中图法分类号:TP183　　**文献标识码**: A　　**文章编号**: 1006-8961(XXXX)XX-0001-14

论文引用格式:Cai Tijian, Huang Yuanxuan, Wang Zhenyu, Hu Cheng, Yi Shengquan, Xie Xin. XXXX. Adaptive face fraud detection model driven by transformer based graph representation learning. Journal of Image and Graphics, XX(XX):0001-0014(蔡体健, 黄远轩, 王振宇, 胡成, 易晟权, 谢昕. XXXX. Transformer图表示学习驱动的自适应人脸欺诈检测模型. 中国图象图形学报, XX(XX):0001-0014)[DOI: 10.11834/jig.250373]

Transformer图表示学习驱动的自适应人脸欺诈检测模型

蔡体健,黄远轩,王振宇,胡成,易晟权,谢昕

华东交通大学信息与软件工程学院　南昌　330013

摘　要:**目的**　针对欺诈人脸中存在的光照、几何比例、运动模式等方面的全局不一致性,本文通过引入数据驱动的图拓扑信息以增强模型的结构感知能力,并借助Transformer的全局自注意力机制弥补图结构在全局信息建模上的局限,提出了一种基于Transformer图表示学习的自适应人脸欺诈检测模型,旨在提升模型捕捉全局空间中不一致欺诈线索的能力。**方法**　该模型通过交替堆叠图神经网络(graph neural networks, GNN)与Transformer层来构建两者的互补优势,GNN通过屏蔽无关节点,将Transformer的盲目全局注意力转变为稀疏、低秩、结构敏感;Transformer凭借长距离依赖拓展了GNN的感受野,使消息传递更远、更灵活,它们的交替堆叠使模型具有更强的关系表示能力。鉴于GNN对图拓扑结构的强依赖性,本模型应用动态K近邻稠密算法来改进邻居节点的采样方法,利用数据驱动来自适应地调整每个节点的连接性(度),提高了模型的结构表示能力。为了提高模型对关键线索的敏感性,本模型使用双重注意力机制对图中的节点和边分别进行加权,对节点的加权可以使模型重点关注携带更多判别信息的图像块,而对边的加权可以突出信息流动的关键途径。这种双重注意力机制模拟了实体及其关系之间的复杂相互作用,使模型能够专注于最突出和最具信息性的特征和联系。为了提高模型抵御新攻击的能力,本文引入改进的元学习优化策略,通过组合的KL散度损失与软目标交叉熵损失函数来促使模型快速适应新的攻击类型或新的场合。**结果**　在MSU、Replay、OULU与CASIA的人脸活体检测跨数据集实验中,平均半总错误率(half total error rate, HTER)超过了所有对比算法;在FF++和Celeb-DF数据集上的域内测试中Acc达到98.93%与98.44%,而在FF++、Celeb-DF与DFD的交叉评估与跨域实验中Acc与曲线下面积(area under the curve, AUC)均高于最新模型。**结论**　本文提出的方法能够有效地捕捉并识别出欺诈样本中存在的细微欺诈线索以及全局不一致性现象,该方法在增强模型对不同数据集的泛化能力和对新环境的适应性方面尤为显著,确保了模型在多变的实际应用场景中均能达到较高的识别性能。本文数据集已在ScienceDB存档,访问链接为https://doi.org/10.57760/sciencedb.j00240.00096.

关键词:人脸欺诈检测;图神经网络;Transformer;元学习优化;动态K近邻稠密算法;全局不一致性

Adaptive face fraud detection model driven by transformer based graph representation learning

Cai Tijian, Huang Yuanxuan, Wang Zhenyu, Hu Cheng, Yi Shengquan, Xie Xin

College of Information and Software Engineering, East China Jiaotong University, Nanchang 330013, China

收稿日期:2025-08-03;**修回日期**:2026-01-30

*通信作者:黄远轩(2023068085404014@ecjtu.edu.cn)

基金项目:国家自然科学基金(项目编号:62162026,62362032);江西省自然科学基金(项目编号:20242BAB25114, 20242BAB26019)

Supported by: Project supported by the National Natural Science Foundation of China (Grant No. 62162026, 62362032); the Natural Science Foundation of Jiangxi Province, China(Grant No. 20242BAB25114, 20242BAB26019).

© 中国图象图形学报版权所有

Abstract: **Objective** The relentless evolution of presentation attacks pose a threat to the security of facial recognition systems, and low-cost deepfake technology has caused widespread social security issues. With the continuous confrontation between forgery and anti-counterfeiting technologies, the difference between real and forged data is becoming increasingly subtle, transient, and sparse. However, when the generated patches (or occluded patches) are mixed with the original image, global inconsistencies in lighting, geometric proportions, motion patterns, and other aspects are inevitable. This paper proposes an adaptive face fraud detection model based on Transformer graph representation learning, which utilizes fraud clues with widespread global spatial inconsistency in fraud samples. This model aims to improve the accuracy and generalization ability of face fraud detection to address increasingly complex security challenges. **Method** Firstly, the model constructs the complementary advantages of GNN (graph neural networks) and Transformer layers by alternately stacking them. GNN shields irrelevant nodes and converts Transformer's blind global attention into sparse, low rank, and structurally sensitive; Transformer expands the receptive field of GNN through long-range dependencies, making message transmission farther and more flexible. The combination of GNN and Transformer makes it easier for the model to capture globally inconsistent fraud clues. Secondly, due to the strong dependence of GNN on graph topology, this model applies the dynamic K-nearest neighbor dense algorithm to improve the sampling method of neighboring nodes. The algorithm adaptively adjusts the connectivity (degree) of each node based on the local feature density distribution in the latent space. In sparse or ambiguous regions, nodes connect to more neighbors to gather broader context; in dense, discriminative regions, connections focus on the most relevant neighbors. This dynamically constructs data-adaptive graph topologies that are inherently more resilient to noise and variations, providing a robust foundation for subsequent processing. Thirdly, this model uses a dual attention mechanism to weight the nodes and edges in the graph separately. The weight of nodes reflects the degree of influence of image patches in different regions on image labels, and adding weights to nodes can enable the model to focus on image patches that carry more discriminative information. Meanwhile, the weight of edges reflects the degree of influence of neighboring nodes on the central node, and the adding weights to edges can highlight the key pathways of information flow. This dual attention explicitly models the complex interplay between nodes and their relationships, allowing the model to focus on the most salient and informative features and connections, thereby dramatically boosting its relational learning capacity and sensitivity to subtle cues. Fourthly, considering that static models fail against novel attacks, this model incorporates meta learning optimization strategies to enable the model to quickly adapt to new types of attacks or new scenarios. The core of this strategy improvement is a new combination of soft label loss functions. In the inner loop training stage, the model utilizes the sensitivity of KL (Kullback-Leibler) Divergence loss function aiming to soft labels to quickly adjust uncertain model parameters, while in the outer loop stage, it utilizes the strong fault tolerance of Soft-Target Cross-Entropy loss function to ensure that the model learns stable and robust feature representations on various data. The improved meta learning algorithm not only demonstrates higher stability during the training process, but also maintains good generalization performance on different datasets. **Result** This article validates the effectiveness of the proposed model on seven major, publicly available benchmarks, which belong to traditional face anti-spoofing or deepfake detection tasks. In the face anti-spoofing experiment, four classic datasets, MSU-MFSD, Replay-Attack, OULU-NPU, and CASIA-FASD datasets, were used for cross-dataset experiments. The average HTER (Half Total Error Rate) obtained from the experiment exceeded all comparison algorithms, indicating that the proposed model has good generalization performance in face anti-spoofing. In the Deepfake detection experiment, FaceForensics++ (FF++) and Celeb-DF (v2) were used for intra-dataset experiments, achieving classification accuracies of 98.93% and 98.44%, respectively; Then, FF++ was used to conduct cross subset experiments within the dataset, and FF++, DFD (Deepfake Detection Dataset), and Celeb-DF were used to conduct cross-dataset experiments. The above cross domain experimental results show that the accuracy and AUC (Area Under the Curve) of this model are higher than those of the latest model. Further ablation experiments confirmed that DBD-KNN, meta learning optimization, TransGNN module, and FFN module in the model are all beneficial for improving model performance. The comparative experiment between KNN (K-Nearest Neighbors) and DBD-KNN (Density Based Dynamic K-Nearest Neighbors) confirms that even if KNN algorithm uses the optimal K value, its performance is still inferior to DBD-KNN algorithm. The DBD-KNN algorithm leads by 5.55% in the HTER (Half Total Error Rate) metric, mainly due to its ability to dynamically adapt to changes in data, thereby more effectively adapting to the characteristics of

