

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2024)09-2764-16

论文引用格式: Sun J T and Shen W Z. 2024. An iris feature-encoding method by fusion of graph neural networks and convolutional neural networks. Journal of Image and Graphics, 29(09):2764-2779(孙金通, 沈文忠. 2024. 图神经网络与CNN融合的虹膜特征编码方法. 中国图象图形学报, 29(09):2764-2779)[DOI:10.11834/jig.230688]

图神经网络与CNN融合的虹膜特征编码方法

孙金通, 沈文忠*

上海电力大学电子与信息工程学院, 上海 201306

摘要: 目的 更具可解释性的虹膜特征编码方法一直是虹膜识别中的一个关键问题,且低质量虹膜样本识别比较困难,图神经网络的发展为此类虹膜图像特征编码带来了新思路。本文提出了一种图神经网络与卷积神经网络融合的虹膜特征编码网络 IrisFusionNet。方法 在骨干网络前添加一个像素级增强模块以消除输入图像不确定性,并使用双分支骨干网络提取虹膜微观与宏观融合特征。训练阶段使用一个独特的联合损失函数对网络参数进行优化;推理阶段使用融合特征匹配策略进行特征匹配。结果 实验结果表明:使用 IrisFusionNet 训练得到的特征提取器在多个公开低质量虹膜数据集上进行测试分别得到了 EER (equal error rate) 和 FAR@FRR = 0.01% 的最佳值 0.27% 与 0.84%,并且将分离度 DI (discriminating index) 提升 30% 以上,识别准确性以及类聚性均远远领先于基于卷积神经网络和其他使用图神经网络模型的虹膜识别优秀算法。结论 本文提出的 IrisFusionNet 应用于虹膜识别任务具有极高的可行性和优越性。

关键词: 虹膜特征编码;图神经网络 (GNN);硬图注意力算子;特征融合;联合损失函数

An iris feature-encoding method by fusion of graph neural networks and convolutional neural networks

Sun Jintong, Shen Wenzhong*

School of Electronic and Information Engineering, Shanghai University of Electric Power, Shanghai 201306, China

Abstract: Objective Iris recognition is a prevalent biometric feature in identity recognition technology owing to its inherent advantages, including stability, uniqueness, noncontact modality, and live-body authentication. The complete iris recognition workflow comprises four main steps: iris image acquisition, image preprocessing, feature encoding, and feature matching. Feature encoding serves as the core component of iris recognition algorithms. The improvement in interpretable iris feature encoding methods have become a pivotal concern in the field of iris recognition. Moreover, the recognition of low-quality iris samples, which often relies on specific parameter-dependent feature encoders, results in a poor generalization performance. The graph structure represents a data form with an irregular topological arrangement. Graph neural networks (GNNs) effectively update and aggregate features within such graph structures. The advancement of GNN led to the development of new approaches for feature encoding of these types of iris images. In this paper, a pioneering iris feature-fusion encoding network called IrisFusionNet, which integrates GNN with a convolutional neural network (CNN), is proposed. This network eradicates the need to implement complex parameter tuning steps and exhibits excellent generalization performance across various iris datasets. **Method** In the backbone network, the previously inserted pixel-level enhancement module alleviates local uncertainty in the input image through median filtering. In addition, global uncertainty was

收稿日期: 2023-09-28;修回日期: 2023-12-26;预印本日期: 2024-01-02

* 通信作者: 沈文忠 7841426@qq.com

mitigated via Gaussian normalization. A dual-branch backbone network was proposed, where the head of the backbone network comprised a shared stack of CONV modules, and the neck was divided into two branches. The primary branch constructed a graph structure from an image using graph converter. We designed a hard graph attention network that introduces an efficient channel attention mechanism to aggregate and update features through utilization of edge-associated information within the graph structure. This step led to the extraction of microfeatures of iris textures. The auxiliary branch, on the other hand, used conventional CNN pipeline components, such as simple convolutional layers, pooling layers, and fully connected layers, to capture the macrostructural information on the iris. During the training phase, the fused features from the primary and auxiliary branches were optimized using a unique unified loss function graph triplet and additive angular margin unified loss (GTAU-Loss). The primary branch mapped iris images into a graph feature space, with the use of cosine similarity to measure semantic information in node feature vectors, L2 norm to measure the spatial relationship information within the adjacency matrix, and graph triplet loss to constrain feature distances within the feature space. The auxiliary branch applied an additional angular margin loss, which normalized the input image feature vectors and introduced an additional angular margin to constrain feature angle intervals, which improved intraclass feature compactness and interclass separation. Ultimately, a dynamic learning method based on an exponential model was used to fuse the features extracted from the primary and auxiliary branches and obtain the GTAU-Loss. The hyperparameter settings during training included the following: The optimization of network parameters involved the use of stochastic gradient descent (SGD) with a Nesterov momentum set to 0.9, an initial learning rate of 0.001, and a warm-up strategy adjusting the learning rate with a warm-up rate set to 0.1, conducted over 200 epochs. The iteration process of SGD was accelerated using NVIDIA RTX 3060 12 GB GPU devices, with 100 iterations lasting approximately one day. For feature matching concerning two distinct graph structures, the auxiliary branch calculated the cosine similarity between the output node features. Meanwhile, the primary branch applied a gate-based method and initially calculated the mean cosine similarity of all node pairs as the threshold for the gate, removed node pairs below this threshold, and retained node features above it to compute their cosine similarity. The similarity between these graph structures was represented as the weighted sum of cosine similarities from the primary and auxiliary branches. The similarity weights of the feature pairs computed using the primary and auxiliary branches were both set to 0.5. All experiments were conducted on a Windows 11 operating system, with PyTorch as the deep learning framework. **Result** To validate the effectiveness of integrating GNNs into the framework, this study conducted iris recognition experiments using a single-branch CNN framework and a dual-branch framework. The experimental outcomes substantiated the superior recognition performance involved in the structural design incorporating the GNN branch. Furthermore, to determine the optimal values for two crucial parameters, namely, the number of nearest neighbors (k) and the global feature dimension within the IrisFusionNet framework, we conducted detailed parameter experiments to determine their most favorable values. k was set to 8, and the optimal global feature dimension was 256. We compared the present method with several state-of-art (SOTA) methods in iris recognition, including CNN-based methods, such as ResNet, MobileNet, EfficientNet, ConvNext, etc., and GNN-based methods, such as dynamic graph representation. Comparative experimental results indicate that the feature extractor trained using IrisFusionNet, which was tested on three publicly available low-quality iris datasets — CASIA-Iris-V4-Distance, CASIA-Iris-V4-Lamp, CASIA-Iris-Mobile-V1.0-S2—to achieve equal error rates of 1.06%, 0.71%, and 0.27% and false rejection rates at a false acceptance rate of 0.01% (FRR@FAR = 0.01%) of 7.49%, 4.21%, and 0.84%, respectively. In addition, the discriminant index reached 6.102, 6.574, and 8.451, which denote an improvement of over 30% compared with the baseline algorithm. The accuracy and clustering capability of iris recognition tasks using the feature extractor derived from IrisFusionNet substantially outperformed SOTA iris recognition algorithms based on convolutional neural networks and other GNN models. Furthermore, the graph structures derived from the graph transformer were visually displayed. The generated graph structures of similar iris images exhibited a high similarity, and those of dissimilar iris images presented remarkable differences. This intuitive visualization explained the excellent performance achieved in iris recognition by constructing graph structures and utilizing GNN methods. **Conclusion** In this paper, we proposed a feature fusion coding a method based on GNN (IrisFusionNet). The macro features of iris images were extracted using the CNN and the micro features of iris images were extracted using GNNs to obtain fusion features encompassing comprehensive texture characteristics. The experimental results indi-

