

## 基于跨流注意力增强中心差分卷积网络的CG图像检测

黄锦坤<sup>1,2</sup>, 黄远航<sup>1,2</sup>, 黄文敏<sup>1,2</sup>, 骆伟祺<sup>1,2</sup>

(1. 中山大学计算机学院, 广东 广州 510006; 2. 广东省信息安全技术重点实验室, 广东 广州 510006)

**摘要:** 随着计算机图形学 (computer graphics, CG) 技术在图像生成领域的日益成熟, 创造的图像逼真程度大幅提升。这些技术虽然在日常生活中被广泛应用并带来诸多便利, 但同时也有着许多安全隐患, 如果使用CG技术生成的伪造图像被恶意使用, 在互联网、社交媒体上广泛传播, 则可能对个人和企业权益造成损害。提出了一种创新的跨流注意力增强中心差分卷积网络, 致力于提高CG图像检测的准确性。模型中构建了一个双流结构, 旨在分别从图像中抽取语义特征和非语义的残差纹理特征。每个流中的普通卷积层被中心差分卷积所替代, 这一改进使模型能同时提取图像中的像素强度信息和像素梯度信息。此外, 通过引入一个跨流注意力增强模块, 该模型在全局层面上增强了特征提取能力, 并促进了两个特征流之间的互补。实验结果证明, 基于跨流注意力增强中心差分卷积网络的CG图像检测方法相比现有方法具有更优的性能。此外, 一系列消融实验进一步验证了所提模型设计的合理性。

**关键词:** 计算机图形学; CG图像检测; 中心差分卷积; 注意力机制

**中图分类号:** TP391

**文献标志码:** A

**DOI:** 10.11959/j.issn.2096-109x.2024083

## Cross-stream attention enhanced central difference convolutional network for CG image detection

HUANG Jinkun<sup>1,2</sup>, HUANG Yuanhang<sup>1,2</sup>, HUANG Wenmin<sup>1,2</sup>, LUO Weiqi<sup>1,2</sup>

1. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, china

2. Guangdong Key Laboratory of Information Security Technology, Guangzhou 510006, china

**Abstract:** With the maturation of computer graphics (CG) technology in the field of image generation, the realism of created images has been improved significantly. Although these technologies are widely used in daily life and bring many conveniences, they also come with many security risks. If forged images generated using CG technology are maliciously used and widely spread on the Internet and social media, they may harm the rights of individuals and enterprises. Therefore, an innovative cross-stream attention enhanced central difference convolutional net-

收稿日期: 2024-04-05; 修回日期: 2024-08-11

通信作者: 骆伟祺, luoweiqi@mail.sysu.edu.cn

基金项目: 国家自然科学基金 (62472458)

**Foundation Item:** The National Natural Science Foundation of China (62472458)

**引用格式:** 黄锦坤, 黄远航, 黄文敏, 等. 基于跨流注意力增强中心差分卷积网络的CG图像检测[J]. 网络与信息安全学报, 2024, 10(6): 96-108.

**Citation Format:** HUANG J K, HUANG Y H, HUANG W M, et al. Cross-stream attention enhanced central difference convolutional network for CG image detection[J]. Chinese Journal of Network and Information Security, 2024, 10(6): 96-108.

work was proposed, aiming at improving the accuracy of CG image detection. A dual-stream structure was constructed in the model, in order to extract semantic features and non-semantic residual texture features from the image. Vanilla convolutional layers in each stream were replaced by central difference convolutions, which allowed the model to simultaneously extract pixel intensity information and pixel gradient information from the image. Furthermore, by introducing a cross-stream attention enhancement module, the model enhanced feature extraction capability at the global level and promoted complementarity between the two feature streams. Experimental results demonstrate that this method outperforms existing methods. Additionally, a series of ablation experiments further verify the rationality of the proposed model design.

**Keywords:** computer graphics, CG image detection, central difference convolution, attention mechanism

## 0 引言

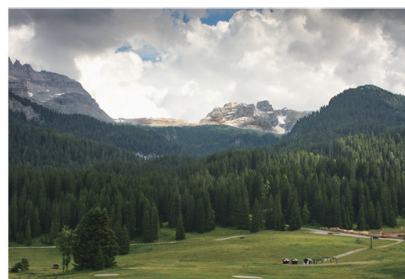
随着图像生成技术的快速发展,人们借助于图像生成技术,可以轻易地生成各种逼真的图像,并将其应用于游戏、电影、广告和媒体等领域,这不仅丰富了人们的日常生活,也带来了许多新的机遇。这其中,计算机图形学(computer graphics, CG)技术扮演了重要的角色。CG技术是一种通过数学算法将二维或三维图形转化为计算机显示器的栅格形式的科学,其利用计算机处理和生成图像、动画和其他视觉效果。其中,三维渲染技术是一种主流的CG技术,在各种图像生成场景中受到广泛使用。随着三维渲染软件和硬件技术的发展,生成的图像和视频效果也越来越好。图1展示了CG图像和真实自然图像。可以看到,由CG技术生成的图像与真实自然图像十分相似,能够很好地模拟自然环境下的光照、色彩、阴影等特点,因而不容易被人眼识别。正因如此,一旦这些逼真的CG图像被人在互联网、社交媒体上恶意地传播,或用于司法证据伪造、学术造假等场景,就会带来一定的安全隐患。因此,如何实现对CG图像的检测,以消除这些安全隐患,就成为重要的研究课题。

目前,相关学者已经提出了一些CG图像的检测方法。早期的工作<sup>[1-6]</sup>主要基于手工特征提取的方式,模型被分为两个部分,由手工设计的特征提取器提取特征,并进一步通过特定的分类算法进行分类。然而,所提取到的特征严重依赖于人们的先验知识,因此性能局限于所提取特征的质量。随着伪造图像的质量越来越高,手工设计特征提取方式变得越来越困难。近年来,深度学习技术迅猛发展,许多基于深度学习的CG图

像检测方法<sup>[7-14]</sup>被提出,在性能上相较传统手工特征提取方法有明显的提升。



(a) CG图像



(b) 真实自然图像

图1 CG图像和真实自然图像

Figure 1 CG image and real natural image

然而,现有的基于深度学习的方法仍存在一些不足。现有的许多方法<sup>[8,12,14]</sup>表明,CG图像虽然在图像内容等语义特征上与真实图像十分相似,但在残差纹理特征等非语义特征中存在许多与真实图像不一致的地方。传统的卷积神经网络(convolutional neural network, CNN)一方面受限卷积层的特性,因而无法显式地对图像中的梯度信息进行提取,限制了提取残差纹理特征能力;另一方面,卷积层较小的感受野使网络无法从全局角度去提取特征,一定程度上也影响了性能。针对CG图像的特点和现有CG图像检测方法存在的不足,本文提出了一种基于跨流注意力增强

中心差分卷积网络的 CG 图像检测方法, 所提方法使用 RGB 流和 SRM (spatial rich model)<sup>[15]</sup>流分别提取图像的语义特征和残差纹理特征。特别地, 在每个流中, 本文使用中心差分卷积<sup>[16]</sup>替换了传统的卷积层, 以便同时提取像素强度信息和像素梯度信息。此外, 本文还设计了一个跨流注意力增强模块, 从全局角度提升特征提取能力, 并对两个流的特征进行跨流增强。在 3 个 CG 图像检测数据集上的检测结果表明, 所提方法优于其他现有方法, 且具备良好的泛化性能。综上所述, 本文的主要贡献包括以下方面。