© 中国图象图形学报版权所有

the dataset. In the loss function combination experiment of meta learning, it was confirmed that the best experimental results can be obtained when selecting the KL divergence loss function for the inner loop and the Soft-Target Cross Entropy loss function for the outer loop. After analyzing the complexity of the model, we found that although the proposed model has significant advantages in computational efficiency and parameter quantity, its detection accuracy still remains at a high level. Finally, by observing the evolution of the topological structure of the graph, it can be found that as the model training deepens, the connections between nodes are no longer limited to local areas, but begin to capture semantic features with distinctive characteristics throughout the entire image range. This reveals the inherent logic of the model in making detection decisions. **Conclusion** The method proposed in this article can effectively capture and identify subtle fraud clues and global inconsistencies present in fraud samples, thereby improving the detection accuracy of the model. This method is particularly remarkable in enhancing the model's generalization ability to different datasets and adaptability to new environments, ensuring that the model can achieve high recognition performance in diverse practical application scenarios. The dataset for this paper has been archived in ScienceDB and can be accessed viahttps://doi. org/10. 57760/sciencedb. j00240. 00096.

Key words: face presentation attack detection; graph neural networks; transformer; meta-learning; density-based dynamic k-nearest neighbors algorithm; global inconsistency

0 引　言

伴随着人脸信息价值的不断提高，人脸欺诈攻击的风险日益增加。一方面，高精度的物理呈现攻击（如照片、视频重放、3D面具等）增加了人脸活体检测（face anti-spoofing，FAS）技术的难度，严重地威胁了人脸识别系统的安全性；另一方面，生成式人工智能（如GANs、扩散模型等）的快速发展催生了高度逼真的Deepfake欺诈，低成本的数字攻击正在重塑安全威胁格局，其造成的破坏远超传统认知。面对这种局面，识别和区分真实人脸与欺诈人脸（包括物理攻击与数字攻击）的人脸欺诈检测技术成为了当前研究的一个热点课题。人脸欺诈线索主要包括两大类（Shang等，2023）：一种是在生成过程中产生的局部伪影，例如屏幕重放时的莫尔纹、镜面反射等，或基于GAN方法的深度伪造时，上采样会产生棋盘格伪影、离散的亮斑等；另一种是在将生成的人脸（或遮挡部分）与原始图像混合时产生的皮肤、光照、几何比例、运动模式等方面与环境的全局不一致性。这两类线索中，全局空间不一致性比局部伪影具有更强的泛化性，而图神经网络（GNN）强大的结构关系学习能力使其更易于捕获这种全局空间不一致性，因此GNN有望成为对抗人脸欺诈攻击的重要工具。

本文的主要贡献可归纳如下：

1）在模型结构上，本文通过将图的拓扑信息注入Transformer，使Transformer能感知模型的整体结构；而将Transformer的全局自注意力机制引入GNN能有效降低结构噪声的影响，它们的互补可增强模型的关系表示能力。并且在GNN层设计了边-点双注意力图卷积（dual-weighted graph convolution network，DW-GCN），对节点与边分别进行可学习加权，提升了模型对细微欺诈线索的敏感性。

2）在构建图拓扑结构时，本文引入了动态K近邻稠密算法（density-based dynamic K-nearest neighbors，DBD-KNN）来改进邻居节点的采样方法，该算法基于潜在空间内的局部特征密度分布自适应地调整每个节点的连接性（度），以构建适应数据的图结构。

3）本文引入了改进的元学习优化方法来提高模型的学习能力。为了在提高模型快速适应能力的同时还保证模型的稳定性，元学习优化的内循环选择了KL散度损失函数，而在外循环使用了更鲁棒的软标签交叉熵损失函数，有效地提升了模型的泛化能力与适应性。

1　相关工作

人脸欺诈检测包括人脸活体检测和深度伪造检测两大类，它们的主要差别在于攻击方式不同，分别对应物理攻击与数字攻击，其任务都是通过分析真实人脸与欺诈人脸之间的差异特征来辨别真伪。Yu等人（2024）指出人脸活体检测和深度伪造检测

© 中国图象图形学报版权所有

任务具有许多共同特征，真实或活体样本具有相似的外观特征和生理特征（如周期性rPPG节奏等），而欺骗/深度伪造样本具有不同类型的手工痕迹和噪声线索，它们的数据集具有互补性，在特征提取和模型学习方面存在较大的关联性，它们可以相互促进，共同提高模型性能。

早期的人脸活体检测技术通过纹理特征、图像质量、颜色失真、生理特征等区别信息来区分活体人脸和欺诈人脸。深度学习模型能够自动学习区分真假人脸的共性规律，成为当前主要的研究方法。Xie等人（2022）利用深度学习模型有效融合多模态信息，显著提升了模型对多种已知攻击的检测能力。为了提高模型的泛化能力，Jia等（2020）提出单边域泛化方法（single-side domain generalization，SSDG），仅使用来自不同域的真实人脸来学习一个广义的特征空间，其中真实人脸的特征分布是紧凑的，而伪造人脸的特征分布在不同域中；此外，设计了一个不对称的三元组损失来约束不同域的伪造人脸被分开，而真实人脸则被聚合，实验证明该模型增强了人脸反欺骗方法在未知场景的泛化能力。Cai等人（2023）在FAS模型中引入条件域对抗模块来实现特征和类层面多个源域的对齐，并通过熵函数来控制样本的优先级，减少难迁移样本对域泛化造成的影响，获得了较好的检测结果。为了提高模型对新型欺诈的适应能力，Shao等人（2020）提出了一种正则化细粒度元学习框架（regularized fine-grained meta face anti-spoofing，RFMeta），选择分类器参数作为元优化参数，元训练时同时模拟多个域迁移，以充分利用丰富的域移位信息。Chuang等（2023）提出了多任务的元学习框架，并将元学习的域间优势和三元组损失的域内优势相结合，提高了活体检测模型的泛化性能。