cate that our method considerably improved the accuracy and clustering of iris recognition and obtained a high feasibility and generalizability without necessitating complex parameter tuning specific to particular datasets.

Key words: iris feature coding; graph neural network (GNN); hard graph attention operators; feature fusion; unified loss function

0 引言

虹膜识别是一种基于虹膜丰富纹理特征进行身份确认的生物识别技术,作为最准确、最可靠的身份认证方法之一,具有稳定性、唯一性、非接触和活体识别等优点,并已广泛应用于公安、金融、医疗和住宅管理等诸多领域(王蕴红等,2002)。

完整的虹膜识别流程一般包括虹膜图像采集、图像预处理、特征编码和特征匹配4个步骤,其中,特征编码是虹膜识别算法的核心部分。目前应用较广泛的虹膜特征提取方法是由Daugman开创的技术(Daugman,2009),因其具有计算复杂度低、较高匹配效率和准确率的特点占据了虹膜识别技术主导地位。在Daugman开创性研究的启发下,近10年来,离散余弦变换(discrete cosine transform, DCT)(Monro等,2007)、离散傅立叶变换(discrete Fourier transform, DFT)(Miyazawa等,2008)和序数度量(Sun和Tan,2009)等多种特征描述方法在虹膜识别任务上均表现出不俗的性能。但是这些方法难以区分虹膜的微观特征与宏观特征的重要信息,并且对采集到的虹膜图像有较高约束要求,困难样本识别效果也一直不理想。

近些年来,卷积神经网络(convolutional neural network, CNN)的强大特征提取能力对虹膜识别的发展起到了强力推动:Gangwar和Joshi(2016)提出DeepIris,首次将CNN应用于虹膜识别领域,其对CNN应用于虹膜识别的一系列问题(CNN网络输入尺寸、CNN网络规模、不同虹膜分割方法的影响、跨数据集测试等)进行了详细实验和讨论;Zhao和Kumar(2017)将基于全卷积网络(fully convolutional networks, FCN)的UniNet模型应用于虹膜识别,在不同数据集上进行实验,表现出了较好的泛化性能;Gangwar等人(2019)提出了一个用于表达可见光和近红外光虹膜图像的CNN框架DeepIrisNet2,应用Spatial Transformer对虹膜图像进行了数据增强,并提出了一种二进制编码策略;陈英等人(2023)融合

虹膜与眼周特征,基于CNN提出了一种鲁棒性更强的虹膜特征表征方法。尽管CNN能够捕捉虹膜图像纹理特征,但难以表达局部特征间的关联。

图结构数据是一种广泛存在且具有非规则拓扑结构的数据形式,因其复杂性和异质性,图结构数据的分析一直是备受关注的问题(岳含霖等,2022)。图神经网络(graph neural network, GNN)是图域上的深度学习方法,因其强大的表征能力,广泛应用于图分析任务(侯磊等,2023),常用的GNN包括图卷积神经网络(graph convolutional network, GCN)(Kipf和Welling,2016)、图注意力网络(graph attention network, GAT)(Veličković等,2018)、GraphSAGE(graph sample and aggregate)(Hamilton等,2018)和DeepGCNs(deep graph convolutional networks)(Li等,2023)等。现实生活中许多实际任务都可以抽象成由节点、边、全局特征组成的图结构数据,在视觉领域,图像也不例外。Ren等人(2020)提出了一种动态图表示方法(dynamic graph representation, DGR),在遮挡虹膜识别任务中取得了一定效果;Han等人(2022)通过patch-merging与KNN(K nearest neighbors)方法将图像表示为图结构数据,提出面向视觉任务的GNN骨干网络Vision GNN,在ImageNet数据集图像分类任务中取得了超越CNN与Transformer的效果;Han等人(2023)放弃传统基于pairwise进行图构建的方法,使用超图(hypergraph)方法封装图像信息,提出ViHGNN(vision hypergraph neural network),在ImageNet数据集图像分类任务及COCO2017(common objects in context 2007)数据集目标检测任务中均取得SOTA(state of the arts)效果。

