过计算式 (14) 的梯度 ∇L_{dm}，更新有约束图像拼接检测与定位网络的参数

10) **END FOR**

3　实验分析

3.1　训练与实现

3.1.1　训练集与测试集的生成策略

在有监督训练中，需要大量具有标签掩模图的训练样本对。然而，现有的图像取证数据集规模不足以支撑有约束图像拼接检测与定位网络的训练，大多数图像取证数据集中的图像只包含篡改图像，不包含篡改来源图像，而获取靠人工篡改的图像对及其像素级的标签成本又过高。考虑到有约束图像拼接检测与定位方法主要利用图像的相似性特征进行匹配和定位，篡改区域较为明显或失真的边缘和语义上的不一致对检测结果影响较小，因此现有约束图像拼接检测与定位方法均采用了合成数据集进行模型的训练[12-13]。本研究使用 MS COCO 数据集自动生成合成图像。MS COCO 数据集主要应用于物体检测和语义分割等任务，其提供了大量包含物体分割标注的图像，有 82 783 幅训练图像和 40 504 幅测试图像（2014 版本）。本研究在 MS COCO 的训练图像上生成合成的训练图像对，在其测试图像上生成测试图像对，训练对和测试对完全是 MS COCO 2 个分开的子集，图像均被变换为 512×512 的尺寸。

对于合成训练对的生成，随机挑选一幅图像中的一个标注区域，经过不同的变换后粘贴到另外一个随机挑选的训练图上，由此获得 3 组训练对（2 组正样本对，1 组负样本）。采用 5 种不同的变换，即平移、旋转、缩放、明暗、形变。具体来讲，所有的复制区域都经过了平移变换，范围为 $\mathbb{U}(-512,512)$，并以 50% 的概率会经历其他的变换。旋转变换的取值为 $\mathbb{U}(-30,30)$，缩放变换取值为 $\mathbb{U}(0.5,4)$，明暗变换取值 $\mathbb{U}(-32,32)$，形变取值 $\mathbb{U}(0.5,2)$。选取的粘贴区域需符合一个基本要求，即其面积需要大于图像面积的 1% 且小于图像面积的 50%。由于 MS COCO 中的图像往往包含多个标注区域，因此，在 MS COCO 的训练图像上遍历 5 次，生成了 1 035 255 对训练对，其中 1/3 为前景对，1/3 背景对，以及 1/3 负样本对。

测试合成图像对的生成，也采取了类似的策略。主要在前景对上测试定位效果，这是因为背景对事实上是最简单的且没有经过任何变换的形式。根据粘贴区域的面积比例，测试对主要分为 3 组，即复杂集（粘贴区域 1%~10%）、一般集（10%~25%）和简单集（25%~50%）。

3.1.2　网络的实现和训练策略

所提方法训练测试代码均在 PyTorch 上实现。尺度可变有约束图像拼接检测与定位网络首先仅依靠空间交叉熵损失进行训练，即式 (15)，特征提取模块的参数由用来进行图像分类的 VGG16 进行初始化。在合成训练对上，尺度可变有约束图像拼接检测与定位网络经过 3 轮（epoch）训练，批大小设定为 16。本研究生成了 1 035 255 对训练对，3 轮训练需要耗时近 121 h，经实验观察在第 3 轮训练后期，IoU 值在 0.71 附近波动，达到收敛状态，如图 3 所示。在尺度可变的有约束图像拼接检测与定位网络预训练阶段，采用了 Adadelta 优化算法以及 PyTorch 的默认设置。对经过预训练的有约束图像拼接检测与定位网络，采用块级

对抗优化，即采用算法 1 对预训练网络进行优化。块级对抗学习优化进行了 1 轮训练，耗时 46 h，采用了 Adam 优化算法，判别网络的学习率设定为 0.000 2，检测和定位网络优化学习率设定为 0.000 01，对抗损失权重η为 0.01。图 4 给出了预训练模型与对抗优化模型的得分分布统计，将 0.0~1.0 的得分每 0.1 划分为一个区间，统计生成定位结果在每个得分区间上的像素点数量的占比。

