计算机科学
COMPUTER SCIENCE

位置增强与频域分量交互的深度伪造检测方法

孟思雨, 牛春翔, 谭荃戈, 王蓉

引用本文

孟思雨, 牛春翔, 谭荃戈, 王蓉. 位置增强与频域分量交互的深度伪造检测方法[J]. 计算机科学, 2026, 53(4): 445-453.

MENG Siyu, NIU Chunxiang, TAN Quange, WANG Rong. Deepfake Detection Method Based on Positional Enhancement and Frequency Domain ComponentInteraction [J]. Computer Science, 2026, 53(4): 445-453.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.250700070

位置增强与频域分量交互的深度伪造检测方法

孟思雨　牛春翔　谭荃戈　王　蓉
中国人民公安大学信息网络安全学院　北京 100038
(1983160587@qq.com)

摘　要　随着深度伪造技术的快速发展,伪造人脸图像和视频在社交媒体上频繁出现。然而,这些技术也被恶意利用,严重威胁社会安全。现有检测方法在已知数据集的伪造人脸检测中表现良好,但在面对未知数据集的伪造人脸时,检测效果却显著下降。针对这一问题,提出了一种位置增强与频域分量交互的深度伪造检测方法,旨在提高深度伪造人脸检测算法的鲁棒性及泛化性。首先,采用 Vision Transformer 作为骨干网络,从全局角度捕捉伪造痕迹;其次,设计动态局部特征提取模块,利用卷积进行逐通道逐点局部特征提取,并根据每个像素在特征表示中的重要性进行动态加权,精细化局部特征,提高对局部特征的感知能力;同时,构建多尺度特征提取与位置增强模块,采用多膨胀率卷积获取多尺度特征,引入位置增强机制强化像素间的位置信息关联,有效提取不同区域的多尺度信息;然后,设计全局-局部频域分量交互模块,通过频域分解注意力机制实现不同频域分量之间的信息交互,捕捉全局与局部特征之间的依赖关系,以获取在伪造人脸图像质量下降时 RGB 空间中消失的伪影;最后,设计像素关系相似度损失函数计算像素间的位置关系损失,并结合交叉熵损失函数构建联合损失函数,提高深度伪造人脸检测的准确性。实验结果表明,所提方法在 FF++和 Celeb-DF 数据集上的 AUC 指标分别达到 99.29%和 78.62%,其能有效提升深度伪造人脸检测算法的鲁棒性与泛化性。

关键词:特征提取;位置增强;频域分量交互;联合损失;深度伪造检测
中图分类号　TP391.41

Deepfake Detection Method Based on Positional Enhancement and Frequency Domain Component Interaction

MENG Siyu,NIU Chunxiang,TAN Quange and WANG Rong
College of Information and Cyber Security,People's Public Security University of China,Beijing 100038,China

Abstract　With the rapid development of Deepfake technology,forged facial images and videos generated by such techniques have become increasingly prevalent on social media platforms. However,these technologies are also being maliciously exploited,posing serious threats to social security. Although existing detection methods perform well in detecting Deepfake faces on in-domain datasets,their performance significantly degrades when applied to unseen datasets. To address this issue,a Deepfake detection method based on positional enhancement and frequency domain component interaction is proposed,aiming to improve the robustness and generalization of facial forgery detection. Firstly,vision Transformer is employed as the backbone network to capture forgery traces from a global perspective. Secondly,the dynamic local feature extraction module is designed,utilizing channel-wise and point-wise convolutional operations for local feature extraction. This module dynamically weights features based on pixel-level importance in feature representation,thereby refining local features and enhancing the ability to perceive local features. Concurrently,the multi-scale feature extraction and positional enhancement module is constructed,which acquires multi-scale features through multi-dilated convolutions and introduces a positional enhancement mechanism to strengthen positional correlations between pixels,effectively extracting multi-scale information from different regions. Then,the global-local frequency domain component interaction module is developed,implementing information exchange between different frequency components through the frequency domain decomposition attention mechanism. This captures dependencies between global and local features to identify artifacts that disappear in RGB space when fake facial image quality degrades. Finally,the pixel relationship similarity loss function is designed to calculate positional relationship losses between pixels and is combined with cross-entropy loss to construct the joint loss function to improve detection accuracy. Experimental results demonstrate that the proposed method achieves AUC

到稿日期:2025-07-14　返修日期:2025-09-04
基金项目:高等学校学科创新引智基地项目(B20087)
This work was supported by the Discipline Innovation and Talent Introduction Base Program for Institutions of Higher Education(B20087).
通信作者:王蓉(dbdxwangrong@163.com)

scores of 99.29% and 78.62% on FF++ and Celeb-DF datasets respectively, proving its effectiveness in enhancing the robustness and generalization of facial forgery detection.

Keywords Feature extraction, Positional enhancement, Frequency domain component interaction, Joint loss, Deepfake detection

1 引言

近年来，深度伪造人脸技术快速发展，其生成的图像和视频逼真度不断提升，在娱乐、影视和虚拟现实等领域展现出广阔的应用前景[1-2]。然而，该技术也存在被滥用的风险，可能导致虚假新闻传播、隐私侵犯以及诈骗等问题。因此，开发高效的深度伪造人脸检测算法具有重要的现实意义。

传统深度伪造人脸检测方法主要基于卷积神经网络(Convolutional Neural Networks, CNNs)，聚焦于局部特征的捕捉。Zhao 等[3]设计的多个空间注意力头和纹理特征增强模块，显著提升了对细微伪造特征的检测能力。Zhang 等[4]以 EfficientNet 为主干网络提取空间特征，结合可学习的空域富模型(Spatial Rich Model, SRM)滤波器捕获噪声特征，通过生成噪声注意力图并与空间特征融合，再利用通道注意力机制增强特征表示能力。Guo 等[5]通过 5 个深度可分离卷积层聚焦局部操纵线索，结合多层感知机与最大池化操作捕捉全局特征，将局部与全局特征连接后通过深度可分离卷积处理成混合特征，并利用全局特征生成注意力矩阵优化特征，最终完成检测任务。Zhang 等[6]通过融合多层次卷积特征，促使模型捕捉更全面的伪造线索，进而提升伪造图像鉴别的有效性。Wang 等[7]从人脸视频中提取精准的面部特征点，并通过长短期记忆网络捕获五官区域的不规则运动，最终实现了对压缩人脸视频真伪的有效鉴别。然而，CNN 受限于局部感受野，可能学习到特定归纳偏差，影响模型的泛化能力。近年来，Transformer[8]在建模长距离依赖关系方面优势显著，并在视觉任务中取得成功，如 Vision Transformer(ViT)[9]。一些研究开始尝试在 CNN 基础架构中嵌入 Transformer 层，通过融合两者的优势提升伪造人脸检测算法的性能。Zhou 等[10]将 EfficientNet 作为骨干网络，引入不同尺寸图像块的 ViT，以提升深度伪造检测算法的泛化能力。Khormali 等[11]采用基于 ViT 架构的特征提取器，通过自监督对比学习进行预训练，并结合图卷积网络与 Transformer 构建判别模块以识别伪造特征。此外，还有一些研究通过引入频率信息来增强特征表达能力。Zhou 等[12]从浅层特征中提取中高频信息以强化浅层伪造表征，缓解网络深层中高频线索丢失的问题，并从中层全局特征中捕获中高频信息以丰富伪造表征。Lai 等[13]通过扰动训练数据的振幅谱生成高频带多样性更强的增强图像，扩展高频特征的变化范围，同时引入伪造伪影一致性学习策略引导判别性特征学习。Zhang 等[14]通过隐写分析增强模型提取高频噪声特征，将高频噪声伪影融入空间纹理特征建模过程，并借助特征加权机制抑制背景噪声干扰。Huang 等[15]通过卷积神经网络学习图像空间域中的细微操纵痕迹，同时聚焦于图像频域中高频信息对应的操纵痕迹，以实现对图像操纵行为的精准检测。