为探索效果更好的虹膜识别方法,本文将GNN作为主分支结构并添加CNN作为辅分支,同时对虹膜图像宏观特征与微观特征进行编码,提出了虹膜特征融合编码方法IrisFusionNet。具体来说,本文主要贡献概括如下:1)针对虹膜图像采集因素引起的不确定性(如采集设备情况、环境因素、采集距离和采集者姿态等),设计像素级图像增强模块,通过局部增强和全局增强,来减少噪声对识别效果的影响。

2)首次将硬图注意力机制应用于虹膜识别方法,提出了一种全新的双分支特征融合骨干网络。其中主分支的图转换器和结合高效通道注意力机制的硬图注意力网络从图像中构建图结构中节点与边并实现图信息提取,辅分支通过传统卷积方法获得全局特征,利用图结构不规则拓扑性,使局部变化对融合特征在识别中的干扰最小化。3)提出一个独特的联合损失函数 GTAU-Loss(graph triplet and additive angular margin unified loss)。主分支 graph triplet loss 与辅分支 AAM loss(additive angular margin loss)动态学习策略对网络进行联合优化,通过扩大特征空间类间间距,保持模型识别准确指标的同时,显著提升聚类效果。4)在3个公开低质量虹膜数据集上进行详尽的虹膜识别实验。实验结果表明,IrisFusionNet在虹膜识别任务上取得极佳准确性和泛化性,同时通过虹膜图结构可视化,解释了本方法能够取得极佳效果的原因。

1 算法设计

本节将通过IrisFusionNet中的基本模块来搭建模型结构,并对损失函数以及匹配策略进行详尽分析。

1.1 算法模型结构

图1为IrisFusionNet网络结构示意图。归一化

虹膜图像首先经过像素级增强(pixel-level enhancement, PLE)模块,以消除采集因素对特征编码产生的不确定性影响,而后通过由3个基本块(basic block)与1个卷积层组成的卷积模块(Conv module),基本块结构如图2所示。通过卷积模块后,每一幅尺寸 $128 \times 256 \times 1$ 的归一化虹膜图像被映射为 $16 \times 32 \times 96$ 的feature map,此后经历以下两个分支:

1)主分支以卷积模块输出的feature map作为输入,图转换器(graph converter)从feature map中产生图结构,图结构中每个节点都包含一个特征向量,边的权值表示节点之间的关系,称为feature graph。feature graph被送入ECA-HGAT,通过高效通道注意力层(efficient channel attention layer, ECA layer)和硬图注意力层(hard graph attention layer, HGAT layer)构建图神经网络更新图结构中的节点特征,具体内容将在1.3节中介绍。

2)辅分支由一个basic block和两个全连接层组成,与卷积模块共同构成了一个常规的CNN管道,以提取输入图像的全局特征,表1中给出了IrisFusionNet中不同基本块的参数设置。

值得注意的是,如图1所示,IrisFusionNet采用不同层的输出得到融合特征,训练阶段传入损失函数用于优化网络参数,以及推理阶段提取虹膜编码特征用于计算相似度分数。