1) 提出了跨流注意力增强中心差分卷积网络, 该网络采用了双流结构和跨流注意力增强模块, 能够很好地提取图像中的语义特征与非语义的残差纹理特征。

2) 将中心差分卷积引入 CG 图像检测任务中, 利用中心差分卷积的特性, 同时提取图像中的像素强度信息和像素梯度信息, 提升模型的 CG 图像检测性能。

3) 在 3 个公开的 CG 图像检测数据集上进行了实验。相较于目前现有的 CG 图像检测方法, 本文所提方法在 3 个数据集上均能取得最好的检测效果, 且具备较好的泛化性能。此外, 本文还进行了消融实验验证了本文模型中各模块的合理性。

## 1 相关工作

CG 检测方法主要可以分为两个类别: 基于手工特征提取的传统方法和基于深度学习技术的方法。

### 1.1 基于手工特征提取的传统方法

早期的 CG 图像检测方法主要基于手工特征提取的方式实现。例如, Ng 等<sup>[1]</sup>利用分形和微分几何相关的特征来实现 CG 图像检测。Lyu 等<sup>[2]</sup>提出的方法利用了一阶和高阶小波统计特征, 并使用一个支持向量机 (support vector machine, SVM)<sup>[17]</sup>作为分类器。Wang<sup>[3]</sup>使用了小波滤波后数据的直方图生成特征, 取得了较好的成果。Li 等<sup>[4]</sup>基于二阶差分统计量来检测 CG 图像。Wu 等<sup>[5]</sup>从 HSV 颜色空间提取特征, 并使用 Fisher 线性判别 (Fisher linear discriminant, FLD)<sup>[18]</sup>进行分类。Fan 等<sup>[6]</sup>认为图像边

缘信息对于 CG 图像取证任务具有很好的指导意义。

### 1.2 基于深度学习技术的方法

得益于深度学习技术的发展, 许多基于深度学习网络的 CG 图像检测方法被提出, 取得了更好的 CG 图像检测性能。例如, Quan 等<sup>[7]</sup>利用一个参数可学习的  $7 \times 7$  卷积层作为模型的输入层, 以提取图像中的高频信息, 提升模型的性能。Yao 等<sup>[8]</sup>在模型中使用了三组高通滤波器, 来增强模型感知图像中高频信息的能力。He 等<sup>[9]</sup>提出了一个双流的 CNN 来提取图像特征, 并结合了循环神经网络 (recurrent neural network, RNN) 来增强模型性能。He 等<sup>[10]</sup>设计了一个双流网络, 并使用基于注意力机制的特征融合来实现 CG 图像的检测。Zhang 等<sup>[11]</sup>在模型中设计了一个混合相关模块, 结合了图像的通道和像素相关性, 提升检测性能。Quan 等<sup>[12]</sup>在模型中使用了空域富模型滤波器<sup>[15]</sup>从图像的三通道分别进行特征提取, 并重点关注了模型的泛化能力, 探讨了模型针对未知数据的检测能力。Yao 等<sup>[13]</sup>提出的 CGNet 使用迁移学习的方式, 将预训练的 VGG 作为特征提取器, 并使用注意力机制融合不同层的特征。Bai 等<sup>[14]</sup>构建了一个大型数据集 LSCGB, 并提出了一个纹理感知网络 (texture-aware network) 来检测 CG 图像和自然图像间的纹理差异。

## 2 本文方法

本文所提的跨流注意力增强中心差分卷积网络的模型结构如图 2 所示, 模型中模块的具体结构如图 3 所示, 其中  $\oplus$  表示按像素相加,  $\ominus$  表示按像素相减,  $\otimes$  表示按通道拼接, NI 表示自然图像, CDC 表示中心差分卷积,  $C$  表示输出通道数,  $s$  表示步长。整个模型可以分为 4 个部分, 分别是 RGB 流、SRM 流、跨流注意力增强模块、分类模块。本文模型首先使用 RGB 流和 SRM 流分别提取图像的语义特征和残差纹理特征, 然后使用跨流注意力增强模块增强两个流的特征, 最后使用分类模块融合两个流的特征并进行分类。下面将对模型的各个组成部分进行详细介绍。

### 2.1 中心差分卷积

现有的深度学习模型大多使用卷积层作为特

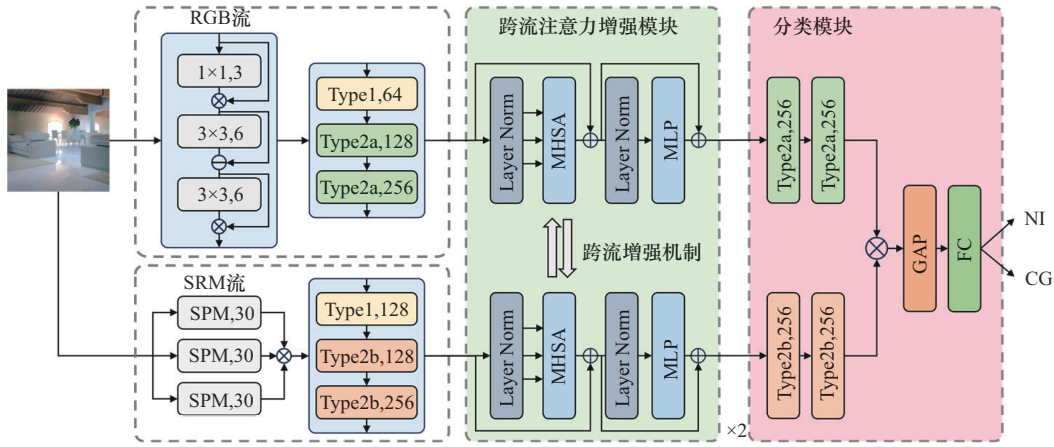


图2 模型结构

Figure 2 The structure of the model

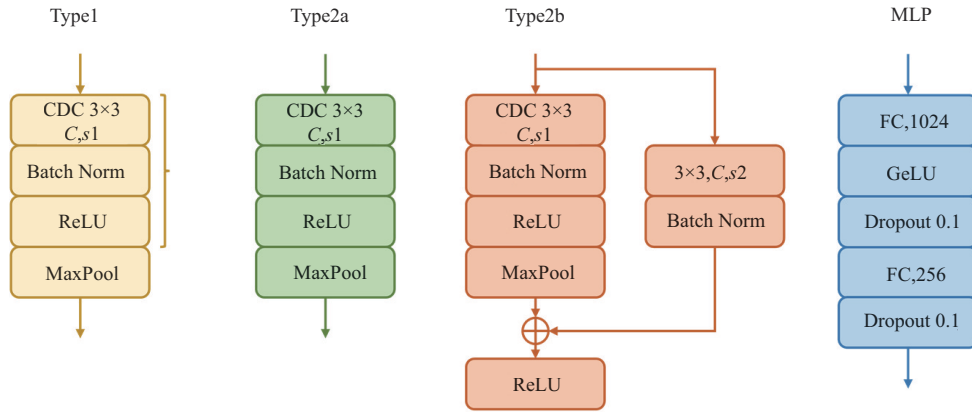


图3 模型中模块的具体结构

Figure 3 The structure of module in the model

征提取的手段。通过堆叠多个卷积层，网络可以很好地从图像中提取具体的语义特征。然而，卷积层更加擅长提取图像中的像素强度特征，而不能显式地提取图像中的梯度信息。由于图像的梯度信息能够很好地表征图像的纹理区域特征，因此普通的卷积层在特征提取时容易忽略这些区域的特征信息，对非语义信息的提取能力较弱。受到现有工作<sup>[16,19]</sup>的启发，本文在模型中引入了中心差分卷积（central difference convolution, CDC），同时提取图像中的像素强度特征和像素梯度特征，以提升模型性能。