与以往仅在 CNN 架构中嵌入少量 Transformer 层的研究不同，本文以 ViT 为主干网络，借助其全局建模特性实现对图像伪造痕迹的精准捕捉。Miao 等[16]采用纯 ViT 架构，通过特征袋方法对图像块间的关联关系进行编码，在无需额外掩码监督的条件下实现对伪造特征的有效学习。然而，单纯依赖纯 ViT 架构进行深度伪造检测任务还存在一定的局限性：1)Transformer 虽擅长全局信息建模，但可能忽略对检测至关重要的局部细节，如细微缺陷、边缘不平滑等；2)不同伪造算法生成的伪造区域大小各异，固定图像块大小的检测方法难以有效捕捉不同区域的伪造痕迹；3)随着伪造人脸图像质量的下降，伪影可能在 RGB 空间中消失，单靠多头注意力机制难以有效建模长距离依赖关系。

为解决上述问题，本文提出了一种位置增强与频域分量交互的深度伪造检测方法。首先，采用 Vision Transformer 作为骨干网络，从全局角度捕捉伪造痕迹；其次，设计动态局部特征提取模块，利用逐通道与逐点卷积提取局部特征，并根据每个像素在特征表示中的重要性进行动态加权，突出关键信息，增强对局部细节的捕捉能力；同时，构建多尺度特征提取与位置增强模块，利用不同膨胀率的卷积提取多尺度特征，引入位置增强机制强化像素间的位置关系，有效提取不同区域的多尺度信息；然后，提出全局-局部频域分量交互模块，通过频域分解注意力机制实现不同频域分量间的信息交互，更全面地捕捉图像中的全局与局部依赖关系，有效解决伪影在 RGB 空间中可能消失的问题；最后，设计像素关系相似度损失函数计算像素间的位置关系损失，并结合交叉熵损失函数构建联合损失函数，提高检测的准确性。本文的主要工作如下：

1)设计动态局部特征提取模块，利用卷积进行逐通道逐点局部特征提取，并根据每个像素在特征表示中的重要性赋予其动态权重，实现对局部细节的精确捕捉；

2)构建多尺度特征提取与位置增强模块，通过不同膨胀率的卷积提取多尺度特征，引入位置增强机制强化像素间的位置信息关联，有效提取不同大小区域的伪造痕迹；

3)设计全局-局部频域分量交互模块，通过频域分解注意力机制实现不同频域分量间的信息交互，有效捕捉全局与局部之间的依赖关系，提取在伪造人脸图像质量下降时 RGB 空间中消失的伪影；

4)设计像素关系相似度损失函数，计算像素间的位置关系损失，并结合交叉熵损失函数构建联合损失函数，提高检测的准确性。

2 本文方法

本文提出了一种基于位置增强与频域分量交互的深度伪造检测方法，其以 Vision Transformer 为骨干网络，主要由动态局部特征提取、多尺度特征提取与位置增强、全局-局部频域分量交互和联合损失函数等模块组成，如图 1 所示。首先，动态局部特征提取模块利用卷积实现逐通道逐点局部特征提取，并依据每个像素在特征表示中的重要性进行动态加权以

突出关键信息；同时，采用不同膨胀率的卷积提取多尺度特征，捕捉不同大小区域的伪影，并引入位置增强机制强化像素间的位置关系，以解决膨胀操作可能导致的像素关系被破坏问题；然后，将多尺度特征与动态局部特征融合，使生成的 Token 同时携带多尺度信息与局部信息。上述过程依次重复 3 次，并逐步减少特征图的空间维度。在此基础上，将提取的特征输入全局-局部频域分量交互模块，利用频域分解注意力机制实现不同频域分量的信息交互，建立全局与局部间的多种依赖关系，并将这些信息与动态局部特征提取模块输出的局部特征融合，以提取 RGB 空间中可能消失的伪造痕迹；最后，构建联合损失函数，进一步提升模型对人脸图像真实性的判断能力。

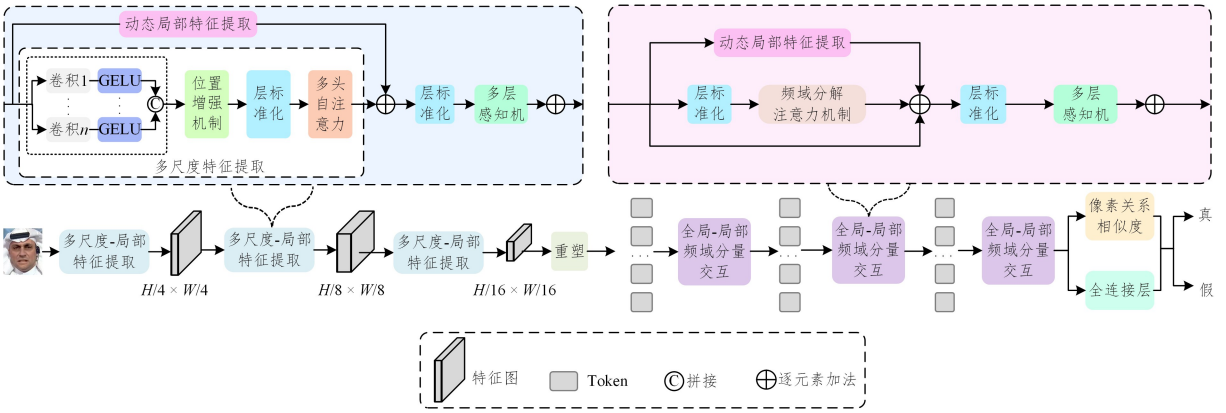

图 1　位置增强与频域分量交互的深度伪造检测方法的结构

Fig. 1　Structure of Deepfake detection method based on positional enhancement and frequency domain component interaction

2.1　动态局部特征提取

ViT 的自注意力机制侧重于关注全局特征，而忽略了对深度伪造人脸检测至关重要的局部细节。为此，设计动态局部特征提取模块，利用卷积进行逐通道逐点局部特征提取，并通过门控机制动态调整每个像素的重要性，突出关键局部特征。该模块的结构如图 2 所示。