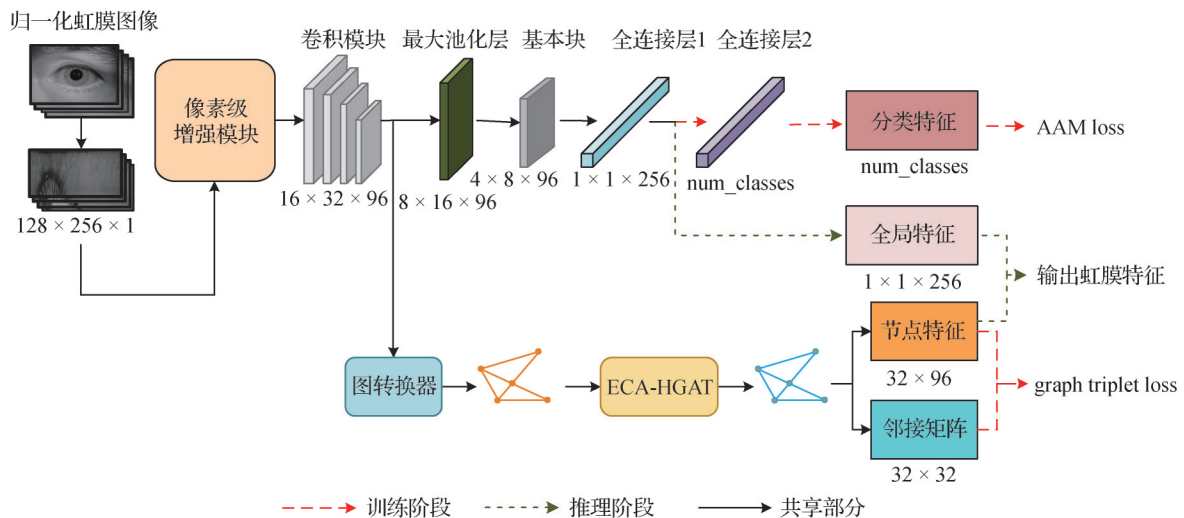


图1 IrisFusionNet模型结构

Fig. 1 Model structure of IrisFusionNet

1.1.1 像素级增强模块

在实际应用场景中,诸如采集设备参数以及采集环境变换等未知采集因素,会引起采集图像与实

际图像之间的不确定性差异。对于虹膜识别系统来说,不确定性差异导致的虹膜纹理扭曲通常会严重威胁识别性能。为了解决这一问题,本文设计了一

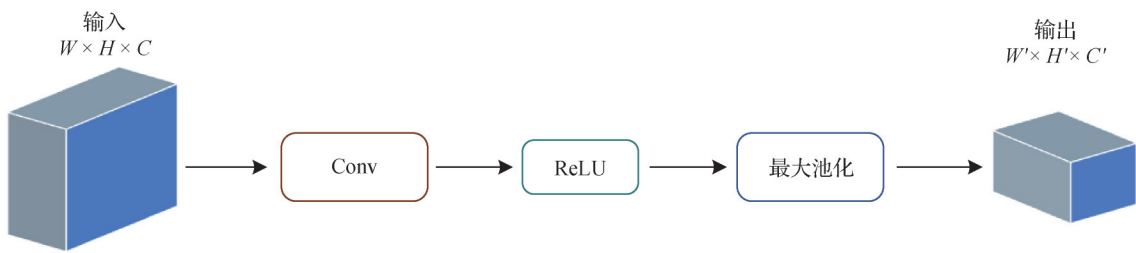


图2 基本块结构
Fig. 2 Structure of basic block

表1 不同 basic block 中的参数细节

Table 1 Details of parameters in different basic blocks

名称	Type/kernel/stride/ padding	输出尺寸
basic block0	Conv0	Conv/5×9/1/2×4
	MaxPool0	MaxPool/2/2/0
basic block1	Conv1	Conv/5×7/1/2×3
	MaxPool1	MaxPool/2/2/0
basic block2	Conv2	Conv/5×5/1/2×2
	MaxPool2	MaxPool/2/2/0
basic block3	Conv3	Conv/5×5/1/2×2
	MaxPool3	MaxPool/2/2/0
basic block4	Conv4	Conv/5×5/1/2×2
	MaxPool4	MaxPool/2/2/0

个像素级增强(pixel-level enhancement, PLE)模块,对每个归一化虹膜图像进行局部特征和全局统计量的增强,来缓解这类不确定性因素带来的影响。

对于局部区域,计算该区域的中值作为中心位置像素,窗口内的非线性平滑操作能够从局部角度减轻了图像噪声不确定性。为了避免过度平滑,训练中采用3×3大小滑动窗口来进行中值滤波。

由于采集过程的不可重复性,每幅图像都有其特定亮度和对比度(Xiao等,2013)。对于全局不确定性,PLE模块根据整幅归一化图像的一阶和二阶统计信息对图像进行校正,具体过程为:设 X_i 为宽为 H ,高为 W 的第 i 个归一化图像,得到增强图像的过程具体为

$$X_i^E(h,w)=\frac{X_i(h,w)-X_i^M}{X_i^{SD}} \tag{1}$$

式中, $X_i^M=\frac{\sum_{h,w}X_i(h,w)}{H\times W}$ 和 $X_i^{SD}=\sqrt{\frac{\sum_{h,w}[X_i(h,w)-X_i^M]^2}{H\times W}}$

分别为一阶和二阶统计信息, h 和 w 分别表示水平和垂直方向位置。

1. 1. 2 图转换器

图结构数据可以通过表达语义信息的节点特征与表达空间关系信息的邻接矩阵所表示,因此,如图3所示,图转换器通过以下两个步骤将feature map转换为feature graph。

1)获得节点特征。首先将feature map通过空间

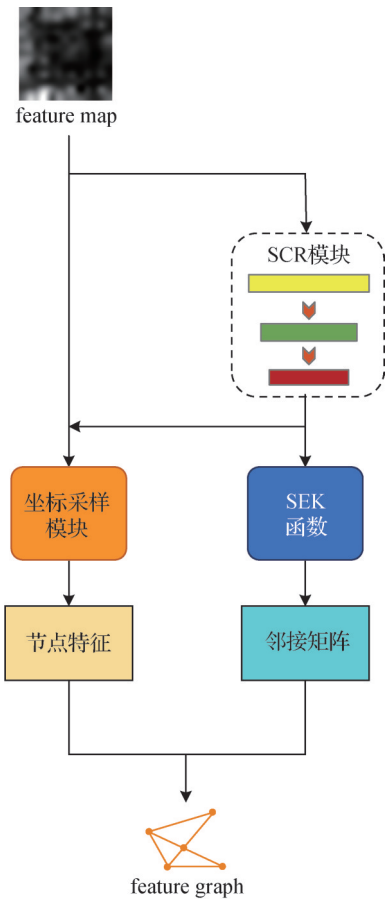


图3 图转换器结构
Fig. 3 Structure of graph converter

坐标回归(spatial coordinates regression, SCR)模块, 该模块由两个 basic block 和两个全连接层组成, 输入 $W \times H \times C$ 的 feature map 经回归转换为 $N \times 2$ 的空间坐标。根据二维空间坐标, 通过双线性插值从特征图中采样每个节点的特征向量。具体为

$$f_{\text{COORD}} = \text{SCR}(\mathbf{F}^M) \quad (2)$$

式中, $\mathbf{F}^M$ 为 feature map, $f_{\text{COORD}} \in \mathbf{R}^{N \times 2}$ 为二维空间坐标, SCR 表示空间坐标回归模块内运算。

图特征向量表示为

$$\mathbf{F}_{(i,j)}^C(c) = \text{BILINEAR}(\mathbf{F}^M(i, h, c), f_{\text{COORD}}) \quad (3)$$

式中, $\mathbf{F}^C$ 表示图特征向量, i, j 为节点的二维空间坐标, c 为每个特征向量所在通道索引, $c \in \{1, 2, \dots, C\}$, C 为 feature map 的通道数。

2) 生成邻接矩阵。本文使用基于空间距离指数核(spatial exponential kernel, SEK)函数, 根据节点空间位置生成邻接矩阵, 具体为

$$\mathbf{A}(a, b) = \begin{cases} 0 & d_{\text{coord}} > \theta \\ \exp(-d_{\text{coord}}^2) & d_{\text{coord}} \leq \theta \end{cases} \quad (4)$$