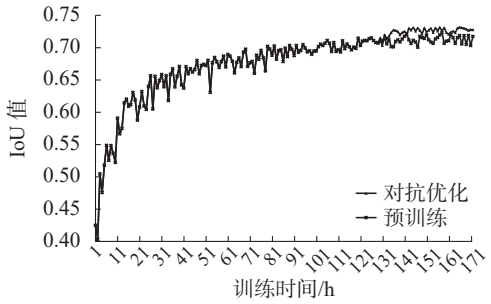

图 3　预训练与对抗优化 IoU 值变化趋势

Fig. 3　IoU scores of pre-training process and adversarial optimization process

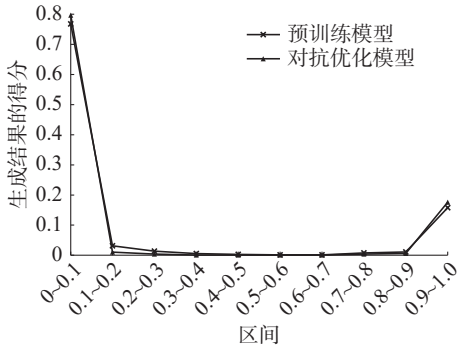

图 4　预训练模型与对抗优化模型得分分布统计

Fig. 4　Score distribution of the pretrained model and the adversarial optimized model

根据图 4 的统计结果，由于预训练模型在 [0,0.1] 和 [0.9,1.0] 具有较高的分布，而且在这一分布区间已经可以很好地区分篡改区域和非篡改区域，对抗学习的目的是通过对抗优化的方法使得在 [0.1,0.9] 区间取值的得分能够更少，在 [0,0.1] 和 [0.9,1.0] 更多，故式 (16) 中设定 $\eta_1 = 0.9$，$\eta_0 = 0.1$。

3.1.3　测试数据集及评价指标

所提方法主要在 3 个数据集上进行测实验证。

1) 合成数据集[13]：该集合中图像的篡改区域可能经过多种变换操作，根据篡改区域面积比例分为 3 个子集，即复杂集、一般集和简单集。每个子集有 3 000 对图像，合成数据集共有 90 00 对图像。

2) 成对 CASIA 数据集：Wu 等[12] 通过将 CASIA TIDEv2.0 数据集的 1 821 幅拼接图像和其篡改来源图像进行成对组合生成了 3 642 对正样本，同时通过成对组合 7 491 幅 CASIA 定义类别范围内的图像整理了 5 000 对负样本，该数据集由于缺乏标签掩模图仅用作检测效果验证。

3) MFC2018 数据集：MFC2018 竞赛提供的测试集，即 MFC2018 Eval Part1 Ver1，包含 1 327 对正样本对和 16 673 对负样本对，竞赛还提供了测试代码可以对整体的检测效果进行打分。

评价指标：在评价定位效果方面，采用了像素级 IoU（intersection over union）、MCC（matthews correlation coefficient）、NMM（nimble mask metric）。IoU 经常被用于语义分割和物体检测的评价，这里仅对篡改区域的 IoU 得分进行衡量，对所有测试对计算平均 IoU、MCC 和 NMM 得分。检测效果评价主要采用准确率、召回率和 F1-score、AUC（area under curve）和 EER（equal error rate）。

3.2　定位效果对比

在合成数据集上的比较，主要通过 IoU、MCC 和 NMM 衡量算法篡改定位的准确程度，运算结果见表 1。

表 1　合成数据集定位效果对比结果

Table 1　Localization result comparison on the synthetic datasets

对比算法	图像尺寸	复杂集			一般集			简单集		
		IoU	MCC	NMM	IoU	MCC	NMM	IoU	MCC	NMM
DMVN[12]	256×256	0.277 2	0.353 3	−0.438 2	0.681 8	0.757 0	0.404 2	0.819 8	0.854 4	0.677 0
DMAC[13]	256×256	0.511 4	0.630 8	0.033 5	0.827 9	0.881 5	0.684 0	0.922 2	0.939 5	0.868 5
DMAC-adv	256×256	0.543 3	0.658 4	0.102 6	0.831 7	0.883 3	0.687 7	0.923 7	0.941 1	0.865 5
AttentionDM[14]	256×256	0.722 8	0.810 8	0.479 3	0.960 0	0.969 0	0.938 0	0.898 0	0.932 0	0.825 0
VSAD	256×256	0.697 6	0.784 3	0.451 6	0.880 1	0.917 8	0.801 5	0.951 2	0.961 6	0.933 1
VSSD	256×256	0.708 5	0.791 3	0.469 8	0.886 0	0.919 3	0.812 5	0.951 1	0.961 7	0.933 6
预训练模型	256×256	0.715 2	0.799 8	0.489 1	0.886 2	0.921 7	0.817 1	0.952 6	0.963 4	0.936 2