普通的卷积层主要用于提取像素的强度信息，其计算可以分为两步。① 采样：从特征图中采样一块与卷积核尺寸相同的区域。② 聚合：对该区域内的像素值进行加权求和。其计算式可以用式(1)表示：

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n) \quad (1)$$

其中， $x$ 表示输入特征图， $y$ 表示输出特征图， $w$ 表示卷积核的参数， $p_0$ 表示当前像素点， $\mathcal{R}$ 表示感受野的区域 $P_n$ 。与普通卷积相比，中心差分卷积将差分操作和卷积操作相结合，从而将图像的梯度信息显式地引入卷积操作中，增强了对像素梯度信息的提取能力。中心差分卷积的示意如图4所示，与普通卷积的不同之处在于，中心差分卷积聚合的是中心区域的导向梯度。中心差分卷积的计算式可用式(2)表示，相当于将感受野内的所有像素都减去中心像素值后，再进行卷积操作。

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot (x(p_0 + p_n) - x(p_0)) \quad (2)$$

在具体实现上，中心差分卷积通常与普通卷积进行结合，以同时提取图像中的像素强度信息

和像素梯度信息。具体地,该广义的中心差分卷积的计算公式可用式(3)表示:

$$y(p_0) = \theta \cdot \sum_{p_n \in \mathcal{R}} w(p_n) \cdot (x(p_0 + p_n) - x(p_0)) + (1 - \theta) \cdot \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n) \quad (3)$$

其中,超参数  $\theta \in [0,1]$  表示像素强度信息和像素梯度信息之间的比例,  $\theta$  越大,像素梯度信息的重要性就越大。根据 Yu 等<sup>[16]</sup>的工作,本文设置超参数  $\theta = 0.7$ 。对式(3)进行化简后,可以得到式(4):

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n) - \theta \cdot x(p_0) \cdot \sum_{p_n \in \mathcal{R}} w(p_n) \quad (4)$$

在后文中,本文所述的中心差分卷积均指由式(4)所表示的广义中心差分卷积。

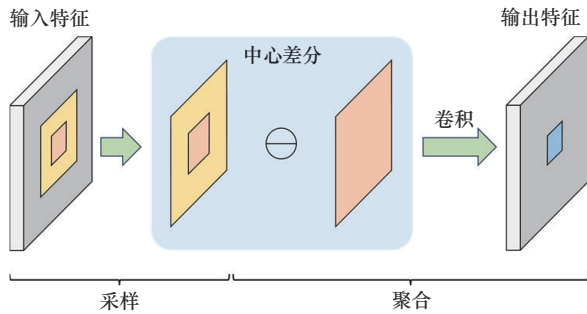


图4 中心差分卷积示意  
Figure 4 Central difference convolution

## 2.2 双流特征提取

与其他的图像分类任务类似,用于CG图像检测的模型需要从图像中提取有利于区分CG图像和真实图像的特征。而图像中所包含的特征可以分为两类:一类是与图像内容相关的语义特征,另一类是图像中的纹理细节等非语义特征。由于普通的单流网络通常难以同时关注两类特征,本文参考了一些现有方法<sup>[9-10,20]</sup>的思路,使用了一个双流结构来提取更加丰富的图像特征,以提升模型的性能。如图2所示,所提方法包含一个RGB流和SRM流,两个流分别以原图和残差纹理图像作为输入,以提取图像中的语义内容信息和残差纹理信息。在两个流中,本文引入了中心差分卷积,在提取语义特征的同时,进一步强化对纹理特征的提取能力。

### 2.2.1 RGB 流

为了提取图像中的语义内容信息,本文模型中设计了一个RGB流。如图2所示,所提方法的RGB流由一个预处理模块、一个Type1模块和两个Type2a模块组成。在RGB流中首先使用了预处理模块对输入图像进行预处理,以得到特征更为丰富的特征图。预处理模块由3个普通卷积层组成,首先一个 $1 \times 1$ 卷积对输入的3通道图像进行重组,得到3通道特征图,并与输入图像拼接后得到6通道的特征图 $F$ ;紧接着,第一个 $3 \times 3$ 卷积对特征图进行进一步的处理,得到6通道的输出特征图,并用特征图 $F$ 减去输出特征图得到 $\tilde{F}$ ;最后,使用第二个 $3 \times 3$ 卷积层对 $\tilde{F}$ 进行处理,得到6通道的输出特征图,并与 $\tilde{F}$ 按通道拼接,得到12通道的特征图 $\hat{F}$ ,作为预处理模块的输出。

在预处理模块之后,本文接续使用了一个Type1模块和两个Type2a模块对特征图进行特征提取和下采样操作。Type1模块和Type2a模块的结构定义如下。

**Type1 模块:** 该模块包含两个由中心差分卷积、批归一化层(batch normalization, BN)和ReLU层组成的中心差分卷积层组,并在最后一层使用一个大小为 $3 \times 3$ 、步长为2的最大池化层进行下采样。需要注意的是,第二个中心差分卷积组的输入通道数等于模块的输出通道。

**Type2a 模块:** 该模块结构与Type1模块类似,区别在于其仅包含一个中心差分卷积层组,最后一层则同样使用一个大小为 $3 \times 3$ 、步长为2的最大池化层进行下采样。

在RGB流中,Type1模块的输入/输出通道数为12/64,两个Type2a模块的输入/输出通道数分别为64/128、128/256。特征图在经过一个Type1模块和两个Type2a模块后,完成了3次下采样操作,特征图的长宽变为输入图像的1/8,通道数由12变为256。在完成RGB流的特征提取后,得到的特征图将会作为跨流注意力增强模块的其中一个输入。