式中, $\mathbf{A}$ 为 $N \times N$ 的邻接矩阵, $N = 32$ 表示节点数,

$d_{\text{coord}} = \|\text{coord}_a - \text{coord}_b\|_2$ 为两个节点之间绝对距离, $\text{coord} \in \mathbf{R}^2$ 为节点二维空间坐标, θ 为节点感受野尺度。若输入图像上两个节点感受野重叠, 则它们之间会产生边, 重叠区域越大, 边的权值就越大。

经过上述两步, SCR 模块选取 feature map 中最重要的位置构成 feature graph 的节点, SEK 函数则通过节点感受野来建立边, 从而将 $W \times H \times C$ 的 feature map 转换为包含节点特征 $\mathbf{F} \in \mathbf{R}^{N \times C}$ 与邻接矩阵 $\mathbf{A} \in \mathbf{R}^{N \times N}$ 的 feature graph。

1.1.3 ECA-HGAT

为了实现节点特征高效聚合更新, 本文提出一个新的图注意力网络 ECA-HGAT。ECA-HGAT 中存在两个基本层: 高效通道注意力层(ECA layer)与硬图注意力层(HGAT layer)。

1) ECA layer。通道注意力机制已经被证明在提升 CNN 性能方面具有显著效果(Liu 等, 2021), 其中 ECA-Net(Wang 等, 2020)作为轻量级模块插入模型中取得了模型效果提升, 因此本文将 ECA 模块推广到图神经网络中, 使用不降维的局部节点跨通道信息交互策略, 设计出 ECA layer。如图 4 所示。

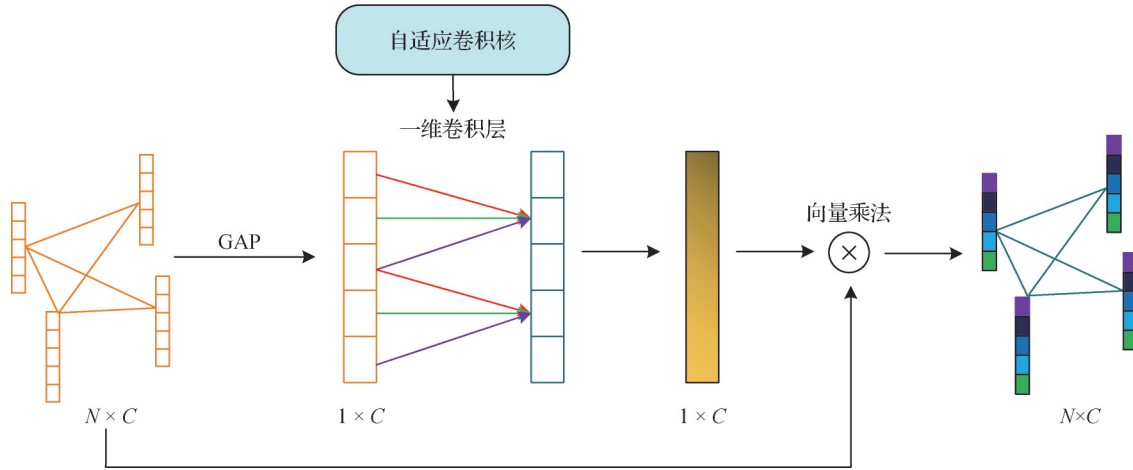


图4 ECA layer结构

Fig. 4 Structure of ECA layer

首先, 通过全局平均池化(global average pooling, GAP)计算每个通道的平均能量。具体为

$$\mathbf{E}(c) = f_{\text{GAP}}(\mathbf{G}) = \frac{1}{N} \sum_{i=1}^N f_i(c) \quad (5)$$

式中, $\mathbf{E} \in \mathbf{R}^C$ 表示平均能量向量, $\mathbf{f} \in \mathbf{R}^C$ 为通道特征向量, $\mathbf{G}$ 表示 feature graph, N 为节点数, $c \in \{1, 2, \dots, C\}$ 为每个特征向量所在通道的索引。然后通过一个一维卷积(one-dimension convolution,

Conv1D)层可以实现局部通道间信息交互, 同时为了高效地进行通道间通信, 由输入通道数经过一个自适应映射(adaptive mapping of kernel size) $\psi(c)$ 得到合适的卷积核大小, 具体为

$$\mathbf{I}(c) = f_{\text{CID}_k}(\mathbf{E}) \quad (6)$$

式中, $\mathbf{I}(c)$ 表示经过局部通道通信后的特征向量, $f_{\text{CID}_k}(\cdot)$ 表示卷积核大小为 k 的一维卷积运算。

$$k = \psi(c) = \left\lceil \frac{\log_2(C) + 1}{2} \right\rceil_{\text{odd}} \quad (7)$$

式中, $|A|_{\text{odd}}$ 表示取不小于 A 的最小奇数。最后通过向量乘法实现跨通道信息交互, 具体为

$$f'_i(c) = \delta(I(c)) \times f_i(c) \quad (8)$$

式中, f'_i 为跨通道融合特征向量, $\delta(\cdot)$ 为非线性激活函数, 此处使用 ReLU (rectified linear unit), i 是节点的索引, $i \in \{1, 2, \dots, N\}$ 。