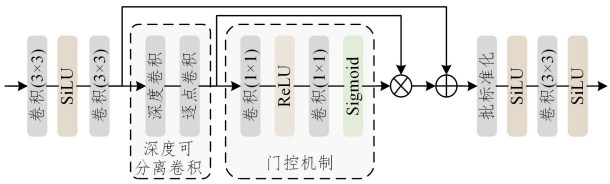

图 2　动态局部特征提取模块的结构

Fig. 2　Structure of the dynamic local feature extraction module

将每个模块的输入特征表示为 $f_i \in \mathbb{R}^{H_i \times W_i \times C_i}$。首先，$f_i$ 通过两次 3×3 卷积和 *SiLU* 提取浅层纹理特征，如式(1)所示：

$$f_i^{\text{shallow}} = Conv_{3\times3}(SiLU(Conv_{3\times3}(f_i))) \tag{1}$$

接着，使用深度可分离卷积提取局部特征并保留其空间信息，如式(2)所示：

$$f_i^{\text{depth}} = PointwiseConv(DepthwiseConv(f_i^{\text{shallow}})) \tag{2}$$

其中，$DepthwiseConv(\cdot)$表示对每个输入通道独立进行卷积操作，建模通道内空间关联；$PointwiseConv(\cdot)$表示对深度卷积的输出通道进行线性组合，建模通道间关联。然后，特征 f_i^{depth} 通过两层 1×1 卷积和激活函数，动态调整每个像素的重要性，内层采用 1×1 卷积结合 ReLU 激活函数来增强模型的非线性表达，外层为 1×1 卷积和 Sigmoid 激活函数，如式(3)所示：

$$f_i^{\text{gate}} = Sigmoid(Conv_{1\times1}(ReLU(Conv_{1\times1}(f_i^{\text{depth}})))) \tag{3}$$

将特征 f_i^{gate} 与 f_i^{depth} 逐元素相乘，并与浅层纹理特征 f_i^{shallow} 逐元素相加，以增强关键特征，同时抑制次要特征，如式(4)所示：

$$f_i^{ml} = f_i^{\text{gate}} \cdot f_i^{\text{depth}} + f_i^{\text{shallow}} \tag{4}$$

最后，f_i^{ml} 通过批标准化、3×3 卷积与两次 SiLU 激活函数，强化特征表达并补充 Vision Transformer 可能忽略的局部细节，如式(5)所示：

$$f_i^{\text{local}} = SiLU(Conv_{3\times3}(SiLU(BN(f_i^{ml})))) \tag{5}$$

2.2　多尺度特征提取与位置增强

2.2.1　多尺度特征提取

由于不同人脸伪造算法生成的伪造区域不同，本文引入膨胀卷积操作来捕获多尺度伪造痕迹，如图 1 所示。将第 i 个多尺度特征提取模块的输入特征表示为 $f_i \in \mathbb{R}^{H_i \times W_i \times C_i}$。该输入特征 f_i 首先通过不同膨胀率的卷积层提取多尺度特征。3 个多尺度特征提取模块的膨胀卷积信息如图 3 所示。对于第 j 个膨胀卷积层，其输出的特征表示为：

$$f_{ij}^{\text{multi}} = GELU(Conv_{ij}(f_i)) \tag{6}$$

$$GELU(x) = x \cdot \Phi(x) \tag{7}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \tag{8}$$

其中，f_{ij}^{multi} 表示第 i 个多尺度特征提取模块中第 j 个膨胀卷积层的输出特征，$GELU(\cdot)$为非线性激活函数，$\Phi(\cdot)$为标准正态分布的累积分布函数。相较于 ReLU 对负输入进行硬截断易导致细粒度信息丢失及 Sigmoid 强压缩输出范围会加剧梯度消失的问题，GELU 通过概率加权机制自适应保留小值特征，既维持了模型的非线性表达能力，又有效缓解了梯度畸变。在本文多尺度伪造痕迹提取场景中，GELU 可避免膨胀卷积捕获的细节特征(如边缘、纹理等)因硬截断受损，同时能增强关键特征的响应，为后续特征拼接与 Token 转换提供了鲁棒性更强的多尺度伪造痕迹表征。接着，将同一模块内不同膨胀卷积层的输出特征拼接，转换为 Token，得到多尺度特征 $f_i^{\text{multi}} \in \mathbb{R}^{(H_i \times W_i) \times C_i}$。

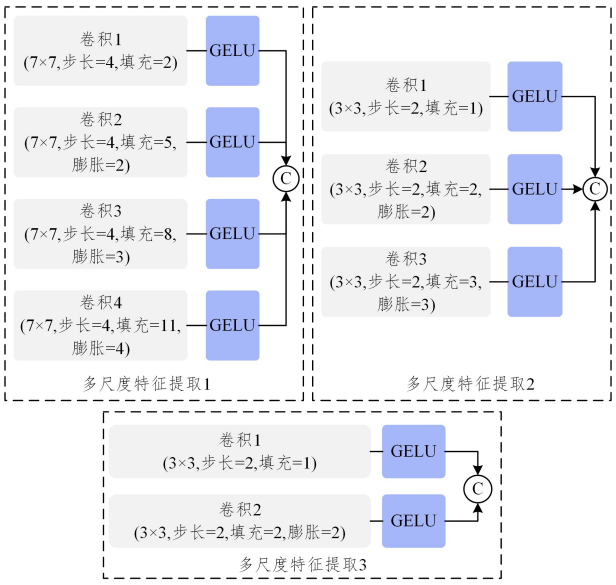

图 3 3 个多尺度特征提取模块的膨胀卷积信息

Fig. 3 Dilated convolution information of three multi-scale feature extraction modules

2.2.2 位置增强机制

由于膨胀操作可能破坏像素间的原始位置关系，因此设计位置增强机制来增强像素间的位置关联性，其结构图如图 4 所示。多尺度特征 f_i^{multi} 首先通过 3 个线性层分别映射到查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$。接着，引入位置编码嵌入层将序列位置索引 $positions$(取值为 0,1,…,$N-1$)映射到嵌入空间，并分别添加到 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 中，如式(9)和式(10)所示：

$$\boldsymbol{Q}_P = Q + Embedding(positions) \tag{9}$$

$$\boldsymbol{K}_P = K + Embedding(positions) \tag{10}$$

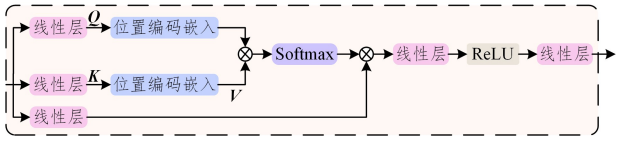

图 4 位置增强机制的结构

Fig. 4 Structure of the positional enhancement mechanism

然后，通过矩阵乘法将添加位置信息后的 $\boldsymbol{Q}_P$ 与 $\boldsymbol{K}_P$ 相乘，并应用 Softmax 得到注意力权重，如式(11)所示：

$$\boldsymbol{A} = \text{Softmax}(\boldsymbol{Q}_P \cdot (\boldsymbol{K}_P)^{\mathrm{T}}) \tag{11}$$

接着，通过矩阵乘法将注意力权重 $\boldsymbol{A}$ 应用于值 $\boldsymbol{V}$，并通过两层线性层和 ReLU 激活函数，得到位置增强后的特征 $f_i^{\text{position}} \in \mathbb{R}^{(H_i \times W_i) \times C_i}$，如式(12)所示：

$$f_i^{\text{position}} = Linear(RELU(Linear(\boldsymbol{A} \times \boldsymbol{V}))) \tag{12}$$