### 2.2.2 SRM 流

目前CG图像的质量已经相当逼真,在色彩

等方面都能够达到以假乱真的程度。但在具体的纹理细节、物体边缘等区域仍容易存在伪影等不自然的特征。图像的高频信息由于包含了较多的纹理信息，因此往往被用于检测CG图像。受到先前的CG图像检测方法<sup>[12]</sup>与隐写分析方法<sup>[19,21-22]</sup>的启发，本文使用SRM<sup>[15]</sup>中的30个线性高通滤波核输入对输入的RGB图像的3个通道分别进行滤波，然后使用截断函数TLU<sup>[23]</sup>限制残差图像的数值范围，并最终得到90个残差图像，用于后续的特征提取。这30个滤波核具体包括一阶滤波核、二阶滤波核、三阶滤波核、EDGE 3×3滤波核、SQUARE 3×3滤波核、EDGE 5×5滤波核和SQUARE 5×5滤波核这7种滤波核以及它们的旋转变体。图5展示了使用SRM滤波核对比图5(a)进行滤波的结果。可以看到，这些残差图像之间存在一定的差异，但其保留的残差信息主要都存在于原始图像的纹理区域，因此更有利于网络学习CG图像与真实自然图像之间在纹理区域的不一致性。

在使用SRM滤波核提取了残差图像后，本文使用一个Type1模块和两个Type2b模块来进行进一步的特征提取。其中，Type1模块与RGB流中所使用的一致。Type2b模块类似于Type2a模块，不同之处在于，Type2b模块采用了一个残差连接，形成了双分支结构，具体结构如图3所示。Type2b模块的主分支与Type1a模块相同，包含一个由中心差分卷积、批归一化层和ReLU层组成的中心差分卷积层组，并使用一个大小为3×3、步长为2的最大池化层进行下采样；残差连接分支则包含一个步长为2的3×3普通卷积和一个批归一化层。两个分支的特征按像素相加后，再通过一个ReLU层，得到最后的输出。

在SRM流中，Type1模块的输入/输出通道数为90/128，两个Type2b模块的输入/输出通道数分别为128/128、128/256。SRM残差图像在经过一个Type1模块和两个Type2b模块后，完成3次下采样操作，特征图的长宽变为输入图像的1/8，通道数由90变为256。在完成SRM流的特征提取后，得到的特征图将会作为跨流注意力增强模块的其中一个输入。

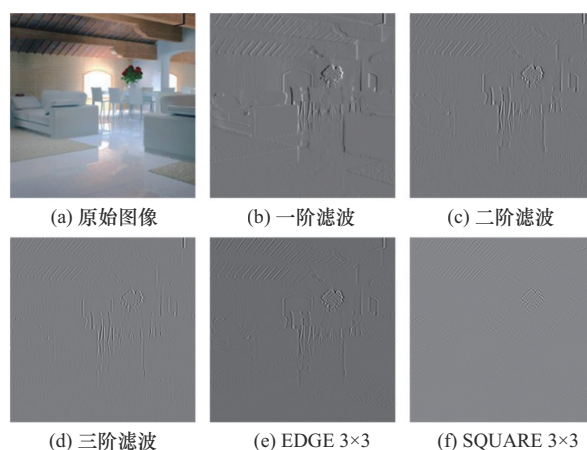


图5 使用SRM滤波后的残差图像

Figure 5 Residual image after SRM filtering

## 2.3 跨流注意力增强模块

2021年，Dosovitskiy等<sup>[24]</sup>提出的ViT（vision transformer）将Transformer<sup>[25]</sup>引入计算机视觉领域。Transformer相较传统的卷积神经网络，采用了多头自注意力机制，并可以从全局的角度去提取特征，以获得更好的表现。鉴于此，本文在模型中引入了Transformer结构，设计了一个跨流注意力增强模块。如图2所示，跨流注意力增强模块包含两个流，两个流结构相似，均由层归一化（layer normalization, LN）层、多头自注意力（multi-head self-attention, MHSA）模块和多层感知（multi-layer perception, MLP）机组成。在两个流的多头自注意力模块之间，本文设计了一个跨流注意力增强机制，使两个流的特征可以相互增强。在将特征图输入模块前，还需要将输入的维度进行调整，以适应模块的结构。具体地，对于两个流的输入 $X$ 和 $Y$ ，需要将维度由 $[b, c, h, w]$ 调整为 $[b, h \times w, c]$ ，其中， $b$ 表示批大小（batch size）， $c$ 表示通道数， $h$ 和 $w$ 表示特征图的高和宽。在每一个流中，首先使用多头自注意力模块进行特征提取，这个过程可用式(5)表示：

$$\hat{T} = \text{MHSA}(\text{LN}(T)) + T \quad (5)$$

其中， $T$ 表示其中一个流的输入。 $T$ 首先通过层归一化层对特征在通道维度进行归一化，然后通过多头自注意力模块进行特征提取，最后通过一个残差连接，将多头自注意力模块的输出与输入 $T$ 相加得到输出 $\hat{T}$ 。多头自注意力模块是基于自注

意力 (self-attention, SA) 机制的, 经典的自注意力计算式可用式(6)表示:

$$SA = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (6)$$

其中,  $Q$ 、 $K$ 、 $V$ 是通过输入特征进行可学习的线性变换得到的,  $d_k$ 表示向量的长度。自注意力机制计算时, 首先使用 $Q$ 和 $K$ 进行点积, 然后通过一个 $\text{softmax}$ 函数进行归一化计算出自注意力权重矩阵, 最后将自注意力权重矩阵同 $V$ 进行矩阵乘法, 得到最终的输出。对于多头自注意力, 在计算时会根据设定的自注意力头的数量 $n$ , 将 $Q$ 、 $K$ 、 $V$ 分别分割成 $n$ 份, 然后分别计算自注意力, 最后将各个头的输出拼接起来作为最终的输出。在本文模型中,  $n$ 设置为8。

为了使两个流的特征之间可以相互增强, 相互融合, 本文对多头自注意力模块的自注意力权重矩阵计算过程进行了调整。如图6所示, 本文设计了一个跨流增强机制。具体地, 令  $\text{dots}_i = \frac{Q_i K_i^T}{\sqrt{d_k}}, i \in \{1, 2\}$ , 本文通过式(7)对两个流的权重矩阵进行增强:

$$\begin{aligned} \text{dots}'_1 &= \text{dots}_1 + \sigma_1 \times \text{dots}_2 \\ \text{dots}'_2 &= \text{dots}_2 + \sigma_2 \times \text{dots}_1 \end{aligned} \quad (7)$$

其中,  $\sigma_i$ 为可学习参数, 并初始化为1, 在训练阶段会自适应地调整自身的值, 以自适应地融合另一个流的自注意力权重矩阵, 得到最终的权重矩阵  $\text{dots}'_i$ 。通过跨流增强的方式, 使两个流的特征之间相互关联, 更有利于特征提取。

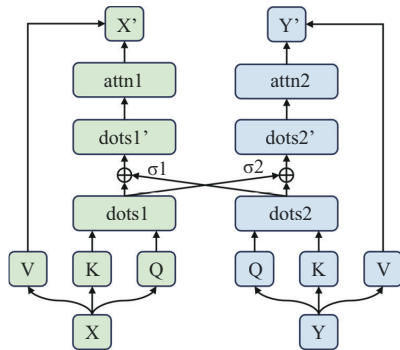


图6 跨流增强机制

Figure 6 Cross-stream enhancement mechanism

在多头自注意力模块后边, 紧接着一个层归一化层和一个多层感知机。多层感知机由全连接

层、GeLU激活函数和Dropout层组成, 其具体结构如图3所示。与多头自注意力模块的计算类似, 其计算式如式(8)所示。输入 $\tilde{T}$ 首先通过层归一化层对特征在通道维度进行归一化, 然后通过多层感知机进行特征提取, 最后通过一个残差连接, 将多层感知机的输出与输入 $\tilde{T}$ 相加得到一个流最终的输出 $\hat{T}$ 。

$$\hat{T} = \text{MLP}(\text{LN}(\tilde{T})) + \tilde{T} \quad (8)$$

在模型中, 本文重复使用了两次跨流增强模块来进行特征增强。在将输出特征输入后续模块之前, 还需要将两个流的特征维度由 $[b, h \times w, c]$ 转换回 $[b, c, h, w]$ , 以符合后续模块的输入形式。