续表 1

对比算法	图像尺寸	复杂集			一般集			简单集		
		IoU	MCC	NMM	IoU	MCC	NMM	IoU	MCC	NMM
对抗优化模型	256×256	0.731 5	0.813 6	0.503 4	0.898 0	0.930 9	0.823 6	0.959 3	0.968 8	0.937 1
DMAC-adv-SR256	512×512	0.642 6	0.749 6	0.311 5	0.852 1	0.897 0	0.742 2	0.936 8	0.950 4	0.902 4
DMAC-adv-SR128	512×512	0.691 1	0.793 0	0.419 8	0.860 2	0.904 7	0.774 2	0.934 9	0.948 4	0.910 8
AttentionDM-SR256	512×512	0.755 7	0.838 7	0.560 8	0.900 5	0.933 9	0.841 0	0.959 7	0.968 6	0.946 5
AttentionDM-SR128	512×512	0.760 8	0.847 7	0.581 9	0.895 3	0.931 3	0.841 6	0.954 0	0.963 9	0.942 2
VSAD	512×512	0.758 3	0.834 6	0.558 1	0.88 63	0.920 1	0.795 2	0.953 7	0.963 7	0.928 9
VSSD	512×512	0.779 3	0.851 2	0.603 3	0.900 1	0.929 8	0.831 1	0.954 4	0.965 5	0.930 2
预训练模型	512×512	0.793 2	0.865 4	0.632 4	0.904 6	0.935 3	0.839 0	0.956 6	0.966 5	0.934 2
对抗优化模型	512×512	0.803 3	0.875 0	0.642 4	0.910 9	0.940 6	0.848 4	0.962 7	0.971 0	0.944 1

3.2.1　分步验证

本研究所提算法主要对比了“尺度可变的有约束图像拼接检测与定位网络”的“预训练模型”和“对抗优化模型”，以及“VGG +可扩展相关计算+ASPP+解码器（VGG + Scalable Correlation Computation+ASPP+ Decoder, VSAD）”版本和“VGG +可扩展相关计算+基于空间注意的 ASPP+解码器（VGG +Scalable Correlation Computation+Spatial Attention based ASPP+ Decoder, VSSD）”版本的消融实验，并验证了输入分别为 256×256、512×512 时的测试结果。

1) VSAD 与 AttentionDM 均采用 VGG 进行特征提取，采用 ASPP 和逐层上采样的 Decoder，只替换了关联性计算模块为尺度可变关联性计算。

2) VSSD 进一步将 ASPP 模块增加了空间注意力增强。

3)“预训练模型”指将 VSSD 的逐层上采样 Decoder 替换为基于深度可分离卷积和残差结构的定位结果重构网络，即本研究“尺度可变的有约束图像拼接检测与定位网络”经过预训练后的模型。

4)“对抗优化模型”是指经过本研究所提块级别对抗优化后的“尺度可变的有约束图像拼接检测与定位网络”模型。

DMVN、DMAC、DMAC-adv、AttentionDM 只能处理 256×256 大小的图像，DMAC-adv-SR256、DMAC-adv-SR128、AttentionDM-SR256、AttentionDM-SR128 为滑动窗口版本[13]，处理图像大小为 512×512。