位置增强特征 f_i^{position} 先经层归一化处理，再输入多头自注意力机制，并与动态局部特征提取模块提取的关键局部信息进行融合，使 Token 能够同时携带多尺度特征与局部特征，如式(13)所示：

$$f_i^{\text{fuse}} = MHSA(LN(f_i^{\text{position}})) + DE(f_i) \tag{13}$$

其中，$MHSA(\cdot)$为多头自注意力机制，$DE(\cdot)$为动态局部特征提取。最后，f_i^{fuse} 经过多层感知机与层归一化，输入到下一个多尺度特征提取与位置增强模块中，如式(14)所示：

$$f_{i+1} = LN(MLP(f_i^{\text{fuse}})) \tag{14}$$

上述过程依次重复 3 次，每阶段通过下采样逐步降低特征的空间维度。最终生成的特征映射被展平为 Token，输入到接下来的全局-局部频域分量交互模块中。

2.3 全局-局部频域分量交互

鉴于某些伪造痕迹在图像质量下降时可能在 RGB 空间中难以察觉，但在频域中仍可被有效检测，本文设计了全局-局部频域分量交互模块，旨在通过频域分解注意力机制实现全局与局部信息的高效交互，其结构图如图 1 所示。将第 i 个全局-局部频域分量交互模块的输入特征表示为 $t_i \in \mathbb{R}^{(H/16 \times W/16) \times D}$。首先，对该特征进行层标准化，抑制特征分布差异。然后，将其输入到频域分解注意力机制中，以实现全局与局部信息的交互。频域分解注意力机制的结构如图 5 所示。对标准化后的特征进行快速傅里叶变换，得到频域特征，如式(15)所示：

$$x_i = \mathcal{F}[LN(t_i)] \tag{15}$$

其中，$\mathcal{F}[\cdot]$为快速傅里叶变换，x_i 为第 i 个全局-局部频域分量交互模块的频域特征。然后，将 x_i 按频率区间划分为多个频域分量，覆盖低频到高频子带，如式(16)所示：

$$x_{i1}, x_{i2}, \cdots, x_{in} = chunk(x_i) \tag{16}$$

其中，x_{ij} ($j=1,\cdots,n$)为第 i 个全局-局部频域分量交互模块的第 j 个频域分量，n 为频域分量的数量。在本文中，n 被设置为 8。对于每个分量 x_{ij} ($j=1,\cdots,n$)，通过 3 个线性层将其映射为查询 $\boldsymbol{Q}$、键 $\boldsymbol{K}$ 和值 $\boldsymbol{V}$，并计算注意力，如式(17)所示：

$$x_{ij}^{\text{interaction}} = \text{Softmax}\left(\frac{\boldsymbol{Q}_{ij} \times (\boldsymbol{K}_{ij})^{\mathrm{T}}}{\sqrt{D}}\right) \times \boldsymbol{V}_{ij} \tag{17}$$

其中，$x_{ij}^{\text{interaction}}$ 是第 i 个全局-局部频域分量交互模块的第 j 个频域分量的全局与局部信息交互的结果。将所有频域分量的全局与局部信息交互结果拼接为一个张量，如式(18)所示，其中包含了全局与局部信息之间的各种依赖关系。

$$x_{ij}^{\text{interaction}} = concat(x_{i1}^{\text{interaction}}, x_{i2}^{\text{interaction}}, \cdots, x_{in}^{\text{interaction}}) \tag{18}$$

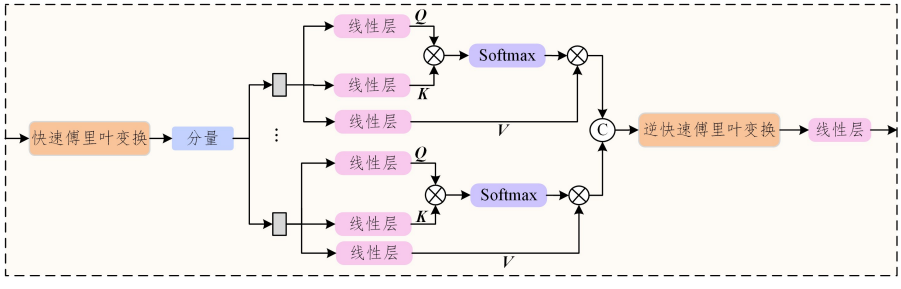

图 5 频域分解注意力机制的结构

Fig. 5 Structure of the frequency domain decomposition attention mechanism

然后，通过逆快速傅里叶变换将其转换回空域，得到频域分解注意力机制的输出 $t_i^{\text{interaction}}$，如式(19)所示：

$$t_i^{\text{interaction}} = Linear(\mathcal{F}^{-1}[x_i^{\text{interaction}}]) \tag{19}$$

其中，$\mathcal{F}^{-1}[\cdot]$表示逆快速傅里叶变换，$t_i^{\text{interaction}}$ 为第 i 个全局-局部频域分量交互模块的频域分解注意力机制的输出。接着，将频域结果、原始输入、动态局部特征残差融合，如式(20)所示，提取在伪造图像质量下降时可能在 RGB 空间中消失的伪影。

$$t_i^{\text{fuse}} = t_i^{\text{interaction}} + t_i + DE(t_i) \tag{20}$$

其中，$DE(\cdot)$表示动态局部特征提取。最后，t_i^{fuse} 经过层标准化和多层感知机，得到第 i 个全局-局部频域分量交互模块的输出，如式(21)所示：

$$t_{i+1} = MLP(LN(t_i^{\text{fuse}})) + t_i^{\text{fuse}} \tag{21}$$

3 个全局-局部频域分量交互模块依次堆叠，最后一个全局-局部频域分量交互模块输出的特征将被输入到像素关系相似度和全连接层中，以获得最终的分类结果。

2.4 联合损失函数

鉴于深度伪造技术生成的人脸图像日益逼近真实人脸的视觉真实度，传统的单一分类损失函数(如交叉熵损失)难以有效捕捉伪造痕迹的细微特征。同时，真实人脸图像的相邻像素在纹理结构与颜色过渡上具有天然的高度相似性(对应较低的损失值)，而伪造人脸图像由于生成或篡改过程的干扰，往往丢失了这种自然相关性(对应较高的损失值)。基于此，本文设计像素关系相似度损失函数，并将其与交叉熵损失函数相结合，构建联合损失函数，如式(22)所示：

$$\mathcal{L} = \lambda \mathcal{L}_{\text{pixel}} + \mathcal{L}_{\text{cls}} \tag{22}$$

其中，$\mathcal{L}_{\text{pixel}}$为像素关系相似度损失函数；$\mathcal{L}_{\text{cls}}$为交叉熵损失函数；$\lambda$ 为平衡超参数，用于调节两部分损失的相对重要性。默认情况下，λ 被设置为 0.001。像素关系相似度损失函数如式(23)所示：

$$\mathcal{L}_{\text{pixel}} = \sum_{i=1}^{N}(1 - CosineSimilarity(y_i, y_j)) \tag{23}$$