## 2.4 分类模块

在使用跨流注意力增强模块对两个流的特征进行增强后, 本文使用分类模块对两个流的特征进行融合, 并进行最后的分类任务。如图2所示, 本文首先在RGB流和SRM流分别使用两个Type2a模块和两个Type2b模块进行下采样和特征提取; 随后将两个流的特征在通道维度进行拼接; 最后使用一个全局平均池化层和一个全连接层得到最终的分输出。

## 2.5 损失函数

CG图像检测任务可以视为一个二分类任务, 与现有的二分类网络训练方式类似, 本文所提方法使用了二值交叉熵 (binary cross entropy, BCE) 作为损失函数, 损失函数 $L$ 的表达式如式(9)所示。其中,  $y$ 表示真实的标签,  $\hat{y}$ 表示模型相应的输出。

$$L = -(y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})) \quad (9)$$

## 3 实验结果及分析

### 3.1 数据集设置

本文在3个公开的数据集上进行了实验, 这3个数据集分别是LSCGB<sup>[14]</sup>、SPL2018<sup>[9]</sup>和DsTok<sup>[24]</sup>。3个数据集的简要介绍如表1所示。

LSCGB数据集<sup>[14]</sup>由Bai等于2021年发布, 是一个大规模的CG图像检测数据集, 包括71 168幅CG图像和71 168幅真实自然图像, 共142 336幅图像。其中, CG图像的来源包括3D游戏、3D电影、现有的CG检测数据集以及GAN生成图

像, 真实自然图像的来源包括公开的图像检测数据集、CG检测数据集、不含后期特效的电影、图片网站中获取。对于LSCGB数据集的数据划分, 本文沿用了Bai等提供的数据划分方式, 其中, 训练集包含99 638幅图像, 验证集包含14 235幅图像, 测试集包含28 463幅图像。

SPL2018数据集<sup>[9]</sup>由He等于2018年发布, 包括6 800幅CG图像和6 800幅真实自然图像, 共计13 600幅图像。其中, CG图像的来源包括网络收集和通过超过50种渲染软件中生成, 真实自然图像则来自不同种类的相机拍摄的图像, 数据集中图像的分辨率从266×199至2 048×3 200。对于SPL2018数据集的数据划分, 本文按照10:3:4的比例将数据集随机分成训练集、验证集和测试集。

DsTok数据集<sup>[26]</sup>由Tokuda等于2013年发布, 包括4 850幅CG图像和4 850幅真实自然图像, 共计9 700幅图像。数据集中所有的图像均从网络上进行收集。对于DsTok数据集的数据划分, 本文按照3:1:1的比例将数据集随机分成训练集、验证集和测试集。

表1 3个公开数据集的简要介绍

Table 1 Brief introductions of three public datasets

| 数据集     | CG<br>图像/幅 | 真实自然<br>图像/幅 | 分辨率                     | 数据划分   |
|---------|------------|--------------|-------------------------|--------|
| LSCGB   | 71 168     | 71 168       | 75×75~<br>7 680×4 230   | 7:1:2  |
| SPL2018 | 6 800      | 6 800        | 266×199~<br>2 048×3 200 | 10:3:4 |
| DsTok   | 4 850      | 4 850        | 512×512~<br>4 200×3000  | 3:1:1  |

由于各数据集中图像分辨率存在较大的差异, 为了减少由于图像尺度变化所带来的影响, 对于所有的数据集, 本文首先将图像的尺寸使用双线性插值方式进行等比例缩放, 使图像的短边长度等于512。紧接着, 在训练和测试时, 分别使用随机裁剪和中心裁剪的方式从图像中裁剪出224×224的图像块作为网络输入。

### 3.2 评价指标

为了验证所提方法的有效性, 本文使用了ACC、TPR和TNR这3个评价指标来评估模型的性能。其中, ACC是准确率, TPR是正样本准

确率, TNR是负样本准确率, 3个评价指标的计算式如下。

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

$$TPR = \frac{TP}{TP + FN} \quad (11)$$

$$TNR = \frac{TN}{FP + TN} \quad (12)$$

其中, TP (TN) 表示正样本 (负样本) 被模型正确分类的数量; 而FN (FP) 表示正样本 (负样本) 被模型错误分类的数量。具体地, 本文将真实自然图像作为正样本, CG图像作为负样本。

### 3.3 模型实现细节

本文在3个现有的CG图像检测数据集上进行实验, 3个数据集使用了完全相同的实验设置。本文所提方法使用PyTorch来实现, 并使用4张NVIDIA TITAN Xp显卡进行训练。在训练策略上, 本文使用Adam优化器, 初始学习率设置为0.000 2, 训练BatchSize设置为32, Epoch设置为120。在训练过程中, 本文使用余弦退火策略对学习率进行衰减, 其中超参数设置:  $T_{\max} = 120$ ,  $\eta_{\min} = 1 \times 10^{-7}$ 。

### 3.4 对比模型

为了验证所提方法的有效性, 本文选取了8个专门用于CG图像检测的方法进行实验对比, 具体包括ScNet<sup>[7]</sup>、Yao<sup>[8]</sup>、SPL2018<sup>[9]</sup>、He<sup>[10]</sup>、Zhang<sup>[11]</sup>、QuanNet<sup>[12]</sup>、CGNet<sup>[13]</sup>和TextureNet<sup>[14]</sup>。为了保证对比的公平性, 所有对比方法所使用的图像大小、数据划分等均与所提方法保持一致。

### 3.5 公开数据集检测结果

为了验证所提方法在CG图像检测任务上的效果, 本文在3个数据集进行了实验对比, 实验结果如表2所示。所提方法首先在数据量最大的LSCGB数据集上对各方法的性能进行了评估。可以看到, 8个对比方法中有3个方法的正确率超过了90%, 其中QuanNet的表现最佳。而与现有方法相比, 所提方法在3个评价指标上均取得了最好的实验结果, ACC、TPR、TNR这3个指标分别达到了98.33%、98.57%、98.09%, 与在LSCGB数据集上表现第二佳的QuanNet相比, 所提方法在3个指标上分别提升了1.43%、1.57%和1.29%。由上述分析表明, 所提方法在LSCGB数据集上能够取得非常高的性能, 与其他现有方