AttentionDM 中采用的是一种结合了平均值、最大值和排序值，并对关联位置有约束的一种关联性计算模块，DMVN、DMAC 和 AttentionDM 均采用这种关联性计算模型，但是这种关联性计算模型导致所提网络只能处理固定尺寸大小的图像，DMVN、DMAC 和 AttentionDM 均只能处理 256×256 大小的图像。在替换为关联性计算模块为尺度可变的关联性计算后，VSAD 在处理 256×256 大小的图像时，比 AttentionDM 效果有一定的下降，但是在处理 512×512 大小的图像时，却能取得比 AttentionDM 滑动窗口版本即 AttentionDM-SR256 更优的结果，而且滑动窗口版本运算效率偏低，VSAD 可以直接处理 512×512 大小的图像，效率更高。如果在空洞空间金字塔池化前后加入空间注意力增强，即 VSSD，效果可以进一步提升；在采用基于深度可分离卷积和残差结构的定位结果重构网络后，“预训练模型”有了进一步提升，特别是“对抗优化”模型在合成数据集上已经可以取得明显优于 AttentionDM 的效果。由表 1 也可以看出，在 256×256 的合成图像中，预训练模型可以取得与 AttentionDM 接近的定位效果，经过对抗训练优化后，对抗优化模型则可以取得优于 AttentionDM 的结果。在 512×512 的图像中，预训练模型和对抗优化模型均可以取得更好的定位效果，而且无须滑动窗口重复计算，运算效率也更高。“VSAD”版本尽管可以处理任意尺度图像，但是在处理 256×256 的图像时，效果有明显下降。而加上空间注意力增强空洞空间金字塔池化模块和基于深度可分离卷积和残差结构的定位结果重构网络后，“预训练模型”则可以在 256×256 的图像上取得接近 AttentionDM 的效果，这证明了整体尺度可变的有约束图像拼接检测与定位网络的有效性。而且从不同角度衡量，对抗优化模型在合成数据集上均具有明显优势。

3.2.2　对抗优化有效性验证

为了进一步证明对抗优化的有效性，图 3 提

供了利用式 (15) 交叉熵损失训练 4 轮的“预训练”曲线，以及使用交叉熵损失预训练 3 轮后利用式 (14) 采用对抗优化方法训练一轮的“对抗优化”曲线，在合成数据集的 256×256 复杂集上的 IoU 值的变换趋势，证明了所提网络结构以及块级对抗优化模型的有效性。可以看出使用交叉熵损失进行预训练在第 3 轮训练的末期已经收敛，IoU 值在 0.71 上下波动，而进一步进行第 4 轮训练，IoU 值也没有明显变化；但是在第 4 轮进行块级对抗优化，其 IoU 值则会进一步提高。为了验证，块级对抗学习机制对预训练的尺度可变有约束图像拼接检测与定位网络进行优化，可以使得生成的定位结果与标签值分布更加接近，图 4 统计了所提网络在合成数据集上输入图像大小为 256×256 时生成结果的得分分布，可以看到“对抗优化模型”相比“预训练模型”在 [0,0.1] 与 [0.9,1] 区间有更多的分布。块级对抗优化的目标就是让生成结果拟合$\widetilde{\boldsymbol{G}}_t$的分布（取值在 [0,0.1] 与 [0.9,1] 区间），实验证明块级对抗优化的有效性。

3.2.3　尺度可变性验证

为了进一步验证所提方法，特别是“尺度可变的关联性计算”的有效性，图 5 给出了在合成数据集复杂集上 IoU 值随输入图像尺度变化趋势，其中 AttentionDM 网络仅能处理 256×256 大小的图像，这里用虚线表示其 IoU 值的水平，可以看到仅将关联性计算模块替换为“尺度可变的关联性计算”时，在 256×256 尺度下 VSAD 的效果要差于 AttentionDM，但是随着输入图像尺度的增大，则可以超过 AttentionDM 在 256×256 尺度下的 IoU 值的水平；为了能够在各个尺度下均取得更优的水平，本研究对 VSAD 进一步从网络结构和优化方式上进行改进，最终的对抗优化模型则可以在 256×256 尺度下，IoU 值也更高，而且随着输入图像尺度的增大，定位精确度也进一步提高。

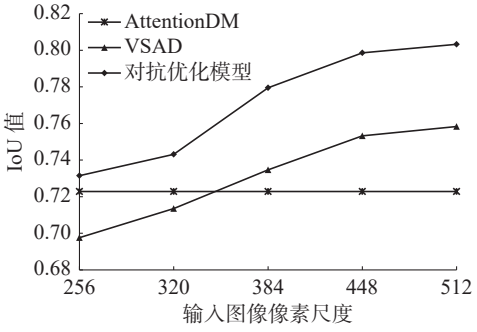

图 5　合成数据集复杂集上 IoU 值随输入图像尺度变化趋势

Fig. 5　IoU scores with images of different scales on the difficult set of the synthetic dataset