深度伪造检测是随着deepfake伪造技术应运而生的，由于deepfake伪造成本低和精度高的特点，其在虚假内容生成和欺诈行为中具有更高的危险性，也使得深度伪造检测技术面临更严峻的挑战。Chollet等（2017）利用XceptionNet比较了视频的全帧和提取人脸部分对模型训练的影响，结果显示基于人脸训练的模型效果远远好于全帧模型。Salman等（2025）提出AWARE-Net，基于三种先进架构（Xception、Res2Net101和EfficientNet-B7）组成的两层集成框架，通过动态加权机制组合多个模型的预测，增强了模型的检测性能。Afchar等（2018）通过MesoNet框架对模型进行了对抗性评估，研究发现，在数据分布相同的测试集上，检测器能够实现很高的检测准确率；然而，当面对未知或经过篡改的数据集时，模型的表现显著下降，尤其是那些特征重合程度不高的数据集，迁移能力较差，导致检测效果不佳。为了提高模型的适应性，Guan等人（2023）提出了一种基于元学习的多特征通道域加权框架（multi-feature channel domain-weighted，MCW），利用图像的RGB域和频域信息，通过为特征图上的通道分配元权重来增强模型检测目标的泛化能力，所提出的MCW框架在少样本学习场景中具有更好的微调潜力。随着生成式技术的发展，深度伪造检测也正朝着多模态特征融合的方向发展（Yao等，2025）。文献（Yan等，2022）利用空域、频域和时域等多模态数据共同进行欺诈检测；Long等（2025）也提出一种基于多域融合的多分支网络框架（MBMD），利用空间域，时空域与频域信息挖掘全面细致的伪造线索。然后也有文献（Shang等，2023）指出篡改的视频中不可避免地产生的空间伪影，而时间不一致性是随机出现在假视频中，并且频域的欺诈线索受到图像压缩影响较大，因此目前基于空域的人脸欺诈检测仍然是主要研究方法。

与传统的卷积神经网络（convolutional neural networks，CNN）相比，图神经网络（GNN）（Li等，2019）通常具有更弱的归纳偏置，不仅能捕获对象特征，而且能较为灵活地捕获不规则和复杂的对象关系，因此，当前GNN在图像处理方面的应用潜力在不断增长。目前已有一些文献将图神经网络用于深度伪造检测。文献（Khali等，2023）所提出的DFGNN模型是GNN与ResNet的结合，该模型通过将图像分割成小块（节点），通过连接最近的邻居来构建图结构，然后利用图卷积层和前馈神经网络（feed-forward network forward pass module，FFN）模块来提取多尺度的区别特征，从而实现对深度伪造的检测。GNN和Transformer（Zhang等，2024）的结合能够有效弥补彼此的不足，一方面图的拓扑信息注入Transformer，使Transformer感知模型的整体结构；而Transformer的全局自注意力机制能降低结构噪声的影响。文献（Khormali等，2024）所提出的模型（self-supervised graph transformer，SSGT）将GNN与Transformer模型相结合，通过自监督对比学习来

© 中国图象图形学报版权所有

预训练模型，再通过TransGNN模块更好地提取伪造特征，以增强模型的可解释性和可信度。为了进一步提高模型的适应性，文献（Mandal等，2022）将元学习策略应用于GNN网络，该文根据节点的度来选择元学习参数，它认为high-degree节点的结构信息丰富，其表示被认为是准确的，所以high-degree节点被用作元训练集（meta training）来学习共享先验。

总之，随着欺诈人脸攻击技术的发展，人脸活体检测技术面临着细微的全局特征提取能力不足、泛化能力弱以及适应性较差等问题。为此，本文提出了一种基于Transformer图学习的自适应人脸欺诈检测方法，利用GNN非结构化数据表示的优势和Transformer的全局特征提取能力，增强模型的检测精度，提高模型的可信度；通过引入元学习策略提高模型的适应能力，使得模型在面对少量未知攻击样本的新任务时，能够快速调整模型参数提高模型的准确性、鲁棒性、快速适应能力等。

2 本文方法

本文提出了一种基于Transformer图学习的自适应人脸欺诈检测模型，其整体架构如图1所示。该模型应用动态K近邻稠密算法来自适应调整节点的度，以构建适应数据的图结构；通过边-点双注意力图卷积来获得更重要的特征；利用Transformer的长距离特征提取能力，来增强模型全局特征提取能力；再引入元学习优化策略，通过组合的KL散度损失与软标签交叉熵损失函数，进一步提高模型的鲁棒性和适应性。

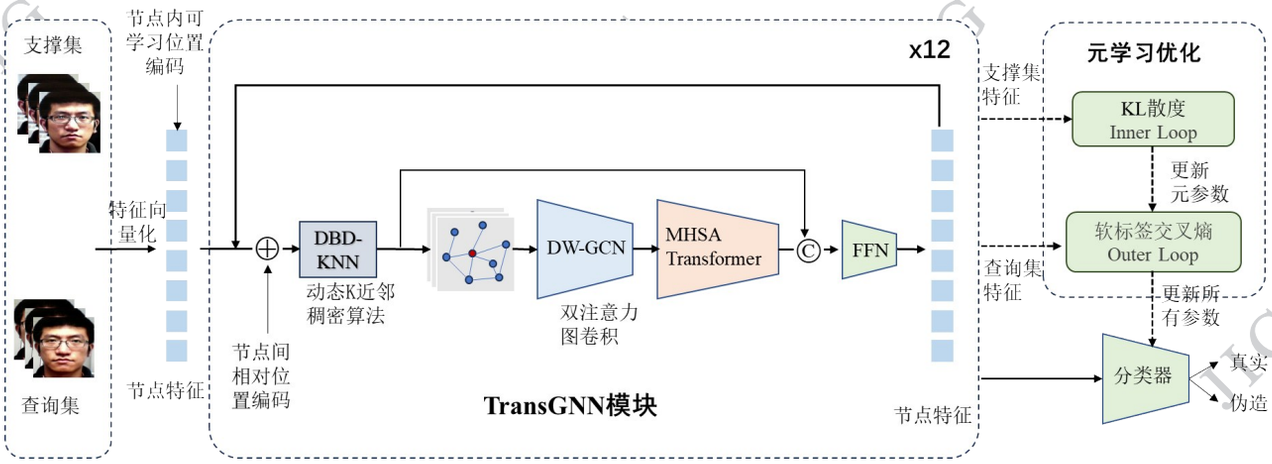


图1 基于Transformer图学习的自适应人脸欺诈检测模型框架

Figure 1 Adaptive face fraud detection model framework based on Transformer graph learning

2.1 由图像到图

不同于传统卷积神经网络，图神经网络只接受图的输入。图定义为$\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{E})$，其中$\boldsymbol{V}$为图中各个顶点的集合，$\boldsymbol{E}$为图中各个边的集合，图可以看成节点和边的一个总集。在图像到图的转化过程中，本文将人脸图像划分为N个不重叠的图像块，然后经过两组Conv-BN-ReLU，将特征向量化，得到特征向量序列$\boldsymbol{V}=\{\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_i,\cdots,\boldsymbol{v}_N\}$。为了保留二维图像的空间信息，根据规则的网格图像特征，本文为每个节点添加了节点内像素点间的可学习位置编码，以及节点间的相对位置编码，产生带位置编码的节点属性。在节点属性中引入位置编码可以丰富节点属性，能显著提升GNN对图像结构信息的捕获能力。

在构造图拓扑结构时，本模型将根据带位置编码节点属性的相似性来采样节点。本文通过改进的动态K近邻稠密算法（DBD-KNN）来控制节点的度（即K值）。传统的KNN算法采用固定的、人工设置的K值，而DBD-KNN算法能够根据数据的平均局部密度自适应地选择最优的K值。DBD-KNN的计算过程如下：首先基于节点属性$\boldsymbol{v}_i$计算得到节点间成对距离，并计算得到每个节点到其他所有节点的平均距离d_i，再根据平均距离计算各节点的逆密度$\rho_i=\frac{1}{d_i}$，并将逆密度进行归一化，然后根据归一化逆密度的中位数$\tilde{\rho}$来确定样本的K值，其计算公式如下：

© 中国图象图形学报版权所有

$$K = round\left(\left(\frac{\tilde{\rho} - \min(\tilde{\rho}_i)}{\max(\tilde{\rho}_i) - \min(\tilde{\rho}_i)}\right) \times k_{diff} + k_{min}\right) \quad (1)$$