表2 公开数据集实验结果

Table 2 The experimental results on public datasets

| 方法                         | LSCGB         |               |               | SPL2018       |               |               | DsTok         |               |               |
|----------------------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                            | ACC           | TPR           | TNR           | ACC           | TPR           | TNR           | ACC           | TPR           | TNR           |
| ScNet <sup>[7]</sup>       | 87.48%        | 88.72%        | 86.24%        | 91.78%        | 92.25%        | 91.31%        | 84.07%        | 83.92%        | 84.23%        |
| Yao <sup>[8]</sup>         | 86.44%        | 88.85%        | 86.06%        | 91.56%        | 87.06%        | 96.06%        | 81.75%        | 78.25%        | 85.26%        |
| SPL2018 <sup>[9]</sup>     | 91.00%        | 91.03%        | 90.98%        | 94.09%        | 91.50%        | 96.69%        | 86.13%        | 87.01%        | 85.26%        |
| He <sup>[10]</sup>         | 85.95%        | 84.11%        | 87.78%        | 88.66%        | 97.88%        | 89.44%        | 81.13%        | 72.89%        | 89.38%        |
| Zhang <sup>[11]</sup>      | 95.19%        | 95.56%        | 94.82%        | 93.44%        | 93.63%        | 93.25%        | 90.72%        | 90.52%        | 90.93%        |
| QuanNet <sup>[12]</sup>    | 96.90%        | 97.00%        | 96.80%        | 95.56%        | 95.19%        | 95.94%        | 92.78%        | 92.06%        | 93.51%        |
| CGNet <sup>[13]</sup>      | 93.83%        | 94.55%        | 93.10%        | 95.59%        | 95.94%        | 95.25%        | 94.79%        | 93.81%        | 95.77%        |
| TextureNet <sup>[14]</sup> | 94.37%        | 94.58%        | 94.16%        | 93.22%        | 93.06%        | 93.37%        | 95.05%        | 95.46%        | 94.64%        |
| 所提方法                       | <b>98.33%</b> | <b>98.57%</b> | <b>98.09%</b> | <b>96.94%</b> | <b>96.63%</b> | <b>97.25%</b> | <b>96.70%</b> | <b>96.49%</b> | <b>96.91%</b> |

法相比提升较大, 这表明了所提方法在数据量较大的数据集上的有效性。

为了进一步验证所提方法在数据量较小的数据集上的效果, 本文在 SPL2018 和 DsTok 这两个数据集上对各方法性能进行了评估。其中, 在 SPL2018 数据集上, 所提方法取得了最优的结果, ACC、TPR、TNR 这 3 个指标分别为 96.94%、96.63%、97.25%, 与表现第二佳的 CGNet 相比, 3 个指标分别提升了 1.35%、0.69%、2.00%。而在 DsTok 数据集上, 所提方法同样取得了最佳的实验结果, 3 个评价指标分别为 96.70%、96.49%、96.91%, 与第二佳的 TextureNet 相比, 3 个指标分别提升了 1.65%、1.03%、1.27%。在 SPL2018 和 DsTok 这两个数据量较小的数据集上的实验结果表明, 在训练数据相对较少的情况下, 所提方法同样能够取得比较佳的性能, 这表明所提方法能够很好地适应不同的数据集尺度, 并取得相当不错的 CG 图像检测性能。

### 3.6 跨数据集检测结果

由于在真实场景下, 可能无法获取 CG 图像的来源和种类, 因此需要 CG 检测方法具备一定的泛化性能, 以应对未知类型的 CG 图像。为了验证所提方法的泛化性能, 本文设计了一组跨数据集检测的实验。具体地, 本文首先将所提方法和其他现有方法在 LSCGB 数据集上进行训练, 然后在 SPL2018 和 DsTok 这两个数据集上进行测试, 比较在两个数据集上的 ACC 指标。具体的实验结果如表 3 所示, 可以看到, 由于不同数据集间数据分布存在一定的差异, 各方法在跨数据集检测

中都存在性能下降的情况, 但所提方法仍然取得了最佳的效果。具体来说, 所提方法在 DsTok 数据集上取得了最优的实验结果, ACC 达到了 98.56%, 相较第二名的 QuanNet 提升了 3.25%; 在 SPL2018 数据集上虽然未能取得最佳的效果, 但与最优的方法 ScNet 差距不大, 且 ScNet 在 DsTok 数据集上的表现并不优秀。而在两个数据集上的平均表现上, 所提方法相较其他方法也具有较大的提升, ACC 达到了 88.97%, 与第二名的 TextureNet 相比, 提升了 1.65%。由上述实验结果表明, 所提方法不仅在数据集内能够取得优异的实验结果, 在跨数据集检测的能力上也具有不错的表现, 这表明所提方法具有很好的泛化性。

表3 跨数据集实验结果

Table 3 The experimental Cross-dataset results

| 方法                         | ACC           |               |               |
|----------------------------|---------------|---------------|---------------|
|                            | SPL2018       | DsTok         | Avg           |
| ScNet <sup>[7]</sup>       | <b>80.78%</b> | 77.89%        | 79.34%        |
| Yao <sup>[8]</sup>         | 64.94%        | 72.11%        | 68.53%        |
| SPL2018 <sup>[9]</sup>     | 55.44%        | 84.95%        | 70.20%        |
| He <sup>[10]</sup>         | 75.16%        | 71.34%        | 73.25%        |
| Zhang <sup>[11]</sup>      | 40.00%        | 90.77%        | 65.39%        |
| QuanNet <sup>[12]</sup>    | 64.62%        | 95.31%        | 79.97%        |
| CGNet <sup>[13]</sup>      | 80.44%        | 93.04%        | 86.74%        |
| TextureNet <sup>[14]</sup> | 80.00%        | 94.64%        | 87.32%        |
| 所提方法                       | 79.37%        | <b>98.56%</b> | <b>88.97%</b> |

### 3.7 计算复杂度对比

由于模型的计算复杂度会影响模型部署时的

资源占用，本文对各方法的计算复杂度进行了对比，如表4所示。其中MAC为乘加累积操作数（multiply-accumulate operation），FLOP为浮点运算次数（floating point operation），Param为参数量。可以看到，几个较早提出的方法<sup>[7-11]</sup>在浮点运算量和参数量上相对较少，但整体的模型性能与较新的方法<sup>[12-14]</sup>相比存在较大的差距；而所提方法与几个较新的方法相比，在浮点运算量上并没有明显提升，但在参数量上更少。这说明所提方法通过有效的模型设计，在不明显增加计算负担的情况下提升了CG图像检测性能。

表4 计算复杂度对比

Table 4 Computational Complexity Comparison

| 方法                         | MAC(G) | FLOP(G) | Param(M) |
|----------------------------|--------|---------|----------|
| ScNet <sup>[7]</sup>       | 1.59   | 3.17    | 43.22    |
| Yao <sup>[8]</sup>         | 0.11   | 0.22    | 0.04     |
| SPL2018 <sup>[9]</sup>     | 1.62   | 3.24    | 0.70     |
| He <sup>[10]</sup>         | 3.70   | 7.41    | 0.78     |
| Zhang <sup>[11]</sup>      | 0.71   | 1.42    | 0.06     |
| QuanNet <sup>[12]</sup>    | 16.54  | 33.08   | 68.90    |
| CGNet <sup>[13]</sup>      | 13.89  | 27.78   | 10.18    |
| TextureNet <sup>[14]</sup> | 21.10  | 42.20   | 21.49    |
| 所提方法                       | 23.99  | 47.97   | 7.99     |