3.2.4　运算效率验证

表 2 对比了几种有约束图像拼接检测与定位算法的运算效率，对比算法均在合成数据集上进行运算测试，测试环境为 Intel(R) Core(TM) i7-5930K CPU @ 3.50 GHz, 64 GB RAM 和一块 GPU (TITAN X)。所提方法的输入图像大小为 256×256，运算速度略慢于 DMAC，略快于 AttentionDM。但是，正如前述所分析，DMAC 与 AttentionDM 无法处理任意尺度的图像，当运算 512×512 大小的图像时，需要采用滑动窗口策略，重复运算明显增多，效率明显下降，但所提尺度可变的有约束图像拼接检测与定位网络可直接处理 512×512 大小的图像，运算效率具有明显优势，事实上预训练模型与对抗优化模型只是参数不同，结构是一致的，因此运算效率相同。

表 2　运行效率分析

Table 2　Computational time analysis

对比算法	运算时间	图像大小
DMVN	0.296 8	256×256
DMAC/ DMAC-adv	0.028 8	256×256
AttentionDM	0.030 6	256×256
预训练/对抗优化模型	0.029 3	256×256
DMAC-adv-SR256	0.684 8	512×512
DMAC-adv-SR128	3.111 6	512×512
AttentionDM-SR256	0.732 5	512×512
AttentionDM-SR128	3.391 1	512×512
预训练/对抗优化模型	0.199 3	512×512

3.3　检测效果对比

3.3.1　成对 CASIA 数据集

在 Wu 等[12] 的研究中，由于缺乏其他 CISDL 算法，直接与复制移动篡改检测算法进行对比（即 Christlein 等[24]、Luo 等[25]、Ryu 等[26]、Cozzolino 等[27]），本实验直接借鉴了 Wu 等[12] 的研究中公布的得分，对比结果见表 3。预训练模型及对抗优化模型的篡改概率（即检测得分）主要计算篡改区域的平均得分。对于每个生成掩模图，计算得分大于 0.5 的区域的平均得分$\{s^{(k)}|k=1,2\}$，最后的篡改概率则为 2 个生成掩模图的平均值$(s^{(1)}+s^{(2)})/2$。在 CASIA 数据集上，所提方法可以取得较好的效果，特别是准确率是各个深度学习方法中最高的，其中，对抗优化模型能够取得最高的 F_1-score 得分 0.935 4。Christlein 等[24]、Luo 等[25]、Ryu 等[26]、Cozzolino 等[27]4 个对比算法不依赖于合成数据训练集，采用的是人工设计的特征

进行匹配和定位，为有约束图像拼接检测与定位任务提供了一个“基线”效果。可以看出依赖于合成数据集训练的网络（DMVN、DMAC、AttentionDM 以及本研究所提方法），在 CASIA 这种公开的、人工篡改的图像取证数据集上，均可以取得优于这些“基线”模型的效果，这也说明通过合成数据集训练的模型在实际图像取证获取的图像上具有良好的鲁棒性。这是因为有约束图像拼接检测与定位主要挖掘的是篡改区域和篡改来源区域的视觉相似性特征，不同于其他的基于深度学习的篡改图像定位方法主要挖掘像素的统计分布不一致特征，视觉相似性特征鲁棒性相对更强，跨库效果更好。

表 3　成对 CASIA 数据集检测效果对比结果

Table 3　Detection result comparison on the paired CASIA datasets

算法	准确率	召回率	F_1-score
Christlein等 [24]	0.516 4	0.829 2	0.636 4
Luo等[25]	0.996 9	0.535 3	0.696 6
Ryu等[26]	0.961 4	0.589 5	0.730 9
Cozzolino等[27]	0.989 7	0.633 4	0.772 5
DMVN-loc	0.915 2	0.791 8	0.849 1
DMVN-det	0.941 5	0.790 8	0.859 6
DMAC	0.925 5	0.866 8	0.895 2
DMAC-adv	0.965 7	0.857 6	0.908 5
AttentionDM	0.928 8	0.920 4	0.924 6
预训练模型	0.990 9	0.870 7	0.926 9
对抗优化模型	0.987 2	0.888 8	0.935 4