其中，N 为图像的像素总数；y_i 表示第 i 个像素；y_j 为 y_i 的邻域像素；$CosineSimilarity(\cdot)$为余弦相似性，用于衡量两个像素之间的相似度，值越接近 1，表示两像素相关性越强。交叉熵损失函数如式(24)所示：

$$\mathcal{L}_{\text{cls}} = y\log\hat{y} + (1-y)\log(1-\hat{y}) \tag{24}$$

其中，y 设置为 1 表示伪造人脸，设置为 0 则表示真实人脸；$\hat{y}$ 是模型的预测值。

3 实验

3.1 数据集与评价指标

3.1.1 数据集

本文将对比实验分为数据集内评估、跨数据集评估和跨操纵方法评估 3 个部分，以检验所提方法的泛化能力。

1)数据集内评估：采用 FaceForensics++(FF++)[17]数据集进行训练和测试。该数据集包含 1 000 个原始视频和 4 000个伪造视频，伪造视频由 Deepfakes，Face2Face，FaceSwap和 NeuralTextures 这 4 种伪造技术生成。遵循 FF++，随机选择 720 个视频用于训练，140 个用于验证，140 个用于测试，对每个视频采样 270 帧。针对 FF++数据集中真实样本与伪造样本呈现 1:4 的失衡问题，对真实样本进行 4 次重复采样[18]。这种样本均衡处理既能避免模型在训练过程中因过度关注伪造样本的损失而忽视对真实样本特征的学习，又能有效防止伪造样本特征对真实样本特征分布空间的挤压，促使模型同步优化两类样本的分类误差，从数据层面增强模型的泛化性。

2)跨数据集评估：为验证模型的跨域泛化能力，在 FF++数据集上完成训练后，选择 Celeb-DF[19]，DFDC[20]，WildDeepfake[21]和 UADFV[22]这 4 个未知数据集对模型的泛化能力进行评估。Celeb-DF[19]包含 590 个原始视频和 5 639 个伪造视频。DFDC[20]包含 119 197 个视频，其中 19 197 个为真实片段。WildDeepfake[21]包含 3 509 个伪造视频序列和 3 805 个真实视频序列。UADFV[22]包含 49 个真实视频和 49 个伪造视频。每个数据集随机选择真实和伪造各 20 000 帧进行测试。

3)跨操纵方法评估：采用 GID-DF 和 GID-F2F 两种评估协议，分别在 FF++数据集的其余 3 种伪造方法上进行训练，但在 Deepfakes 和 Face2Face 上进行测试。训练集、验证集和测试集的划分与 FF++一致。

3.1.2 评价指标

本文采用接收者操作特征曲线(Receiver Operating Characteristic，ROC)的曲线下面积(Area Under Curve，AUC)与准确率(Accuracy，ACC)作为评估指标。由于所提出的方法是基于图像级检测机制，因此通过单张图像生成预测分数。图像级检测方法不仅能够检测伪造图像，还能拓展至伪造视频的识别。文中与最先进的方法的比较结果除了来自于相关文献，否则会明确指出结果是通过运行其代码得出的。

3.2 参数设置

使用 RetinaFace[23]裁剪面部区域，并将输入图像的大小调整为 224×224。采用 AdamW 优化器，初始学习率设置为 5×10^{-3}，权重衰减系数为 0.001，批量大小设为 64，使用 PyTorch 和 NVIDIA GeForce RTX 4090 对整个模型进行 300 个 epoch 的训练。在第一个 epoch 中，使用 warm-up 策略将学习率从 1×10^{-6}增长到 5×10^{-3}，之后的 epoch 中学习率将持续下降。

3.3 对比实验

3.3.1 数据集内评估

所提方法在 FF++数据集上与其他先进方法的性能对比结果如表 1 所列。本文方法在参数量较少且推理速度较快的情况下，其 AUC 指标超越了所有对比方法，较 SPSL[24]提升了 3.93 个百分点。SPSL[24]虽从频域挖掘伪造痕迹，但由于不同伪造方法生成的伪造区域特征差异显著，且该方法未充分考虑多尺度特征，因此性能受限；相比之下，本文方法通过频域信息提取与多尺度特征融合策略，显著提升了检测性能。FFD[14]通过可训练的 SRM 卷积提取高频噪声特征以构建噪声模型，并结合多尺度卷积进行处理，但其仅依赖 SRM 提取高频噪声的方式存在固有局限性；本文方法则通过频域分解注意力机制，实现了不同频率信息的跨尺度交互，能够捕捉 RGB 空间中难以察觉的频域异常特征，因此其性能优于 FFD[14]。本文方法在 AUC 指标上较 WaViT-CDC[25]高出

0.44个百分点。WaViT-CDC[25]通过中心差分卷积捕捉局部细粒度差异，但其依赖固定的差分运算逻辑；相比之下，本文设计动态局部特征提取模块，通过逐通道卷积提取局部特征，并基于像素在特征表示中的重要性进行动态加权。这种自适应加权机制能更精准地突出伪造区域的关键信息，较固定差分运算具备更强的灵活性与针对性，尤其适用于FF++数据集等包含微妙伪造痕迹的检测任务。尽管HFI-Net[26]，LDFnet[5]和LiSiam[27]采用了更大的输入尺寸，但是本文方法在AUC指标上仍分别超出这3种方法0.59个百分点、0.33个百分点和0.12个百分点，这得益于频域特征的引入。此外，本文方法与F2Trans[28]的性能相当，AUC指标仅比F2Trans高出约0.01个百分点。

表1 FF++数据集上的实验结果

Table 1 Experimental results on the in-domain dataset

模型	输入大小	参数量	FLOPs(Mac)	推理速度(帧/秒)	*AUC*/%	*ACC*/%
SPSL[24]	299	—	—	—	95.32	91.50
HFI-Net[26]	384	1.13×10^{8}	2.040×10^{10}	—	98.66	95.12
FFD[14]	224	—	—	—	98.70	94.46
WaViT-CDC[25]	224	$5.959\,16\times10^{9}$	2.149×10^{10}	58	98.81	96.35
LDFnet[5]	256	9.36×10^{5}	8.010×10^{8}	236	98.92	96.01
LiSiam[27]	299	—	—	—	99.13	96.51
F2Trans-S[28]	224	$1.175\,2\times10^{8}$	1.972×10^{10}	—	99.18	96.09
F2Trans-B[28]	224	$1.280\,1\times10^{8}$	2.178×10^{10}	—	99.24	**96.60**
本文方法	224	5.05×10^{6}	1.644×10^{10}	21.91	**99.25**	96.20

3.3.2 跨数据集评估

为评估所提方法在面对现实世界未知被操纵面部数据时的泛化性，将模型在FF++数据集上进行训练，并在Celeb-DF，DFDC，WildDeepfake和UADFV这4个未知数据集上进行测试。本文方法与其他先进方法在AUC指标上的对比结果如表2所列，其中带有“*”的模型代表复现其代码的结果。

表2 跨数据集实验的AUC结果

Table 2 AUC results of cross-domain dataset experiments

(%)