2) HGAT layer。在图结构数据的任务中, 图注意力算子 (graph attention operator, GAO) 是常用的学习相邻节点信息权重的方法 (Shankar 等, 2018)。然而, GAO 属于软注意力, 而非硬注意力, 后者只关注最重要的特征, 从而产生更佳表现 (Yu 等, 2021)。因此, IrisFusionNet 对于 feature graph 使用硬图注意力 (hard graph attention, HGAT) 层进行节点特征汇聚更新, HGAT 层为硬图注意力算子 (hard graph attention operator, HGAO) (Gao 和 Ji, 2019) 的变体, 对于输入 feature graph 中的第 i 个节点, 通过一个可训练的投影向量 $p \in \mathbf{R}^c$ 从其一阶邻居节点中选出 k 个最重要的节点, 该层计算为

$$y = \frac{|p \times F^T|}{\|p\|_2} \quad (9)$$

对于 feature graph 中任一节点 i , 有

$$idf_i = \text{Ranking}_k(A_{:,i} \otimes y) \quad (10)$$

$$\hat{F}_i = F(idf_i, :) \quad (11)$$

$$\hat{y}_i = \delta_1(y(idf_i)) \quad (12)$$

$$\tilde{F}_i = \text{diag}(\hat{y}_i) \times \hat{F}_i \quad (13)$$

$$s_i = f_i \times \text{Sum}(\text{softmax}(\delta_2(\tilde{F}_i \times \tilde{F}_i^T \times A_{idf_i,i}))) \quad (14)$$

$$F' = [s_1, s_2, \dots, s_N] \quad (15)$$

结合上述公式, HGAT layer 工作原理表达如下: 式 (9) 计算 F 在 p 上的绝对标量投影, 得到 $y = [y_1, y_2, \dots, y_N]^T$, 其中每个 y_i 用于度量节点 i 的重要性。对于每个节点 i , 式 (10) 中 $A_{:,i}$ 表示 A 的第 i 列, $\otimes$ 表示逐元素矩阵乘法, $\text{Ranking}_k(\cdot)$ 操作会根据节点 i 在 y 中投影值对其一阶相邻节点进行排序, 并选择投影值最大的 k 个节点。令节点 i 所选节点索引为 $idf_i = [i_1, i_2, \dots, i_k]$, 则节点 i 仅对这 k 个节点进行信息汇聚, 而不是所有相邻节点。在式 (11) 中, 通过 idf_i 得到节点 i 相关的新节点特征 $\hat{F}_i =$

$[F_{i_1}, F_{i_2}, \dots, F_{i_k}]$, $F(idf_i, :)$ 表示由 idf_i 索引出 F 的行子集。在式 (12) 中, 使用门操作来控制信息流, 即通过对选定的标量投影值 $y(idf_i)$ 应用 sigmoid 函数来获得门向量 $\hat{y}$, $\delta_1(\cdot)$ 表示 sigmoid 函数。通过在式 (13) 中的矩阵乘法 $\text{diag}(\hat{y}_i) \times \hat{F}_i$, HGAT 能够控制选定节点信息, 使投影向量 p 梯度具有反向传播可训练性。在式 (14) 中通过注意力算子计算节点 i 的响应 s_i , $\text{Sum}(\cdot)$ 表示将输入内所有元素求和, $A_{idf_i,i} \in \mathbf{R}^k$ 表示邻接矩阵 A 第 i 列根据 idf_i 索引得到的子序列, 最后通过式 (14) 构造输出特征矩阵 F' 。值得注意的是, p 针对 F 中所有节点是共享的。

3) ECA-HGAT。ECA-HGAT 的框架如图 5 所示。ECA-HGAT 采用残差结构, 共有 4 层, 其中两层为 ECA 层, 其余为 HGAT 层, 在特征降维层之前将输出添加到输入中。特征降维层对节点特征向量进行降维并添加非线性, 具体为

$$f'_i = \delta(f_i \times W_{\text{rd}}) \quad (16)$$

式中, $W_{\text{rd}} \in \mathbf{R}^{C \times C'}$ 为参数矩阵, C' 为输出特征维度, $\delta(\cdot)$ 为 sigmoid 函数。ECA-HGAT 最终输出 32×96 的节点特征向量和 32×32 的邻接矩阵

1.2 损失函数

为了保证虹膜识别保持较高准确率的情况下,

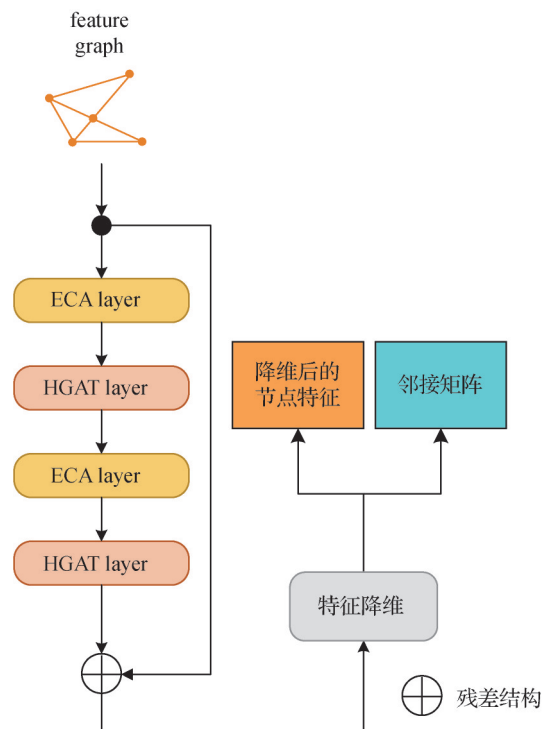


图5 ECA-HGAT结构

Fig. 5 Structure of ECA-HGAT

更好地提升网络聚类效果,本文提出了一个适用于 IrisFusionNet 的损失函数 GTAU-Loss。GTAU-Loss 通过图三元组损失(graph triplet loss)与附加角裕度损失(additive angular margin loss, AAM loss)对网络进行联合优化,1.2.1节与1.2.2节中对图三元组损失和附加角裕度损失分别进行了详细的介绍。