3.3.2　MFC2018 数据集

由表 4 可知，本研究所提方法始终具有更低的 EER 值，即误报率更低，且对抗优化模型在 512×512 输入时，可以获得更高的 AUC 值（0.794 6）和最低的 EER 值（0.205 4）。图 6 和图 7 中的“本研究所提方法”指对抗优化模型。图 6 提供了所提方法随着输入图像尺度不同，检测得分的变化趋势图，可以看到随着输入图像的变大，所提方法可以挖掘更加丰富的信息，AUC 得分会逐渐提高，而 EER 则会逐渐降低。图 7 进行了定位效果的比较和展示，所提方法具有一定的检测和定位优势，对于面积较小的篡改区域，如第 2 和第 3 对图像中的行人和广告牌，所提方法可以给出更加准确的轮廓和边界。第 4 对图像中，存在很多行人的干扰，所提方法误检率相对更低，检测准确度相对更高。

综上所述，本研究所提方法能够取得优于 DMVN、DMAC-adv、AttentionDM 的效果，特别是提出了一种尺度可变的有约束图像拼接检测与定位网络，所设计的尺度可变的关联性计算模型使得该检测与定位网络能够处理任意尺寸的图像，当处理分辨率较大的图像时，不需要像 DMVN、DMAC-adv、AttentionDM 采用滑动窗口策略，所提方法相对滑动窗口的准确度及效率均有提升（表 2）。但是如果只替换 AttentionDM 中的关联计算模块为尺度可变的关联性计算，模型效果会有一定下降，进一步设计解码器部分，使用空间注意力增强空洞空间金字塔池化模块和基于深度可分离卷积和残差结构的定位结果重构网络，在 256×256 则可以达到接近 AttentionDM 的效果（表 1）。由表 1~4 可知，预训练模型尽管其在大尺寸图像（512×512）可以取得不错的效果，但是在 256×256 的图像上，效果仍然略差于 AttentionDM。为此，本研究进一步提出了块级对抗优化机制，设计了块级对抗学习优化网络，使得对抗优化后的模型能够取得优于 AttentionDM 的效果。由此，证明了所提网络结构，以及块级对抗学习优化机制的有效性。

表 4　MFC2018 数据集检测效果对比结果

Table 4　Detection result comparison on the MFC2018 dataset

算法	AUC	EER
DMVN-loc	0.658 4	0.400 0
DMVN-det	0.697 0	0.366 5
DMAC	0.754 2	0.312 3
DMAC-adv	0.751 1	0.309 3
AttentionDM	0.792 2	0.275 6
预训练模型（256×256）	0.732 4	0.267 5
预训练模型（512×512）	0.785 4	0.214 6
对抗优化模型（256×256）	0.747 0	0.253 0
对抗优化模型（512×512）	0.794 6	0.205 4

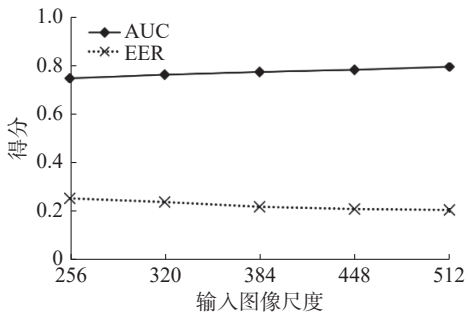

图 6　MFC2018 数据集上本研究所提方法得分随图像尺度变化趋势

Fig. 6　Scores of our method with different image scales on the MFC2018 dataset

图 7　MFC2018 数据集检测与定位效果图对比

Fig. 7　Visual comparisons on the MFC2018 dataset

4　结束语

本研究对有约束图像拼接检测与定位展开研究，提出一种尺度可变的有约束图像拼接检测与定位方法以及块级对抗优化机制。所提尺度可变的有约束图像拼接检测与定位网络主要进行了以下改进：利用强关联因子截断操作，设计了一种基于高效通道注意力增强的尺度可变关联计算模型，从而实现处理任意尺度大小的图像；设计空间注意力增强的空洞空间金字塔池化对关联映射的多尺度信息进行挖掘，设计基于深度可分离卷积和残差结构的定位结果重构网络。此外，为了进一步优化预训练模型，设计了一种块级对抗学习机制，使得生成的定位结果得分分布与标签值分布更加接近。在公开数据集上的大量实验证明了本研究所提方法的有效性。

参考文献:

[1] VERDOLIVA L. Media forensics and DeepFakes: an overview[J]. IEEE journal of selected topics in signal processing, 2020, 14(5): 910−932.