式中，$round(\cdot)$表示四舍五入取整，$\max(\tilde{\rho}_i)$和$\min(\tilde{\rho}_i)$分别是最大、最小归一化逆密度，$k_{diff} = k_{max} - k_{min}$是设置的最大、最小K值的差。根据中位数来确定K值可以避免样本的密度过大或过小的风险。

考虑到计算复杂度的问题，本模型并没有为每个样本调整K值，而是为每批样本设置平均值。若一批样本节点的索引集为B，则$\tilde{\rho}$是一批样本中所有节点的归一化逆密度的中位数，$\tilde{\rho} = quantile(\rho_i, 0.5) \quad i \in B$，所计算得到的$K$值是一批样本的自适应节点的度。

2.2 改进的TransGNN特征提取模块

本模型的TransGNN模块交替地堆叠GNN与Transformer层，进行互补增益，GNN通过屏蔽无关节点，将Transformer的盲目全局注意力裁剪为“按图索骥”的稀疏、低秩、结构敏感聚焦；Transformer凭借长距离依赖拓展了GNN的感受野，把信息聚合从显式边缘中解放出来，使消息传递更远、更灵活。

2.2.1 GNN层

本模型的GNN层采用边-点双注意力图卷积来更新节点信息。该图卷积运算中使用了双重注意力加权，即对节点和边分别进行加权。在人脸欺诈检测任务中，一个样本不同区域的图像块对样本标签的影响是不同的，也就是说不同节点对图标签的贡献差异较大，因此，本模型设置了一个可学习权重，对各节点的影响力进行加权。若节点集特征为$\boldsymbol{V} \in \mathbf{R}^{B \times C \times N}$，其中，$B$为批大小，$C$为通道数，$N$为节点数，则节点可学习权重维度为$\boldsymbol{W}_v \in \mathbf{R}^{1 \times 1 \times N}$，它与节点的点乘$\boldsymbol{W}_v \odot \boldsymbol{V}$可得到各节点的影响值。此外，图运算中的边特征是中心节点与邻居节点的差值，中心节点$\boldsymbol{v}_i$的边特征可表示为$\boldsymbol{e}_{ik} = \boldsymbol{v}_k - \boldsymbol{v}_i|_{(\boldsymbol{v}_k \in N(\boldsymbol{v}_i))}$，式中，$N(\boldsymbol{v}_i)$是中心节点$\boldsymbol{v}_i$的邻居节点集。不同邻居对中心节点的重要程度不同，因此，本模型使用softmax函数对边特征进行归一化加权，边权重为：$\boldsymbol{W}_e^{ij} = \frac{\exp(\boldsymbol{a}^T \boldsymbol{e}_{ij})}{\sum_{k \in N(\boldsymbol{v}_i)} \exp(\boldsymbol{a}^T \boldsymbol{e}_{ik})}$，式中，$\boldsymbol{a}$为可学习参数。softmax函数在对边进行加权的同时，还对权重进行了归一化，也就是确保边权重在所有边的维度上总和为1，边-点双注意力图卷积运算可表示成以下公式：

$$V_i' = \sum_{k=1}^{K} softmax(\boldsymbol{a}^T \boldsymbol{W}_v^i \odot (\boldsymbol{v}_k - \boldsymbol{v}_i)) \quad (2)$$

式中，$\boldsymbol{v}_i'$是更新后节点集$\boldsymbol{V}'$的第i个节点特征，$\boldsymbol{v}_k$是$\boldsymbol{v}_i$的邻居节点，$\boldsymbol{W}_v^i$是第i个节点的可学习权重，softmax函数对边进行归一化加权，最后将加权后的边特征进行求和，利用边加权特征的聚合值来更新节点特征。这种归一化加权可以避免随训练过程产生的内部协变量偏移，使每层的数据分布趋于稳定，可加速模型的训练收敛。

2.2.2 Transformer层

GNN具有较强的结构表示能力，然而其对图拓扑结构的依赖性较强，错误的连接将产生较大的偏差；而Transformer的密集注意力反而能“强行”学到一些鲁棒特征。因此，本模型在GNN层后紧跟着Transformer层，采用了多头自注意力（multi-head self-attention，MHSA）对$\boldsymbol{V}'$进一步更新。对第h个注意力头（共$h = 1, \cdots, H$个），先通过权重矩阵$\boldsymbol{W}_h^Q$、$\boldsymbol{W}_h^K$和$\boldsymbol{W}_h^V$把每个$\boldsymbol{V}'$投影到三个不同的子空间，$\boldsymbol{Q}_h = \boldsymbol{V}'\boldsymbol{W}_h^Q$、$\boldsymbol{K}_h = \boldsymbol{V}'\boldsymbol{W}_h^K$和$\boldsymbol{V}_h = \boldsymbol{V}'\boldsymbol{W}_h^V$，基于自注意力机制将信息聚合，得到第$h$个注意力头的特征输出：

$$\boldsymbol{O}_h = softmax\left(\frac{\boldsymbol{Q}_h \boldsymbol{K}_h^{\top}}{\sqrt{d_k}}\right)\boldsymbol{V}_h \quad (3)$$

式中，$d_k = C/H$为每个注意力头的维度。最后，将各头注意力输出特征进行拼接，再做一次线性映射，得到Transformer层的输出特征：

$$\boldsymbol{V}'' = Concat(\boldsymbol{O}_1, \cdots, \boldsymbol{O}_h, \cdots, \boldsymbol{O}_H)\boldsymbol{W}_O \quad (4)$$

式中，$\boldsymbol{W}_O$是线性映射参数，$Concat(\cdot)$为拼接函数。为了抑制梯度消失现象，本模型通过残差跳连接将Transformer层的输出与输入节点信息在通道维进行拼接，最后通过前向传递模型（FFN），得到TransGNN模块的输出：

$$TransGNN(V) = FFN(Concat(\boldsymbol{V}, \boldsymbol{V}'')) \quad (5)$$

每一轮TransGNN模块的动态更新过程包括：①根据带位置编码的节点特征，利用DBD-KNN算法动态更新图拓扑结构；②经过GNN层，根据边-点双注意力图卷积更新节点特征得到$\boldsymbol{V}'$；③送入Transfromer层后，利用多头自注意力机制更新节点特征得到$\boldsymbol{V}''$；④经过残差跳连接和FFN层后，得到TransGNN模块的输出。这样的动态更新过程将经过多

© 中国图象图形学报版权所有

轮，每一轮动态更新将在节点更新和边地更新之间循环往复，金字塔结构使模型能在更高层次上捕获更抽象、更全局的图结构信息。

2.3 TransGNN模块结构

TransGNN模块金字塔结构如图2所示。

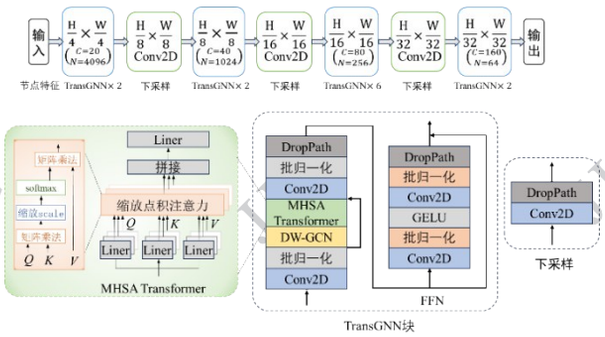


图2 TransGNN模块的结构以及参数设置

Figure 2 Structure and parameter settings of the TransGNN module

图中C表示特征维度，N表示节点数，TransGNN模块后有若干下采样层，使各TransGNN层的特征维度不断增加，而节点数不断减少，形成金字塔结构。每一个TransGNN块由Conv2D卷积、DW-GCN块、MHSA Transformer块、DropPath块以及FFN子块组成，FFN子块由两个Conv2D卷积层及批归一化、GELU激活函数组成。为了提高节点多样性、缓和传统GNN随着层数加深导致的过渡平滑问题，TransGNN模块在多处使用了残差跳连接，显著地提高了模型的训练效率和整体性能。