1.2.1 图三元组损失

IrisFusionNet 的主分支输出中包含两种信息:节点特征向量中的语义信息和邻接矩阵中的空间关系信息。因此,两个特征图(feature graph)相似度度量由两部分组成,具体为

$$S(\mathbf{G}_1, \mathbf{G}_2) = \alpha \times S_{\text{feature}}(\mathbf{G}_1, \mathbf{G}_2) + \beta \times S_{\text{adjacency}}(\mathbf{G}_1, \mathbf{G}_2) \quad (17)$$

式中, $\mathbf{G}_1$ 和 $\mathbf{G}_2$ 表示特征图, S_{feature} 表示语义相似度, $S_{\text{adjacency}}$ 为关系相似度, α 和 β 分别表示 S_{feature} 和 $S_{\text{adjacency}}$ 的加权系数, 训练中 $\alpha = \beta = 1$ 。对于语义相似度, 采用余弦相似度进行度量; 对于关系相似度, 采用 L2 范数进行度量, 具体为

$$S_{\text{feature}}(\mathbf{G}_1, \mathbf{G}_2) = \frac{\mathbf{F}_1 \times \mathbf{F}_2^T}{\|\mathbf{F}_1\|_2 \times \|\mathbf{F}_2\|_2} \quad (18)$$

$$S_{\text{adjacency}}(\mathbf{G}_1, \mathbf{G}_2) = \|\mathbf{A}_1 - \mathbf{A}_2\|_2 \quad (19)$$

式中, $\mathbf{F}_1, \mathbf{F}_2, \mathbf{A}_1, \mathbf{A}_2$ 分别表示特征图 $\mathbf{G}_1$ 与 $\mathbf{G}_2$ 的节点特征与邻接矩阵。在人脸识别中, Schroff 等人(2015)提出了 triplet loss 思想, 以增大类内类间区分度, 具体为

$$\|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^p)\| + m < \|f(\mathbf{x}_i^a) - f(\mathbf{x}_i^n)\| \quad (20)$$

式中, $f(\mathbf{x}_i^a), f(\mathbf{x}_i^p)$ 和 $f(\mathbf{x}_i^n)$ 分别表示选定样本(anchor sample)、选定样本同类样本(positive sample)和选定样本异类样本(negative sample)特征, m 为最小特征距离, 作用是防止所有特征聚合至一个极小的特征空间中。IrisFusionNet 主分支将虹膜样本映射至图特征空间, 参考 triplet loss 的思想, 对于选定图 $\mathbf{G}_{\text{anc}}$, 其同类图和异类图分别表示为 $\mathbf{G}_{\text{pos}}$ 和 $\mathbf{G}_{\text{neg}}$, 损失优化目标应满足式(2), 具体为

$$D(\mathbf{G}_{\text{anc}}, \mathbf{G}_{\text{pos}}) + d < D(\mathbf{G}_{\text{anc}}, \mathbf{G}_{\text{neg}}) \quad (21)$$

式中, $D(\cdot)$ 表示图特征空间中的距离, d 为最小图特征距离。将式(21)进行变形, 得到最终图三元组损失为

$$L_{\text{GT}} = \max\{M + S(\mathbf{G}_{\text{anc}}, \mathbf{G}_{\text{pos}}) - D(\mathbf{G}_{\text{anc}}, \mathbf{G}_{\text{neg}}), 0\} \quad (22)$$

式中, $S(\mathbf{G}_{\text{anc}}, \mathbf{G}_{\text{pos}})$ 表示选定样本与正类样本之间的

相似度, $S(\mathbf{G}_{\text{anc}}, \mathbf{G}_{\text{pos}})$ 表示选定样本与负类样本之间的相似度, $M = 0.5$ 表示正类边际。

1.2.2 附加角裕度损失

网络最后一个全连接层的输出 $\mathbf{W}^T \mathbf{X} + b$ 为 softmax 的输入, softmax loss 可表达为

$$L_{\text{S1}} = -\frac{1}{P} \sum_{i=1}^P \ln \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{x}_i + b_{y_i})}{\sum_{j=1}^n \exp(\mathbf{W}_j^T \mathbf{x}_i + b_j)} \quad (23)$$

式中, $\mathbf{W}_{y_i}^T \mathbf{x}_i + b_{y_i}$ 表示输出预测类别为真实类别, $\mathbf{W}_{y_i}^T, \mathbf{x}_i$ 和 b_{y_i} 分别为权重向量、特征向量和偏置值, P 表示每批训练样本数, n 表示样本类别数。令式(23)中偏置 $b = 0$, 内积转化为余弦形式, 并将权重 $\mathbf{W}$ 和特征 $\mathbf{X}$ 的 L2 范数归一化, 式(23)可以表示为

$$L_{\text{S2}} = -\frac{1}{P} \sum_{i=1}^P \ln \frac{\exp(\cos(\theta_{y_i}))}{\exp(\cos(\theta_{y_i})) + \sum_{j=1}^n \exp(\cos(\theta_{y_j}))} \quad (24)$$

式中, θ_{y_i} 表示输出特征与相应类中心权重向量之间的夹角。由于原始 softmax 中未考虑优化使得类内具有高度相似性而类间具有显著差异性。虹膜识别中经常面临很多由姿势、光照等引起类内差异大类间差异小的问题, 因此为了提升区分非同类虹膜的能力, IrisFusionNet 在辅分支中使用了附加角度边界的损失函数 AAM loss (Deng 等, 2019), 添加附加角裕度约束后, AAM loss 表达式为

$$L_{\text{AAM}} = -\frac{1}{P} \sum_{i=1}^P \ln \frac{\exp(s \cos(\theta_{y_i} + m))}{\exp(s \cos(\theta_{y_i} + m)) + \sum_{j=1}^n \exp(\cos(\theta_{y_j}))} \quad (25)$$