模型	Celeb-DF	DFDC	WildDeepfake	UADFV	平均值
UIA-ViT[29]	64.60	65.29	70.26	—	66.72
M2TR*[18]	68.20	69.94	76.30	82.70	74.29
E-TAD[30]	68.95	57.10	66.50	—	64.18
RFFD[31]	69.34	70.03	73.08	—	70.82
TAN-GFD[32]	72.33	73.46	—	—	72.90
RMLD-HFTF[33]	73.63	74.52	—	—	74.08
FoCus[34]	76.13	68.42	73.31	—	72.62
本文方法	**78.62**	**86.75**	**82.16**	**86.75**	**83.57**

从表2中可以看出，本文方法在所有未知测试数据集上的AUC指标均显著优于其他方法。本文方法在3个数据集上的平均AUC值较无监督方法UIA-ViT[29]高出16.85个百分点。UIA-ViT[29]采用多元高斯模型估计真伪特征分布，基于马氏距离生成伪造位置图，并结合ViT的多头注意力机制构建一致性约束。然而，该方法依赖自监督伪标注信号，导致对细微伪造特征的捕捉能力受限，难以应对未知数据集中新型伪造手段引入的隐蔽性不一致特征，因此泛化性能弱于本文方法。本文方法在DFDC数据集上的AUC指标较E-TAD[30]高出29.65个百分点。E-TAD[30]通过挖掘不同伪造数据的共性，结合纹理与伪影特征判别合成内容，有效弥补了传统方法因忽略特征交互导致的泛化能力不足的问题。但该方法依赖已知特征分布，若未知数据中的纹理与伪影交互模式超出训练数据的覆盖范围，模型将难以有效聚合新特征，导致其对未知伪造场景的适应性不及本文方法。与M2TR[18]，RFFD[31]，TAN-GFD[32]和RMLD-HFTF[33]相比，本文提出的方法在4个未知数据集上的性能均更优。M2TR[18]，RFFD[31]，TAN-GFD[32]和RMLD-HFTF[33]这4种方法均以纹理特征为核心检测线索，通过多尺度或多层次特征处理架构挖掘全局与局部的纹理差异及噪声不一致性，并引入特征融合机制整合多维度信息，以增强模型在跨数据集场景中的泛化能力，避免对特定伪造伪影的过拟合。而本文方法借助频域内的全局与局部特征交互机制，将伪造区域的周期性噪声、相位畸变等隐蔽特征映射至频域空间，通过全局特征捕捉伪造操作引起的整体频率分布异常，同时利用局部特征定位高频细节中的篡改痕迹，能够更深入地挖掘伪造痕迹，因此性能表现更优。本文方法在3个数据集上的平均AUC值较FoCus[34]高出10.95个百分点。FoCus[34]采用全卷积块，结合分类注意力图定位伪造区域，通过多层特征图最大池化提取分类特征，检测多伪造区域，并引入Sobel算子提取边缘相关鲁棒特征。但其有效性依赖于伪造与真实边缘存在差异的假设，若未知数据的边缘处理更接近真实人脸，边缘特征区分力下降会导致伪造痕迹提取失效。本文方法通过频域分解注意力强化伪造痕迹表征，能够有效捕捉伪造面部数据在频域空间中呈现的细微异常特征，如高频成分的不连续性或低频结构的非自然分布，增强了对未知伪造技术的泛化能力，避免了上述局限对性能的影响，从而表现更优。综上，本文方法在面对未知的被操纵面部数据时展现出了良好的泛化能力和鲁棒性。

3.3.3 跨操纵方法评估

为验证本文方法对未知操纵类型的泛化能力，将模型在FF++数据集的3种伪造类型生成的数据集上进行训练，并在另一种伪造类型的数据集上进行测试，实验结果如表3所列。其中，M2TR[18]的实验结果引自F2-Trans[28]。

表3 跨操纵方法实验的AUC结果

Table 3 AUC results of cross-manipulation method experiments

(%)

模型	GID-DF	GID-FF
HFI-Net[26]	86.80	73.01
CFM[36]	88.00	81.40
SCLM[37]	94.10	81.40
SFCF[35]	94.20	82.10
M2TR[18]	94.91	76.99
本文方法	95.23	85.66

与 HFI-Net[26]，SFCF[35] 和 M2TR[18] 相比，本文方法在未知操纵类型检测中表现出更优的性能。HFI-Net[26]，SFCF[35] 和 M2TR[18] 均通过频域特征捕捉伪造痕迹，并采用多尺度分析与特征融合的策略。相比之下，本文方法虽然同样运用了多尺度分析与特征融合，但设计了频域分解注意力机制用于特征融合，通过频域维度对不同特征进行整合，充分挖掘不同特征间的互补信息，从而实现了更优的性能。本文方法在 GID-DF 和 GID-FF 数据集上的 AUC 指标较 CFM[36] 分别高出 7.23 个百分点和 4.26 个百分点。CFM[36] 通过无先验数据增强策略抑制特定伪造痕迹，降低了对专家知识的依赖，并借助实例相似度损失与局部相似度损失挖掘伪造样本的关键特征。但其在提炼先验知识时过度依赖已知操纵类型，仅能提取像素级混合边界等浅层特征，难以适应未知伪造技术产生的差异化伪造特征，导致性能弱于本文方法。本文方法在检测未知操纵类型时的泛化能力优于 SCLM[37]。SCLM[37] 通过监督对比学习策略增强模型泛化性，融合 SRM 特征与 RGB 特征以全面提取检测线索，并强化纹理细节与语义信息的融合。然而，其训练过程构建的知识体系可能未能完全覆盖未知操纵的伪造逻辑。相比之下，本文方法通过联合损失函数聚焦像素位置关系约束，从而在泛化性能上显著优于 SCLM[37]。综上，本文方法在面对未知类型的操纵时展现出了良好的泛化能力。

3.4 可视化分析

为进一步理解所提方法，利用 Grad-CAM[38] 技术对模型关注区域进行可视化，以评估其泛化能力。图 6 展示了 FF++ 数据集中的 4 种伪造方法以及 Celeb-DF，DFDC，WildDeepfake 和 UADFV 数据集的可视化热力图。

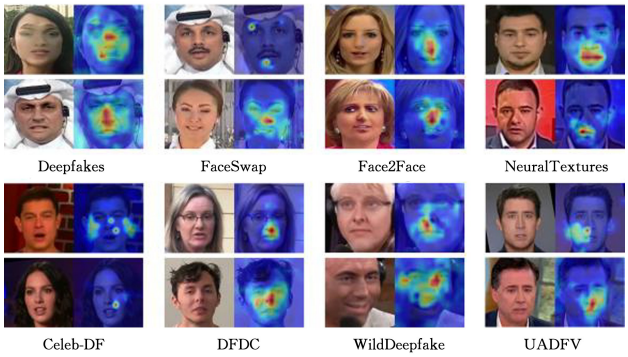

图 6 可视化热力图

Fig. 6 Heatmap visualization

可以看出，所提方法并未集中关注同一特定区域，而是针对不同的伪造类型展现出不同的关注区域。对于数据集内的可视化，Face2Face 和 NeuralTextures 伪造算法主要聚焦于面部的细微部分，如鼻子和嘴巴。从 Face2Face 和 NeuralTextures 的热力图可以看出，本文方法聚焦于鼻子和嘴巴，这进一步证明了模型对细微伪造痕迹的捕捉能力。在进行面部交换时，源图像与目标图像之间的边缘可能会出现模糊或不自然的过渡，本文方法特别关注由 FaceSwap 生成的伪造人脸图像中的这些混合交界处，并可以有效识别不自然过渡区域。对于跨数据集的可视化，本文方法即使是在特定类型的伪造算法生成的数据集上进行训练的，也仍然能够关注到不同面部区域的伪造痕迹，而不局限于固定的面部区域。因此，在由不同类型的伪造算法生成的数据集上进行测试时，模型的泛化能力并未显著下降。通过热力图的定性分析可知，本文方法展现出了较好的泛化性。