2.4 软标签目标损失的元学习优化

本文借助元学习方法强大的学习能力来提高模型的适应性。首先将训练样本按λ:1 − λ的比例随机地划分成不重叠的支持集$\boldsymbol{D}_{supp}$和查询集$\boldsymbol{D}_{query}$，分别应用于模型训练的不同阶段。模型训练包含内循环和外循环两阶段优化过程：

2.4.1 内循环训练阶段

内循环是用支持集$\boldsymbol{D}_{supp}$快速更新任务临时参数。为了能快速适应新任务，本文选择KL散度函数(KL Divergence，KLD)做内循环的目标损失函数，因为该损失函数对软标签较为敏感，有助于快速调整模型参数。借鉴MAML++算法思想，内循环采用多步更新的工作方式，第i频内循环的KL散度目标损失的计算公式如下：

$$\boldsymbol{L}_{supp}^{i}\left(\boldsymbol{\theta}_{i-1},\boldsymbol{D}_{supp}\right)=\sum_{m=1}^{M} y_m \log\left(\frac{y_m}{p_m}\right) \tag{6}$$

式中，M是类别数，y_m和p_m分别是真实的和模型预测的第m类的概率。每步内循环后，都用上一步得到的临时参数$\boldsymbol{\theta}_{i-1}$计算当前步损失函数值$\boldsymbol{L}_{supp}^{i}$，然后利用该损失函数值更新任务临时参数，经过$k$步内循环的参数更新，可得到$k$组参数$\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_k$。

2.4.2 外循环训练阶段

外循环应用查询集$\boldsymbol{D}_{query}$来更新元参数$\boldsymbol{\theta}_0$。为了能更好地学习模型参数，本文的外循环选择软标签交叉熵做目标损失函数，该损失函数能够确保模型在各种数据上都能学习到稳定的特征表示。具体做法是把内循环第i步学习得到的临时参数$\boldsymbol{\theta}_i$作用于查询集$\boldsymbol{D}_{query}$，得到第i步的软标签交叉熵损失：

$$\boldsymbol{L}_{query}^{i}\left(\boldsymbol{\theta}_i,\boldsymbol{D}_{query}\right)=-\sum_{m=1}^{M} y_m \log\left(p_m\right) \tag{7}$$

然后把所有步的损失加权求和，计算公式如下：

$$\boldsymbol{L}_{query}=\sum_{i=0}^{k}\boldsymbol{\alpha}_i\cdot\boldsymbol{L}_{query}^{i} \tag{8}$$

式中，k为内循环的更新步数，$\boldsymbol{\alpha}_i$是第i步损失的权重，越靠后的更新步权重越大。最后将查询集损失$\boldsymbol{L}_{query}$反向传播更新原始元参数。

在内循环阶段，通过KL散度损失函数的优化，可以快速调整模型的元参数；再经过外循环阶段的软标签交叉熵损失函数对类别间的相似性和不确定性进行更细腻的描述，进一步精确地更新模型参数。内循环、外循环的这种分离式的学习过程可以避免了同时进行快速适应和精细微调时可能导致冲突的问题，能够使模型在面对新任务时进行快速调整并持续优化，提高了模型的适应性和鲁棒性。

3 实验与分析

3.1 实验数据及实验环境

本文实验所使用的数据集包括物理攻击和数字攻击两大类。物理攻击类使用了OULU-NPU、CASIA-FASD、Replay-Attack以及MSU-MFSD这4个经典数据集。这些数据集主要包含了打印攻击、重放攻击、面具攻击、剪切照片攻击等，是在不同场景下采用固定相机或手持设备录制产生。数字攻击类使用了DFD(Jiang等，2020)、Celeb-DF(Li等，2020)和FaceForensics++(Rossler等，2019)数据集。DFD包含3068个DeepFake技术生成的换脸视频与363

© 中国图象图形学报版权所有

个原始视频。Celeb-DF聚焦高保真伪造检测，基于名人视频生成5639个高质量换脸样本，优化了面部融合细节。FaceForensics++提供1000真实视频和4000伪造视频，使用了DeepFakes、Face2Face、FaceSwap、NeuralTexture和FaceShifter五种典型深度伪造技术。

对于输入视频，本文首先从视频中随机提取图像帧（2帧/秒），并使用MTCNN算法截取人脸图像，再对每个图像独立应用随机旋转、亮度调整、高斯噪声等方法进行预处理，并使用Mixup（Zhang等，2017）或Cutmix（Yun等，2019）方法来进行数据增强，以提升模型的泛化能力。

本文实验使用了NVDIA GeForc RTX 3080显卡，Pytorch框架。模型输入的人脸图像尺寸为256×256×3，使用随机翻转、颜色抖动、随机擦除等方式对数据进行预处理，batch-size大小设置为64，图像块大小默认为4×4，也就是初始的节点数为64×64。为了开展元学习优化，训练集按7∶3的比例划分为支持集和查询集，内循环、外循环训练阶段的学习率皆设置为0.001。模型优化器为Adamw（Wen等，2015），它是将权重衰减（L2正则化）与Adam优化器结合起来的一种优化器，权重衰减值设置为0.05。初始学习率为2e-3，每训练30轮变为原来的0.1倍。

3.2 跨数据集的人脸活体检测对比实验

为了验证本文所提出的方法在传统人脸活体检测的泛化性能，本文使用OULU-NPU、CASIA-FASD、Replay-Attack以及MSU-MFSD4个数据集进行跨数据集测试实验。这4个数据集各自侧重于不同的攻击类型、采集方法和采集环境，这使得跨数据集的模型评估具有一定挑战性。为方便记录实验结果，数据集依次简记为O、C、I和M。参与比较的算法包括LBPTOP（Freitas等，2014）传统纹理方法；MADDG（Shao等，2019）、SSDG（Jia等，2021）与UDG-FAS（Liu等，2023）域泛化方法；RFMeta（Shao等，2020）元学习方法；DLIF（Yang等，2024）特征解耦方法等经典或最新的算法。性能评价指标使用半总错误率（HTER）和特征曲线下面积（AUC），实验结果如表1所示，在O&C&I to M和O&M&I to C实验中，本文模型的HTER并非最优。分析认为，MSU与CASIA数据集中包含的特定攻击类型与训练集分布差异较大，而本模型更侧重于捕捉光照、几何等空间不一致性线索，对这些特定攻击的局部纹理特征敏感性相对不足。然而，在O&C&M to I与I&C&M to O中，本模型取得最优性能，表明其在多数跨域场景下具有卓越泛化能力。

表1 跨数据集人脸活体检测对比实验

Table 1 Cross-dataset face anti-spoofing comparison experiment

Model	O&C&M to I		O&C&I to M		O&M&I to C		I&C&M to O	
	HTER (%)	AUC (%)	HTER (%)	AUC (%)	HTER (%)	AUC (%)	HTER (%)	AUC (%)
LBPTOP	49.45	49.54	36.90	70.80	42.60	61.05	53.15	44.09
MADDG	22.19	84.99	17.69	88.06	24.5	84.51	27.98	80.02
SSDG	18.21	94.61	16.67	90.47	23.11	85.45	25.17	81.83
RFMeta	17.3	90.48	13.89	93.98	20.27	88.16	16.45	91.16
MTML	8.07	96.85	7.38	96.66	13.2	94.27	8.75	95.95
AFNM	16.03	91.04	10.83	96.75	17.85	89.26	15.67	91.90
UDG-FAS	5.86	98.62	5.95	**98.47**	9.82	96.76	10.97	95.36
DLIF	5.82	98.13	**3.75**	98.33	**6.67**	**97.27**	8.89	96.36
本文模型	**3.23**	**98.97**	6.25	96.09	12.5	95.71	**2.78**	**96.77**