### 3.8 消融实验

为了验证所提方法中各组成部分的有效性，本文设计了一系列的消融实验，消融实验在LSCGB、SPL2018、DsTok这3个数据集上进行，实验设置与前文保持一致。

#### 3.8.1 中心差分卷积的有效性

本文所提方法在RGB流、SRM流和分类模

块中均使用了中心差分卷积来代替传统的卷积层。为了验证模型中所使用的中心差分卷积的有效性，本文设计了一组针对中心差分卷积的消融实验，将模型中的中心差分卷积替换为普通卷积，实验结果如表5中第一行（w/o CDC）所示。可以看到，使用中心差分卷积的模型在性能上优于不使用中心差分卷积的模型。在3个数据集上，使用中心差分卷积能够使ACC分别提升0.27%、0.63%、0.26%，使TPR分别提升0.55%、1.32%、0.61%，但在TNR指标上，则存在小幅的性能下降。由于真实图像往往具有更加真实和复杂的纹理和噪声特征，因此中心差分卷积能够很好地帮助模型提升检测真实图像的能力；而CG图像纹理更平滑、噪声更少，会导致检测性能小幅下降。但从总体上而言，引入中心差分卷积是有利于CG图像检测任务的：对于RGB流，模型在提取中频、低频信息中语义特征的同时，还可以通过中心差分卷积的特性，同时从图像中自适应地提取纹理特征，并学习图像中语义特征与非语义特征之间的联系；对于SRM流，中心差分卷积则进一步强化了对纹理特征的提取能力，捕获更细微的纹理不一致性。

#### 3.8.2 跨流注意力增强模块的有效性

为了验证跨流注意力增强模块的有效性，本文设计了一组消融实验，将跨流注意力增强模块去除，实验结果如表5中第二行（w/o Attention）所示。可以看到，在添加了跨流注意力增强模块后，模型的性能在3个数据集上均存在不同程度的提升，在ACC指标上分别提升了0.41%、0.66%、0.15%，在TPR指标上分别提升了

表5 消融实验

Table 5 Ablation experiments

| Variants      | LSCGB         |               |               | SPL2018       |               |               | DsTok         |               |               |
|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
|               | ACC           | TPR           | TNR           | ACC           | TPR           | TNR           | ACC           | TPR           | TNR           |
| w/o CDC       | 98.06%        | 98.02%        | <b>98.10%</b> | 96.31%        | 95.31%        | <b>97.31%</b> | 96.44%        | 95.88%        | <b>97.01%</b> |
| w/o Attention | 97.92%        | 97.89%        | 97.95%        | 96.28%        | 95.56%        | 97.00%        | 96.55%        | <b>97.32%</b> | 95.77%        |
| w/ RGB流       | 96.84%        | 97.24%        | 96.44%        | 95.44%        | 93.87%        | 97.00%        | 94.48%        | 94.85%        | 94.12%        |
| w/ SRM流       | 98.10%        | 98.27%        | 97.93%        | 95.94%        | 94.75%        | 97.12%        | 95.98%        | 96.29%        | 95.67%        |
| 所提方法          | <b>98.33%</b> | <b>98.57%</b> | 98.09%        | <b>96.94%</b> | <b>96.63%</b> | 97.25%        | <b>96.70%</b> | 96.49%        | 96.91%        |

0.68%、1.07%、-0.83%，在 TNR 指标上分别提升了 0.14%、0.25%、1.14%。这表明跨流注意力增强模块能提升模型的特征提取能力，增强模型 CG 图像检测性能。

### 3.8.3 双流结构的有效性

为了验证双流结构的有效性，本文设计了两组消融实验，分别只保留了模型中的 RGB 流和 SRM 流，实验结果如表 5 中第 3 行和第 4 行所示。可以看到，单流结构的模型性能存在显著下降，与 RGB 单流的模型（w/ RGB 流）相比，所提方法在 LSCGB 数据集上 3 个指标分别提升 1.49%、1.33%、1.65%，在 SPL2018 数据集上分别提升 1.50%、2.76%、0.25%，在 DsTok 数据集上分别提升 2.22%、1.64%、2.79%。与 SRM 单流的模型（w/ SRM 流）相比，所提方法在 LSCGB 数据集上的 3 个指标分别提升 0.23%、0.30%、0.16%，在 SPL2018 数据集上分别提升 1.00%、1.88%、0.13%，在 DsTok 数据集上分别提升 0.72%、0.20%、1.24%。可以观察到，相较单流结构的模型，所提方法具有更好的表现。而去除中心差分卷积的模型和去除跨流注意力增强模块的模型由于同样保留了双流结构，在整体性能表现上优于单流结构的模型，这表明双流结构的模型由于同时结合了图像的语义特征和非语义的残差纹理特征，在 CG 图像检测上具备更优的性能。同时本文也观察到，SRM 单流的模型相较 RGB 单流的模型在 3 个数据集上具有更好的表现，这表明图像中的残差纹理特征对于 CG 图像

检测任务而言更加有效。

为了进一步验证双流结构对于 CG 图像检测的作用，本文分别使用一幅真实自然图像和一幅 CG 图像作为模型输入，并使用 Grad-CAM<sup>[27]</sup>对模型中的特征图进行了可视化，来分析双流结构对于模型性能的影响。特征图可视化对比如图 7 所示，从图 7 可以看到，RGB 流和 SRM 流的特征热力图之间存在明显的差异，而双流特征热力图则将 RGB 流和 SRM 流重点关注的区域进行了融合。这意味着通过双流结构的设计，模型可以从两个不同的角度去提取图像的特征，从而关注到图像中更多的区域，提取出更多有利于 CG 图像检测的图像特征，从而提供更好的性能。

## 4 结束语

为了解决 CG 图像所带来的安全隐患，本文提出了一个用于 CG 图像检测任务的跨流注意力增强中心差分卷积网络。首先使用 RGB 流和 SRM 流分别提取图像的语义特征和残差纹理特征，然后使用跨流注意力增强模块增强两个流的特征，最后使用分类模块融合两个流的特征并进行分类。在模型中，本文使用了中心差分卷积来替代普通的卷积层，从而使模型能够同时提取像素强度信息和像素梯度信息，更好地帮助模型从图像中提取出有利于 CG 图像检测的特征。在 3 个公开的 CG 图像检测数据集上的实验结果表明，本文所提方法能够在 3 个不同的数据集上取得最好的 CG 检测结果，证明了本文所