3.5 消融实验

为验证本文提出的各个模块的有效性，在 FF++ 数据集上完成模型训练后，在 Celeb-DF(CDF) 数据集上对其进行测试，实验结果如表 4 所列。其中，“局部特征”表示动态局部特征提取模块，“多尺度特征”表示多尺度特征提取模块(无位置增强机制)，“频域分解”表示全局-局部频域分量交互模块，“位置增强”表示位置增强机制，“像素关系”表示像素关系相似度损失函数。

表 4 消融实验结果

Table 4 Results of ablation experiment

模块					FF++	CDF
局部特征	多尺度特征	频域分解	位置增强	像素关系	*AUC*/%	*AUC*/%
					50.68	47.44
√					71.06	66.73
	√				61.53	52.39
		√			61.85	57.52
√		√			73.75	65.68
	√		√	√	68.97	63.05
√	√				98.01	74.22
√	√	√			98.83	75.72
√	√	√	√		99.18	75.40
√	√	√	√	√	**99.29**	**78.62**

从表 4 中可以看出，所有模块均能提升模型性能且组合使用时效果最佳。相比基线模型 Vision Transformer，单独使用动态局部特征提取模块时在 FF++ 和 Celeb-DF 数据集上分别将 AUC 值提升了 20.38 个百分点和 19.29 个百分点，有效补充了 ViT 对局部信息捕捉的不足。无位置增强机制的多尺度特征提取模块在 FF++ 和 Celeb-DF 数据集上将 AUC 值分别提升了 10.85 个百分点和 4.95 个百分点，这表明该模块能够有效捕获不同尺度的伪造痕迹。在此基础上添加位置增强机制和像素关系相似度损失函数后，模型性能进一步提升，AUC 值分别高出 7.44 个百分点和 10.66 个百分点。这证明了位置增强机制和像素关系相似度损失函数能够关注像素间的位置关系，有效解决膨胀操作导致的像素位置关系被破坏的问题。单独使用全局-局部频域分量交互模块时，FF++ 和 Celeb-DF 数据集的 AUC 值分别提升了 11.17 个百分点和 10.08 个百分点，进一步验证了频域分量信息交互的有效性。从表 4 中还可以看出，去除任一模块均会导致 AUC 值下降，且基线模型依次添加各模块时 AUC 值持续提升，这充分说明了各模块间具有协同增效作用。

此外，为直观展示所提模块在人脸伪造检测中的作用，对基于不同模块组合策略及基线模型 ViT 学习到的特征分布进行可视化，结果如图 7 所示，其中涉及原始图像和 FF++ 数据集中的 4 种不同伪造方法。图 7(a)—图 7(f) 分别表示在基线模型 ViT 的基础上逐步加入多尺度特征提取模块(无位置增强机制)、动态局部特征提取模块、全局-局部频域分量交互模块、位置增强机制及像素关系相似度损失函数的特征图，特征图中的亮区表示模型重点关注的区域。

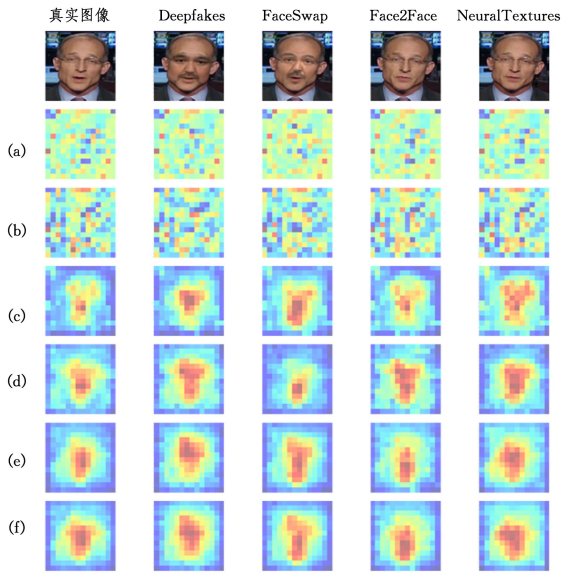

图 7 不同模块组合的特征图

Fig. 7 Feature maps of different module combinations

从图 7 中可以观察到，多尺度特征提取模块关注非特定大小的区域，突出了对多尺度信息的有效捕捉，而在引入动态局部特征提取模块后，模型聚焦于局部细节的识别。在此基础上，结合全局-局部频域分量交互模块，模型关注区域进一步特定化。通过添加位置增强机制与像素关系相似度损失函数，模型基于像素间关系深入挖掘伪造痕迹，能精准识别关键信息。值得注意的是，本文方法生成的特征图在区分不同伪造方法生成的真实与伪造人脸图像时，展现出显著的区分性改进，关注区域差异明显。相比之下，基线模型 ViT 的特征图注意力区域分散，聚焦不足。这一对比充分表明，本文方法在识别真实与伪造人脸方面展现出更高的区分力，验证了其在人脸伪造检测任务中的鲁棒性和泛化性。

结束语 本文提出了一种位置增强与频域分量交互的深度伪造检测方法，以 Vision Transformer 为骨干网络，挖掘与伪造痕迹相关的信息，通过构建动态局部特征提取模块、多尺度特征提取与位置增强模块、全局-局部频域分量交互模块，识别不同区域的伪影，并增强像素间的位置关系；同时，利用联合损失函数优化模型，提升模型识别真实人脸图像与伪造人脸图像的判别能力，从而提高模型的鲁棒性及泛化性。在数据集内、跨数据集以及跨操纵方法上的实验数据结果和可视化展示分析，验证了所提方法的有效性，其相较于许多现有检测方法具有更高的泛化性及鲁棒性。本文实验尽管覆盖了跨数据集与跨操纵方法场景，但尚未充分考虑真实场景中的复杂干扰因素，如低光照、大角度姿态、遮挡等，因此所提模型在真实环境下对伪造样本检测的泛化能力仍需提升。未来将构建跨场景泛化的检测框架，通过融合动态光照、复杂背景、生物特征变异等真实场景干扰因素，构建大规模混合数据集，并探索迁移学习与元学习在伪造人脸检测中的应用，推动相关技术向实际监控等场景落地。

参 考 文 献

[1] THIES J, ZOLLHÖFER M, NIESSNER M. Deferred neural rendering: image synthesis using neural textures [J]. ACM Transactions on Graphics, 2019, 38(4): 66.

[2] THIES J, ZOLLHÖFER M, STAMMINGER M, et al. Face2-Face: Real-Time Face Capture and Reenactment of RGB Videos [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2387-2395.

[3] ZHAO H, WEI T, ZHOU W, et al. Multi-attentional deepfake detection[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2185-2194.