注：加粗数据为最优值

© 中国图象图形学报版权所有

3.3 Deepfake检测对比试验

3.3.1 Deepfake数据集内的对比实验

本节将开展Deepfake数据集内部对比实验，以验证本文提出模型的性能。

参与比较的模型包括基于CNN的ResNet（He等，2016）、Xception（Chollet等，2017）、EfficientNet（Tan等，2019）、AWARE-Net（Salman等，2025）；基于Transformer的VIT（Thing等，2023）、Swin（Liu等，2021）；时空频域相结合的MBMD（Long等，2025）方法；元学习优化的MCW（Guan等人，2023）方法以及图神经网络方法DFGNN（Khali等，2023）、SSGT（Khormali等，2024）。数据集为Celeb-DF和FF++（高质量子集），训练集和测试集按7∶3的比例随机分配，性能评价指标使用分类精度（Accuracy，Acc）和特征曲线下面积（AUC）。实验结果如表2所示，由表2可知本文所提出的模型在Celeb-DF数据集的实验中Acc达到98.93%，AUC达到99.81%；在FF++数据集的实验中Acc达到98.44%，AUC达到99.62%。在两个数据集的实验中，本文提出模型的Acc能达到或接近了当前最新的对比方法，说明本文所提出的模型能够识别细微的伪造线索，包括生成图片时产生的压缩伪影、模糊伪影、几何失真或颜色失真等。

3.3.2 在FF++数据集上的跨子集实验

为了验证本文所提出模型的适应性，本节选择FF++高质量集（C23）做数据集内跨子集实验，包括闭集实验和开集实验两种类型。FF++数据集共有五种欺诈子集：DeepFakes（DF）、Face2Face（F2F）、FaceSwap（FS）、NeuralTextures（NT）、FaceShifter（FSh）。借鉴DFGNN，本节的实验分为两类欺诈检测：身份交换和表情交换。在身份交换欺诈检测实验中，选择FaceSwap（FS）、DeepFakes（DF）和FaceShifter（FSh）三个子集做训练集，另外两个子集做开集实验。而在表情交换欺诈检测实验中，选择Face2Face（F2F）和NeuralTextures（NT）两个子集做训练集，另外三个子集做开集实验。

实验结果如表3所示，由表可知，本文所提出模型的检测指标全面超过DFGNN文档的实验结果，平均Acc超过了13.17%，而平均AUC超过了17.52%，在开集实验中表现尤其明显，平均Acc超过了19.84%，而平均AUC超过了25.32%，说明本文提出的模型具有较好的适应性和泛化性。

3.3.3 Deepfake跨数据集对比实验

为了验证本文所提出的模型在Deepfake检测中

表2 deepfake数据集内部对比实验

Table 2 Comparative experiments on a single deepfake dataset

Model	Celeb-DF		FF++	
	Acc	AUC	Acc	AUC
ResNet	96.96%	99.65%	87.04%	99.21%
Xception	97.68%	99.77%	88.32%	99.70%
HRNet	96.47%	99.87%	88.74%	**99.95%**
VIT	87.21%	97.54%	75.73%	92.29%
BEiT	87.51%	95.90%	86.82%	98.76%
Efficient	96.43%	**99.88%**	87.78%	99.45%
Swin	76.58%	98.84%	87.25%	98.81%
MiniGCNs	98.90%	98.00%	95.09%	–
SSGT	–	–	94.59%	95.16%
DFGNN	93.90%	–	97.16%	–
MBMD	99.11%	99.74%	96.97%	99.54%
AWARE	**100%**	99.92%	96.86%	99.22%
本文模型	98.93%	99.81%	**98.44%**	99.62%

注：加粗数据为最优值，“–”表示该数据在原始文献中未提供

© 中国图象图形学报版权所有

表3　FF++交叉评估实验

Table 3　Cross-validation metrics on FF++

检测类型	训练集	测试集		DFGNN		本文模型	
		闭集	开集	Acc	AUC	Acc	AUC
身份交换	FS+DF+FSh	FS		89.30%	92%	93.75%	97.54%
		DF		85.72%	88%	89.06%	95.74%
		FSh		79.23%	82%	89.06%	98.53%
			F2F	73.50%	75%	84.38%	95.37%
			NT	61.30%	69%	87.50%	98.38%
表情交换	F2F+NT	F2F		80.10%	85%	85.93%	95.30%
		NT		86.30%	90%	95.31%	98.49%
			FS	73.10%	75%	78.13%	91.99%
			DF	69.90%	72%	85.94%	91.53%
			FSh	52.70%	59%	93.75%	99.31%

的泛化性能，本节使用FF++高质量集做训练集，DFD或Celeb-DF做测试集，开展了Deepfake跨数据集的对比实验，结果如表4所示。由表可知，本文提出模型的平均性能取得了最好的结果，说明本模型较擅长捕获生成人脸与原始图像混合时产生的皮肤、光照、几何比例、运动模式等全局不一致性，并且即使面对未知的欺诈类型，也具有较好的适应性。

3.4　消融实验

本节通过消融实验来验证各个组件对于整个模型的影响，所选的组件分为DBD-KNN（对比KNN）、元学习优化、TransGNN模块（对比GNN）以及FFN模块，使用I&C&M to O的跨数据集的实验方案，评估指标使用半总错误率（HTER）。消融实验结果如表5所示。由表可知，逐步加入每一个组件，模型的性能都会逐步提高，这表明所有组件都有利于模型性能的提高。其中，TransGNN模块中包含Transformer子块与图卷积子块，对比方案中没有Transformer子块，该组件对模型影响最大，添加Transformer子块后HTER降低了2.78%；其次DBD-KNN根据数据的局部密度调整节点的度，所构建的拓扑图使模型能够更准确地捕捉和表达数据中的复杂关系；元学习策略主要是在训练过程发挥作用，通过优化模型的初始化和快速适应新任务，提高模型的泛化能力和快速适应性；FFN模块使得模型在深层网络中有效地传递梯度以及增强模型对特征的学习能力，其对模型的影响也较大，该模块使模型的HTER降低了2.34%。

表4　Deepfake跨数据集对比实验

Table 4　Comparative experiments on Deepfake across datasets

训练集	模型	测试集			
		DFD		Celeb-DF	
		Acc	AUC	Acc	AUC
FF++	ResNet	58.73%	0.68	67.60%	0.78
	Xception	68.31%	0.65	75.06%	0.84
	HRNet	60.42%	0.62	70.27%	0.78
	VIT	52.39%	0.76	74.67%	0.84
	BEiT	66.47%	0.70	75.95%	0.84
	Swin	51.90%	0.66	66.08%	0.73
	MCW	**81.70%**	**0.82**	64.20%	0.67
	DFGNN	-	-	73.40%	0.76
	SSGT	-	-	-	0.87
	MBMD	-	-	**85.80%**	0.88
	AWARE	-	-	70.14%	0.88
	本文模型	78.13%	**0.82**	84.38%	**0.89**

注：加粗数据为最优值，“-”表示该数据在原始文献中未提供

© 中国图象图形学报版权所有

表5　不同组件对模型的影响

Table 5　Impact of different components on the mode

TransGNN	DBD-KNN	元学习	FNN	HTER(%)
	√	√	√	11.11
√		√	√	8.33
√	√		√	6.45
√	√	√		5.12
√	√	√	√	**2.78**