[2] COZZOLINO D, VERDOLIVA L. Noiseprint: a CNN-based camera model fingerprint[J]. IEEE transactions on information forensics and security, 2020, 15: 144−159.

[3] WANG Junke, WU Zuxuan, OUYANG Wenhao, et al. M2TR: multi-modal multi-scale transformers for deepfake detection[C]//Proceedings of the 2022 International Conference on Multimedia Retrieval. Newark: ACM, 2022: 615−623.

[4] LIU Yaqi, XIA Chao, ZHU Xiaobin, et al. Two-stage copy-move forgery detection with self deep matching and proposal SuperGlue[J]. IEEE transactions on image processing, 2022, 31: 541−555.

[5] ZHANG Yulan, ZHU Guopu, WU Ligang, et al. Multi-task SE-network for image splicing localization[J]. IEEE transactions on circuits and systems for video technology, 2022, 32(7): 4828−4840.

[6] BAMMEY Q, VON GIOI R G, MOREL J M. An adaptive neural network for unsupervised mosaic consistency analysis in image forensics[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 14182−14192.

[7] LIU Yaqi, GUAN Qingxiao, ZHAO Xianfeng, et al. Image Forgery Localization based on Multi-Scale Convolutional Neural Networks[C]//Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. Innsbruck: ACM, 2018: 85−90.

[8] PENG Bo, WANG Wei, DONG Jing, et al. Image forensics based on planar contact constraints of 3D objects[J]. IEEE transactions on information forensics and security, 2018, 13(2): 377−392.

[9] DAS S, ISLAM M S, AMIN M R. GCA-net: utilizing gated context attention for improving image forgery localization and detection[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022: 81−90.

[10] WANG Junke, WU Zuxuan, CHEN Jingjing, et al. ObjectFormer for image manipulation detection and localization[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 2354−2363.

[11] LIU Yaqi, LYU Binbin, JIN Xin, et al. TBFormer: two-branch transformer for image forgery localization[J]. IEEE signal processing letters, 2023, 30: 623−627.

[12] WU Yue, ABD-ALMAGEED W, NATARAJAN P. Deep matching and validation network: an end-to-end solution to constrained image splicing localization and detection [C]//Proceedings of the 25th ACM international conference on Multimedia. Mountain View: ACM, 2017: 1480−1502.

[13] LIU Yaqi, ZHU Xiaobin, ZHAO Xianfeng, et al. Adversarial learning for constrained image splicing detection and localization based on atrous convolution[J]. IEEE transactions on information forensics and security, 2019, 14(10): 2551−2566.

[14] LIU Yaqi, ZHAO Xianfeng. Constrained image splicing detection and localization with attention-aware encoder-decoder and atrous convolution[J]. IEEE access, 2020, 8: 6729−6741.

[15] MOREIRA D, BHARATI A, BROGAN J, et al. Image provenance analysis at scale[J]. IEEE transactions on image processing, 2018, 27(12): 6109−6123.

[16] YE Kui, DONG Jing, WANG Wei, et al. Feature pyramid deep matching and localization network for image forensics[C]//2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Honolulu: IEEE, 2018: 1796−1802.

[17] WAN Ji, WANG Dayong, HOI S C H, et al. Deep learning for content-based image retrieval: a comprehensive study[C]//Proceedings of the 22nd ACM international conference on Multimedia. Orlando: ACM, 2014: 157−166.

[18] PINTO A, MOREIRA D, BHARATI A, et al. Provenance filtering for multimedia phylogeny[C]//2017 IEEE International Conference on Image Processing. Beijing: IEEE, 2017: 1502−1506.

[19] LI Yuanman, ZHOU Jiantao. Fast and effective image copy-move forgery detection via hierarchical feature point matching[J]. IEEE transactions on information forensics and security, 2019, 14(5): 1307−1322.

[20] 王凯诚, 鲁华祥, 龚国良, 等. 基于注意力机制的显著性目标检测方法[J]. 智能系统学报, 2020, 15(5): 956−963.
WANG Kaicheng, LU Huaxiang, GONG Guoliang, et al. Salient object detection method based on the attention mechanism[J]. CAAI transactions on intelligent systems, 2020, 15(5): 956−963.