[4] ZHANG D, CHEN J, LIAO X, et al. Face Forgery Detection via Multi-Feature Fusion and Local Enhancement[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(9): 8972-8977.

[5] GUO Z, WANG L, YANG W, et al. LDFnet: Lightweight Dynamic Fusion Network for Face Forgery Detection by Integrating Local Artifacts and Global Texture Information[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(2): 1255-1265.

[6] ZHANG K, FAN Z X. Improved Face Forgery Detection Method Based on Adversarial Training[J]. Journal of Chongqing Technology and Business University (Na- tural Science Edition), 2025, 42(4): 88-94.

[7] WANG Y M, HU J, WU X S, et al. Compressed deepfake video detection method based on inconsistent facial motion[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2025, 37(3): 445-452.

[8] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 6000-6010.

[9] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]// Proceedings of the International Conference on Learning Representations. 2021.

[10] ZHOU J, ZHAO X, XU Q, et al. MDCF-Net: Multi-Scale Dual-Branch Network for Compressed Face Forgery Detection[J]. IEEE Access, 2024, 12: 58740-58749.

[11] KHORMALI A, YUAN J S. Self-Supervised Graph Transformer for Deepfake Detection [J]. IEEE Access, 2024, 12: 58114-58127.

[12] ZHOU K, SUN G, WANG J, et al. MH-FFNet: Leveraging Mid-High Frequency Information for Robust Fine-Grained Face Forgery Detection [J]. Expert Systems with Applications, 2025, 276(C).

[13] LAI Z M, ZHANG Y, LI D, et al. Leveraging high-frequency diversified augmentation for general deepfake detection[J]. Journal of Information Security and Applications, 2025, 89: 103994.

[14] ZHANG D Y, QI F F, CHEN J H, et al. Fake face detection based on fusion of spatial texture and high-frequency noise[J]. Chinese Journal Of Electronics, 2025, 34(1): 212-221.

[15] HUANG J S, YANG G M. Face Forgery Detection Method Based on Manipulation Trace Fusion[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2025, 42(4): 80-87.

[16] MIAO C,CHU Q,LI W,et al. Towards Generalizable and Robust Face Manipulation Detection via Bag-of-feature[C]// 2021 International Conference on Visual Communications and Image Processing. 2021:1-5.

[17] RÖSSLER A,COZZOLINO D,VERDOLIVA L,et al. Faceforensics++:Learning to detect manipulated facial images[C]// IEEE/CVF International Conference on Computer Vision(ICCV 2019). 2019:1-11.

[18] WANG J,WU Z,OUYANG W,et al. M2TR:Multi-modal Multi-scale Transformers for Deepfake Detection[C]// Proceedings of the 2022 International Conference on Multimedia Retrieval. 2022:615-623.

[19] LI Y,YANG X,SUN P,et al. Celeb-df:A large-scale challenging dataset for Deepfake forensics[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR 2020). 2020:3204-3213.

[20] DOLHANSKY B,HOWES R,PFLAUM B,et al. The Deepfake detection challenge(DFDC) preview dataset[J]. arXiv:1910.08854,2019.

[21] ZI B,CHANG M,CHEN J,et al. WildDeepfake:A challenging real-world dataset for Deepfake detection[C]// Proceedings of the 28th ACM International Conference on Multimedia. 2020:2382-2390.

[22] YANG X,LI Y,LYU S. Exposing deep fakes using inconsistent head poses[C]// 2019 IEEE International Conference on Acoustics,Speech and Signal Processing(ICASSP 2019). 2019:8261-8265.

[23] DENG J,GUO J,VERVERAS E,et al. Retinaface:Single-shot multi-level face localisation in the wild[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2020:5202-5211.

[24] LIU H,LI X,ZHOU W,et al. Spatial-phase shallow learning:Rethinking face forgery detection in frequency domain[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). 2021:772-781.

[25] BADR N E A,NEBEL J C,GREENHILL D,et al. WaViT-CDC:Wavelet Vision Transformer With Central Difference Convolutions for Spatial-Frequency Deepfake Detection[J]. IEEE Open Journal of Signal Processing,2025,6:621-630.

[26] MIAO C,TAN Z,CHU Q,et al. Hierarchical Frequency-Assisted Interactive Networks for Face Manipulation Detection[J]. IEEE Transactions on Information Forensics and Security,2022,17:3008-3021.

[27] WANG J,SUN Y,TANG J. Lisiam:Localization invariance Siamese network for Deepfake detection[J]. IEEE Transactions on Information Forensics and Security,2022,17:2425-2436.

[28] MIAO C,TAN Z,CHU Q,et al. F2Trans:High-Frequency Fine-Grained Transformer for Face Forgery Detection[J]. IEEE Transactions on Information Forensics and Security,2023,18:1039-1051.

[29] ZHUANG W,CHU Q,TAN Z,et al. UIA-ViT:Unsupervised Inconsistency-Aware Method Based on Vision Transformer for Face Forgery Detection[C]// Computer Vision-ECCV 2022. 2022:391-407.

[30] GAO J,MICHELETTO M,ORRÙ G,et al. Texture and Artifact Decomposition for Improving Generalization in Deep-Learning-Based Deepfake Detection[J]. Engineering Applications of Artificial Intelligence,2024,133(C):108450.

[31] GONG R,HE R,ZHANG D,et al. Robust face forgery detection integrating local texture and global texture information[J]. EURASIP Journal on Information Security,2025,2025(3):1-14.

[32] ZHAO Y,JIN X,GAO S,et al. TAN-GFD:generalizing face forgery detection based on textureinformation and adaptive noise mining[J]. Applied Intelligence,2023,53:19007-19027.

[33] JIANG Q,LIU S,MIAO S,et al. Robust manipulated media localization and detection based on high frequency and texture features[J]. Discover Computing,2025,28(1):1-17.

[34] TIAN J H,CHEN P,YU C,et al. Learning to Discover Forgery Cues for Face Forgery Detection[J]. IEEE Transactions on Information Forensics and Security,2024,19:3814-3828.

[35] ZHENG J S,ZHOU Y C,ZHANG N,et al. A Spatio-Frequency Cross Fusion Model for Deepfake Detection and Segmentation[J]. Neurocomputing,2025,628:129683.

[36] LUO A,KONG C,HUANG J,et al. Beyond the Prior Forgery Knowledge:Mining Critical Clues for General Face Forgery Detection[J]. IEEE Transactions on Information Forensics and Security,2024,19:1168-1182.

[37] DONG F,ZOU X,WANG J,et al. Contrastive Learning-Based General Deepfake Detection with Multi-Scale RGB Frequency Clues[J]. Journal of King Saud University-Computer and Information Sciences,2023,35(4):90-99.

[38] SELVARAJU R,COGSWELL M,DAS A,et al. Grad-CAM:Visual Explanations from Deep Networks via Gradient-Based Localization[C]// 2017 IEEE International Conference on Computer Vision(ICCV). 2017:618-626.

MENG Siyu, born in 2001, postgraduate, is a member of CCF (No. A00849G). Her main research interests include deepfake detection and computer vision.

WANG Rong, born in 1971, professor, Ph.D supervisor, is a member of CCF (No. C4366M). Her main research interests include pattern recognition and computer vision.

(责任编辑:柯颖)