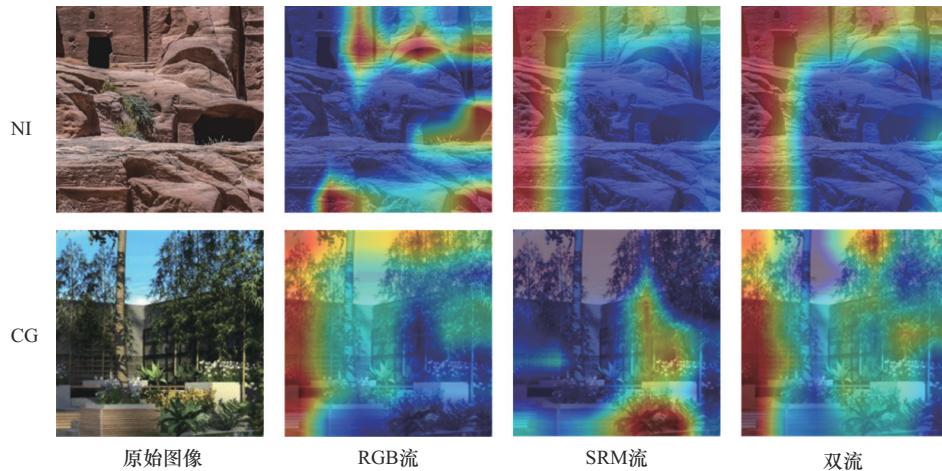


图7 特征图可视化对比

Figure 7 Feature map visualization comparison

提方法的有效性。同时, 本文还验证了所提方法的跨数据集检测能力, 实验结果表明, 所提方法同样能取得较好的结果, 具备很好的泛化性。

虽然所提方法已经取得了相当优异的性能, 但仍存在值得进一步深入研究的地方, 未来的工作将围绕以下3个方面进行展开。

1) 进一步提升模型的鲁棒性。在现实环境下, CG图像同样可能在遭受如JPEG压缩、放缩等后处理操作, 这些后处理操作会破坏图像中的特征, 增加模型对CG图像的识别难度。因此, 提升模型鲁棒性, 有利于提升模型的实用性。

2) 将所提方法拓展至CG视频检测任务。与图像相比, 视频检测更具挑战性, 将所提方法拓展至视频检测能增强模型泛用性。

3) 将所提方法拓展至AIGC图像检测任务。目前, DALL·E 2等图像生成模型的推出, 引起了人们的广泛关注, 如何对这些大模型生成的图像进行检测是重要的研究方向。

## 参考文献:

- [1] NG T T, CHANG S F, HSU J, et al. Physics-motivated features for distinguishing photographic images and computer graphics[C]//Proceedings of the 13th Annual ACM International Conference on Multimedia. New York: ACM, 2005: 239-248.
- [2] LYU S, FARID H. How realistic is photorealistic[J]. IEEE Transactions on Signal Processing, 2005, 53(2): 845-850.
- [3] WANG Y, MOULIN P. On discrimination between photorealistic and photographic images[C]//Proceedings of the 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings. Piscataway: IEEE Press, 2006.
- [4] LI W X, ZHANG T, ZHENG E G, et al. Identifying photorealistic computer graphics using second-order difference statistics[C]//Proceedings of the 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery. Piscataway: IEEE Press, 2010: 2316-2319.
- [5] WU R Y, LI X L, YANG B. Identifying computer generated graphics VIA histogram features[C]//Proceedings of the 2011 18th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2011: 1933-1936.
- [6] FAN S, WANG R, ZHANG Y, et al. Classifying computer generated graphics and natural images based on image contour information[J]. Journal of Information & Computational Science, 2012, 9(10): 2877-2895.
- [7] QUAN W Z, WANG K, YAN D M, et al. Distinguishing between natural and computer-generated images using convolutional neural networks[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(11): 2772-2787.
- [8] YAO Y, HU W T, ZHANG W, et al. Distinguishing computer-generated graphics from natural images based on sensor pattern noise and deep learning[J]. Sensors, 2018, 18(4): 1296.
- [9] HE P S, JIANG X H, SUN T F, et al. Computer graphics identification combining convolutional and recurrent neural networks[J]. IEEE Signal Processing Letters, 2018, 25(9): 1369-1373.
- [10] HE P S, LI H L, WANG H X, et al. Detection of computer graphics using attention-based dual-branch convolutional neural network from fused color components[J]. Sensors, 2020, 20(17): 4743.
- [11] ZHANG R S, QUAN W Z, FAN L B, et al. Distinguishing computer-generated images from natural images using channel and pixel correlation[J]. Journal of Computer Science and Technology, 2020, 35(3): 592-602.
- [12] QUAN W Z, WANG K, YAN D M, et al. Learn with diversity and from harder samples: Improving the generalization of CNN-based detection of computer-generated images[J]. Forensic Science International: Digital Investigation, 2020, 35: 301023.
- [13] YAO Y, ZHANG Z X, NI X, et al. CGNet: Detecting computer-generated images based on transfer learning with attention module [J]. Signal Processing: Image Communication, 2022, 105: 116692.
- [14] BAI W M, ZHANG Z P, LI B, et al. Robust texture-aware computer-generated image forensic: benchmark and algorithm[J]. IEEE Transactions on Image Processing, 2021, 30: 8439-8453.
- [15] FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images[J]. IEEE Transactions on Information Forensics and Security, 2012, 7(3): 868-882.
- [16] YU Z T, ZHAO C X, WANG Z Z, et al. Searching central difference convolutional networks for face anti-spoofing[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 5294-5304.
- [17] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [18] FISHER R A. The use of multiple measurements in taxonomic problems[J]. Annals of Eugenics, 1936, 7(2): 179-188.
- [19] 魏康康, 骆伟祺, 刘明林. 基于中心差分卷积和注意力的空域彩色图像隐写分析[J]. 软件学报, 2024, (12): 1-17.
- [20] 王金伟, 张子荷, 罗向阳, 等. 基于强化样本的伪孪生网络图像篡改定位模型[J]. 网络与信息安全学报, 2024, 10(1): 33-47.
- [21] WANG J W, ZHANG Z H, LUO X Y, et al. Pseudo-Siamese network image tampering localization model based on reinforced samples[J]. Chinese Journal of Network and Information Security, 2024, 10(1): 33-47.
- [22] DENG X Q, CHEN B L, LUO W Q, et al. Fast and effective global covariance pooling network for image steganalysis[C]//Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. New York: ACM, 2019: 230-234.
- [23] WEI K K, LUO W Q, TAN S Q, et al. Universal deep network for steganalysis of color image based on channel representation[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 3022-3036.
- [24] YE J, NI J Q, YI Y. Deep learning hierarchical representations for image steganalysis[J]. IEEE Transactions on Information Forensics

and Security, 2017, 12(11): 2545-2557.

- [24] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [EB]. arXiv Preprint arXiv: 2010.11929, 2020.
- [25] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [26] TOKUDA E, PEDRINI H, ROCHA A. Computer generated images vs. digital photographs: a synergetic feature and classifier combination approach[J]. Journal of Visual Communication and Image Representation, 2013, 24(8): 1276-1292.
- [27] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[J]. International Journal of Computer Vision, 2020, 128(2): 336-359.

#### [作者简介]



**黄锦坤** (1999-), 男, 广东揭阳人, 中山大学硕士生, 主要研究方向为图像伪造检测与定位。



**黄远航** (1997-), 男, 广东茂名, 中山大学博士生, 主要研究方向为数字图像取证。



**黄文敏** (1994-), 男, 广东湛江人, 中山大学博士生, 主要研究方向为人脸属性编辑和图像伪造检测与定位。



**骆伟祺** (1980-), 男, 广东韶关人, 博士, 中山大学教授、博士生导师, 主要研究方向为数字媒体伪造与检测、信息隐写与分析。