注：加粗数据为最优值

3.5　动态K值的影响

为了进一步验证DBD-KNN算法的性能，本节使用OULU-NPU做训练集，Replay-Attack做测试集，做进一步的分析实验。图3展示了一轮迭代训练中K值出现的频次直方图，由图可知，本次实验中出现频数最多的K值为9。

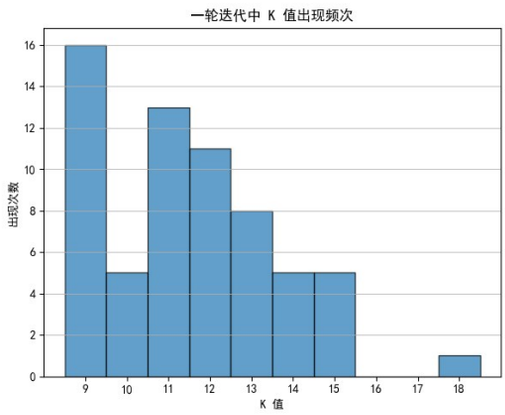


图3　一轮迭代训练中K值出现的频次直方图

Figure 3　Histogram of K values appearing in an epoch training

接着，本节将KNN算法的K值设置为9，然后比较固定K值的KNN算法与动态K值的DBD-KNN算法对模型的影响，所使用的指标是半总错误率(HTER)，结果如表6所示。由表可知：KNN算法即使采用最佳K值，其性能也差于DBD-KNN算法，HTER指标相差5.55%。DBD-KNN算法能够根据数据的密度自适应地选择批样本中的平均最佳K值，从而调整图结构中节点的度，优化的图结构能明显改进模型性能。

3.6　元学习优化中损失函数的选择

在元学习优化过程中，模型在内循环和外循环阶段使用了不同的软标签损失函数：KL散度损失函数和软标签交叉熵损失(soft target cross entropy，STCE)。本节对比使用不同的软标签损失函数的组合效果，以期选择最优的搭配。实验使用了OULU-NPU做训练集，Replay-Attack做测试集，评估指标使用了半总错误率(HTER)。结果如表7所示，由表可知：(1)在内循环选择KLD函数，而外循环选择STCE函数时，实验效果最好，可能的原因是KLD对软标签的敏感性有助于快速调整模型参数，而STCE函数具有较强的容错能力，能够确保模型在各种数据上都能学习到稳定的特征表示，因此它们的组合能得到最优的效果。(2)内循环阶段的损失函数对模型性能的影响是最大的，它将直接影响模型在数据集上的快速适应性，当内循环选择KLD函数时，其实验结果优于选择STCE函数时的结果。

表6　DBD-KNN与KNN算法的对比

Table 6　comparison of DBD-KNN and KNN

方法	KNN	DBD-KNN
HTER(%)	8.33	**2.78**

注：加粗数据为最优值

表7　不同组合软标签损失函数的性能对比

Table 7　Performance comparison of different combinations of soft label loss functions

软标签目标损失	内循环	KLD	KLD	STCE	STCE
	外循环	STCE	KLD	STCE	KLD
性能评价指标	HTER(%)	**4.7**	5.3	6.3	8.5

注：加粗数据为最优值

3.7　模型复杂度与性能分析

为了全面评估本文模型的计算效率与性能，我们将其与VIG，Swin-Transformer，DFGNN等广泛使用的基线模型在FF++数据集内部实验中进行了对比，结果如表8所示。

由表8可知：本文提出的模型在计算效率和参数量上拥有显著优势。由于本文模型在特征维度上远小于其他基线模型，如VIG-TI卷积层最后输出的特征维度为384，DFGNN的特征维度为640，而同网络层本文模型特征维度仅有160，故计算开销与参数量低于基线模型，而在FF++数据集域内实验中依然保持了有竞争力的分类精度。

3.8　模型的可解释性分析

利用图的拓扑结构所具有的直观性，本节展示

© 中国图象图形学报版权所有

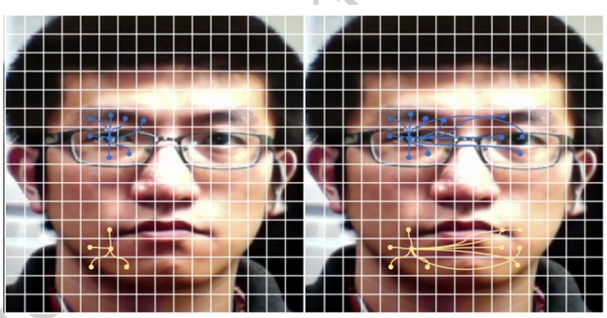
表8　不同模型的复杂度对比

Table 8　Comparison of complexity between different models

模型	参数量(M)	FLOPs (B)	输入尺寸	Acc
DFGNN	27.3	4.6	128×128	97.16%
VIG-TI	7.1	1.3	224×224	94.15
Swin-Transformer	29	4.5	224×224	87.25%
ResNet-50	25.6	4.1	224×224	87.04%
本文模型	5.8	1.18	256×256	**98.44%**

注：加粗数据为最优值

了随着网络层数的加深，图的拓扑结构的演化过程，这一过程直观地揭示了模型做出检测决策的内在逻辑。图4中的(a)和(b)分别展示了进入第1层TransGNN模块和进入第12层TransGNN模块的部分图结构，图中，每个中心节点用相同颜色的曲线连接其邻居节点。由图可知，在网络的浅层阶段，空间位置对图结构影响较大，节点的邻居主要分布在空间相邻的节点间。随着网络深度的增加，节点间的连接不再局限于局部区域，而是开始在整个图像范围内捕捉具有区别特性的语义特征，形成更为广泛的连接。



(a)第1层TransGNN图结构 (b)第12层TransGNN图结构

图4　图拓扑结构的演化

Figure 4　Evolution of Graph Topology Structure

4　结　论

本文提出了一种基于Transformer图学习的自适应人脸欺诈检测模型，该模型结合了GNN强大的结构表示能力和Transformer的全局特征提取能力，增强了模型提取全局空间不一致欺诈线索的能力，提高了模型的泛化性；所改进的动态K近邻稠密算法(DBD-KNN)和元学习优化算法，进一步提高了模型的自适应能力。然而，在欺诈与反欺诈的博弈过程中，更高级的、更多样化的欺诈攻击将始终挑战信息系统的安全性，未来的工作在于：(1)探索如何整合和利用不同模态间的互补信息来增强模型对复杂场景和攻击手段的识别能力，包括视频中图像的空域、时域、频域信息以及音频信息、文字信息、深度信息、生理信号信息、环境信息等；(2)探索利用大语言模型丰富的语义理解和推理能力，辅助并追踪人脸欺诈检测模型决策过程，增强用户交互，提高模型的解释能力和可信度。

参考文献(References)

Afchar D, Nozick V, Yamagishi J and Echizen I. 2018. Mesonet: a compact facial video forgery detection network//Proceedings of 2018 IEEE International Workshop on Information Forensics and Security (WIFS). Hong Kong, China: IEEE: 1-7 [DOI: 10.1109/WIFS.2018.8630787]

Bao H, Dong L, Piao S and Wei F. 2021. Beit: Bert pre-training of image transformers [EB/OL]. [2025-07-22]. https://arxiv.org/abs/2106.08254

Cai T J, Chen F, Liu W, Xie X and Liu Z. 2023. Face anti-spoofing via conditional adversarial domain generalization. Journal of Ambient Intelligence and Humanized Computing, 14(12): 16499-16512 [DOI: 10.1007/s12652-022-03884-z]