[21] 李涛, 高志刚, 管晟媛, 等. 结合全局注意力机制的实时语义分割网络[J]. 智能系统学报, 2023, 18(2): 282−292.
LI Tao, GAO Zhigang, GUAN Shengyuan, et al. Global attention mechanism with real-time semantic segmentation network[J]. CAAI transactions on intelligent systems, 2023, 18(2): 282−292.

[22] CHOLLET F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 1800−1807.

[23] WANG Qilong, WU Banggu, ZHU Pengfei, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 11531−11539.

[24] CHRISTLEIN V, RIESS C, JORDAN J, et al. An evaluation of popular copy-move forgery detection approaches [J]. IEEE transactions on information forensics and security, 2012, 7(6): 1841−1854.

[25] LUO Weiqi, HUANG Jiwu, QIU Guoping. Robust detection of region-duplication forgery in digital image[C]// 18th International Conference on Pattern Recognition. Hong Kong: IEEE, 2006: 746−749.

[26] RYU S J, LEE M J, LEE H K. Detection of copy-rotate-move forgery using zernike moments[M]//Lecture Notes

in Computer Science. Berlin: Springer Berlin Heidelberg, 2010: 51−65.

[27] COZZOLINO D, POGGI G, VERDOLIVA L. Efficient dense-field copy-move forgery detection[J]. IEEE transactions on information forensics and security, 2015, 10(11): 2284−2297.

作者简介：

刘亚奇，助理研究员，博士，主要研究领域为多媒体安全、图像取证、人工智能和模式识别。E-mail：liuyaqi@besti.edu.cn。

蔡强，教授，博士，主要研究方向为计算机图形学、计算几何、科学可视化、智能信息处理。发表学术论文 90 余篇。E-mail：caiq@btbu.edu.cn。

石磊，副研究员，博士，中国人工智能学会智能服务专委会委员，主要研究方向为智能信息处理、大数据分析与挖掘、社交网络搜索及人工智能。E-mail：leiky_shi@cuc.edu.cn。

2024 第二届全国人工智能应用场景创新挑战赛总决赛暨全国人工智能应用场景创新峰会

2024 2nd China's Innovation Challenge on Artificial Intelligence Application Scene Finals(CICAS2024)&National Artificial Intelligence Application Scenario Innovation Summit

为深入贯彻党中央、国务院关于加快人工智能产业创新发展的决策部署，落实国务院《新一代人工智能发展规划》，科技部、工信部等六部委联合印发的《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》，加速人工智能核心技术攻关，着力解决人工智能重大应用和产业化问题，推动人工智能与未来产业、实体经济融合发展，助力粤港澳大湾区建设具有全球影响力的产业科技创新中心，经大赛组委会研究，中国人工智能学会、科技部新一代人工智能发展研究中心定于 2025 年 1 月 17—19 日在深圳举办“‘场景驱动·数智强国’——2024 第二届全国人工智能应用场景创新挑战赛总决赛暨全国人工智能应用场景创新峰会”。现将大赛活动有关安排通知如下：。

时间地点

大赛时间：2025 年 1 月 17—19 日（星期五—星期日）

大赛地点：深圳宝安机场凯悦酒店

大赛官网：www.cicas.cn

主要议程

（一）第二届全国人工智能应用场景创新项目数据集评测暨答辩会

（二）全国人工智能应用场景人才培养与产业化闭门研讨会

（三）全国人工智能应用场景创新挑战赛总决赛暨峰会开幕式

（四）全国人工智能应用场景创新 100 目录清单、全国人工智能应用场景创新 TOP10 最佳实践城市、全国人工智能最具推广示范效应的解决方案等发布

（五）全国人工智能应用场景创新挑战赛专项赛启动仪式

（六）科学家企业家迎新春晚餐会

（七）全国人工智能应用场景创新峰会主论坛、专题分论坛（智能粮食工程专题论坛、智能林草专题论坛、具身智能专题论坛、智能制造专题论坛、智能低碳专题论坛、智能交通专题论坛、智能集成电路专题论坛等）

（八）全国人工智能应用场景创新挑战赛总决赛闭幕式暨颁奖典礼

联系方式：

全国人工智能应用场景创新挑战赛组委会秘书处

联系人：王老师　电话：15726613955（微信同号），郭老师　电话：18118467955（微信同号）

邮箱：zwh@cicas.cn