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE: 1800-1807 [DOI: 10.1109/CVPR.2017.195]

Chuang C C, Wang C Y and Lai S H. 2023. Generalized face anti-spoofing via multi-task learning and one-side meta triplet loss//Proceedings of 2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG). Waikoloa, HI, USA: IEEE: 1-8 [DOI: 10.1109/FG57933.2023.10042751]

Freitas Pereira T D, Komulainen J, Anjos A, De Martino J M, Hadid A, Pietikäinen M and Marcel S. 2014. Face liveness detection using dynamic texture. EURASIP Journal on Image and Video Processing, 2014(1): 2 [DOI: 10.1186/1687-5281-2014-2]

Guan L, Liu F, Zhang R, Liu J and Tang Y. 2023. MCW: a generalizable deepfake detection method for few-shot learning. Sensors, 23(21): 8763 [DOI: 10.3390/s23218763]

He K, Zhang X, Ren S and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Jia Y, Zhang J, Shan S and Chen X. 2020. Single-side domain general-

© 中国图象图形学报版权所有

ization for face anti-spoofing//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 8484-8493 [DOI: 10.1109/CVPR42600.2020.00851]

Jiang L, Li R, Wu W, Qian C and Loy C C. 2020. Deeperforensics-1.0: a large-scale dataset for real-world face forgery detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE: 2889-2898 [DOI: 10.1109/CVPR42600.2020.00296]

Khalid F, Javed A, Ilyas H and Irtaza A. 2023. DFGNN: an interpretable and generalized graph neural network for deepfakes detection. Expert Systems with Applications, 222: 119843 [DOI: 10.1016/j.eswa.2023.119843]

Khormali A and Yuan J. 2024. Self-supervised graph transformer for deepfake detection. IEEE Access, 12: 58114-58127 [DOI: 10.1109/ACCESS.2024.3392512]

Li G, Muller M, Thabet A and Ghanem B. 2019. DeepGCNs: can GCNs go as deep as CNNs? //Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9267-9276 [DOI: 10.1109/ICCV.2019.00936]

Li Y, Yang X, Sun P, Qi H and Lyu S. 2020. Celeb-DF: a large-scale challenging dataset for deepfake forensics//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE: 3204-3213 [DOI: 10.1109/CVPR42600.2020.00327]

Liu Y, Chen Y, Gou M, Huang C T, Wang Y, Dai W and Xiong H. 2023. Towards unsupervised domain generalization for face anti-spoofing//Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 20654-20664 [DOI: 10.1109/ICCV51070.2023.01888]

Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B, et al. 2021. Swin transformer: hierarchical vision transformer using shifted windows//Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE: 10012-10022 [DOI: 10.1109/ICCV48922.2021.00986]

Long M, Yin X, Zhang L B and Peng F. 2025. Multi-branch network based on multi-domain feature fusion for deepfake detection. Journal of Image and Graphics, 30(1): 1-18 (龙敏, 尹曦, 张乐冰, 彭飞. 2025. 基于多域特征融合的多分支网络用于Deepfake检测. 中国图象图形学报, 30(1): 1-18) [DOI: 10.11834/jig.240681]

Mandal D, Medya S, Uzzi B and Aggarwal C. 2022. Meta-learning with graph neural networks: methods and applications. ACM SIGKDD Explorations Newsletter, 23(2): 13-22 [DOI: 10.1145/3468507.3468515]

Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J and Nießner M. 2019. FaceForensics++ : learning to detect manipulated facial images//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 1-11 [DOI: 10.1109/ICCV.2019.00009]

Salman M, Tariq I, Zulfiqar M, Jalal M, Aujla S and Fatima S. 2025. AWARE-NET: Adaptive Weighted Averaging for Robust Ensemble Network in Deepfake Detection//Proceedings of IET International Conference on Secure and Intelligent Computing. London, UK: IET: 526-533 [DOI: 10.1049/icp.2025.1162]

Shang Z, Xie H, Yu L, Zha Z and Zhang Y. 2023. Constructing spatio-temporal graphs for face forgery detection. ACM Transactions on the Web, 17(3): 1-25 [DOI: 10.1145/3580515]

Shao R, Lan X, Li J and Yuen P C. 2019. Multi-adversarial discriminative deep domain generalization for face presentation attack detection//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE: 10023-10031 [DOI: 10.1109/CVPR.2019.01026]

Tan M and Le Q. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, California, USA: PMLR: 6105-6114

Thing V L. 2023. Deepfake detection with deep learning: convolutional neural networks versus transformers//Proceedings of 2023 IEEE International Conference on Cyber Security and Resilience (CSR). Venice, Italy: IEEE: 246-253 [DOI: 10.1109/CSR57506.2023.10225004]

Wen D, Han H and Jain A K. 2015. Face spoof detection with image distortion analysis. IEEE Transactions on Information Forensics and Security, 10(4): 746-761 [DOI: 10.1109/TIFS.2015.2400395]

Xie X H, Bian J T and Lai J H. 2022. Review on face liveness detection. Journal of Image and Graphics, 27(01): 0063-0087 (谢晓华, 丁锦堂, 赖剑煌. 2022. 人脸活体检测综述. 中国图象图形学报, 27(01): 0063-0087) [DOI: 10.11834/jig.210470]

Yao W D, Li P C, Zhao Y and Wu H C. 2025. Review of research on face deepfake detection methods. Journal of Image and Graph- ics, 30(7): 2343-2363 (姚文达, 李盼池, 赵娅, 吴洪超. 2025. 人脸深度伪造检测方法研究综述. 中国图象图形学报, 30(7): 2343-2363) [DOI: 10. 11834/jig. 240586]

Yang J, Yu Z, Ni X, He J and Li H. 2024. Generalized face anti-spoofing via finer domain partition and disentangling liveness-irrelevant factors//Proceedings of ECAI 2024. Santiago de Compostela, Spain: IOS Press: 274-281 [DOI: 10.3233/FAIA240289]

Yu Z, Cai R, Li Z, Yang W, Shi J and Kot A C. 2024. Benchmarking joint face spoofing and forgery detection with visual and physiological cues. IEEE Transactions on Dependable and Secure Computing, 21(5): 4327-4342 [DOI: 10.1109/TDSC.2024.3374532]

Yun S, Han D, Oh S J, Chun S, Choe J and Yoo Y. 2019. CutMix: regularization strategy to train strong classifiers with localizable features//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6023-6032 [DOI: 10.1109/ICCV.2019.00612]

Zhang H, Cisse M, Dauphin Y N and Lopez-Paz D. 2017. Mixup: beyond empirical risk minimization [EB/OL]. [2025-07-22].

© 中国图象图形学报版权所有

https://arxiv.org/abs/1710.09412

Zhang P, Yan Y, Zhang X, Li C, Wang Y, Huang F and Kim S. 2024. TransGNN: harnessing the collaborative power of transformers and graph neural networks for recommender systems//Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington, DC, USA: ACM: 1285-1295 [DOI: 10.1145/3626772.3657769]

作者简介

蔡体健,1968年生,女,副教授,博士,硕士导师,主要研究方向为计算机视觉、深度学习等。

黄远轩,2001年生,男,硕士。主要研究方向为深度学习和人脸欺诈检测。

王振宇,2001年生,男,硕士。主要研究方向为深度学习和Deepfake检测。

胡成,2002年生,男,硕士。主要研究方向为深度学习和Deepfake检测。

易晟权,2000年生,男,硕士。主要研究方向为深度学习和人脸活体检测。

谢昕,1969年生,男,教授,硕士。主要研究方向为深度学习、异常检测等。

© 中国图象图形学报版权所有