式中, s 为缩放因子, m 为角度间隔。不同清晰度图像特征范数不尽相同, AAM loss 对输入特征向量进行归一化, 在训练时仅注重从数据中得到的角度信息, 无需考虑特征向量值大小, 同时增加了低清晰度图像权重, 使得低清晰度图像获得网络更多注意力, 进行反向传播时, 低清晰度图像会产生更大梯度, 因而有利于网络对纹理特征高度相似、难以区分的困难样本进行学习。并且始终存在 m 大小的角度决策边界, 使得类内特征更聚拢、类间间距更大。

1.2.3 GTAU-Loss

固定决策边界会使模型训练过程变得困难, 为了提升模型训练效率, 本文制定了动态学习策略: 在训练初期主要通过主分支图三元组损失 L_{GT} 使模型获得图特征提取能力, 后续学习过程中逐步添加辅

分支 AAM loss, 即 L_{AAM} 以提高模型类内聚合能力。GTAU-Loss 可表示为

$$L_{GTAU} = L_{GT} + W_f \times L_{AAM} \quad (26)$$

$$W_f = (\sqrt[E]{2})^e - 1 \quad (27)$$

式中, L_{GTAU} 表示训练总损失, $W_f \in [0, 1]$ 为动态学习的指数权重函数。本文采用指数函数模型设计动态学习的指数权重函数, $e \in [0, E]$ 表示当前训练的 epoch, E 为训练总 epoch 数。最终 GTAU-Loss 可表示为

$$L_{GTAU} = L_{GT} + ((\sqrt[E]{2})^e - 1) \times L_{AAM} \quad (28)$$

1.3 特征匹配策略

为了充分利用双分支网络提取融合特征中的关键信息, 本文在推理阶段中提出一种针对融合特征的特征匹配策略, 具体来说:

1) 对于辅分支提取出的全局特征, 采用余弦相似度匹配方法;

2) 对于主分支提取出的图节点特征, 应用基于门控的特征匹配方法: 对于 G_A 和 G_B 两个特征图, 分别计算对应节点余弦相似度均值作为门的阈值 τ , 具体为

$$\tau = \frac{1}{N} \sum_{i=1}^N \frac{f_i^A \times f_i^B}{\|f_i^A\|_2 \times \|f_i^B\|_2} \quad (29)$$

式中, $f_i \in \mathbf{R}^C$ 为第 i 个对应节点特征向量, N 为每个特征图中的节点数, C 为节点特征向量长度。若一个节点对相似度小于 τ , 则从两个特征图中分别删除这两个节点, 即在匹配过程中不计算这些相似度较低的节点, 从而对困难图像进行准确识别。

3) 最终采用主分支与辅分支的余弦相似度之积作为最终相似度分数, 计算式为

$$Score_{img} = Score_{sup} \times Score_{main} \quad (30)$$

$$Score_{sup} = \frac{f_{glo}^A \times f_{glo}^B}{\|f_{glo}^A\|_2 \times \|f_{glo}^B\|_2} \quad (31)$$

$$Score_{main} = \frac{1}{N_\tau} \sum_{i=1}^{N_\tau} \frac{f_i^A \times f_i^B}{\|f_i^A\|_2 \times \|f_i^B\|_2} \quad (32)$$

式中, f_{glo} 表示辅分支最后一个全连接层之前输出的全局特征, N_τ 表示经过门节点筛选后图中用于特征匹配的节点数。

2 实验评估

本节将详细介绍 IrisFusionNet 的虹膜识别实验

细节(包括实验平台及环境、数据集、评价指标、预处理和训练策略等), 并通过实验结果及分析证明 Iris-FusionNet 网络模型的性能优势。

2.1 实验细节

2.1.1 实验平台及环境

实验基于 Windows 11 操作系统, GPU 型号为 Nvidia GeForce RTX 3060 12 GB, 深度学习框架为 PyTorch。初始学习率设置为 10^{-3} , 采用 warm-up 预热方法调整学习率, warm-up rate 设置为 0.1, 共训练 200 个 epoch。优化器设置为随机梯度下降(stochastic gradient descent, SGD)算法, 使用 Nesterov 动量, 动量参数设置为 0.9。

2.1.2 数据集

实验选择 3 个识别困难的公开数据集 CASIA-Iris-V4-Distance、CASIA-Iris-V4-Lamp 和 CASIA-Iris-Mobile-V1.0-S2 (分别简称为 Distance、Lamp 和 M1S2) 进行虹膜识别实验。

1) Distance。该数据集(中国科学院自动化研究所, 2020)的图像由距离受试者 3 m 远的设备采集, 包含有来自 142 个对象的 2 446 个脸部样本, 每个样本包含有两只眼睛, 由于采集距离较远, 虹膜纹理清晰度偏低。

2) Lamp。该数据集(中国科学院自动化研究所, 2020)内图像包含来自 819 个对象的 16 212 幅眼部样本, 采集时在受试者附近打开或关闭一盏灯, 由此在不同照明条件下因瞳孔大小变化而引入纹理特征的弹性形变。

3) M1S2。该数据集(中国科学院自动化研究所, 2018)内图像由移动设备于不同距离下进行采集, 它包含来自 400 个对象的 6 000 幅脸部样本, 每个样本包含双眼, 移动设备镜头分辨率较弱, 因而图像质量较低。

训练与测试时均使用预处理后的归一化虹膜图像作为网络输入, 特别地, 在训练时对于原始人脸/人眼图像, 采用极坐标系下不同初始角度进行虹膜归一化的线下数据增强策略来扩增数据集, 提升模型鲁棒性。在剔除了未进行预处理或处理困难的样本后, 得到用于本实验的 3 个数据集有效训练集与有效测试集, 详细统计信息如表 2 所示。

2.1.3 预处理

为了公平比较, 实验对 3 个数据集进行了相同的虹膜定位、虹膜分割和虹膜归一化方法, 具体如下。

表2 不同数据集的统计信息
Table 2 Statistics of different datasets

数据集	有效训练集	有效测试集	类内对数	类间对数
Distance	7 674	2 399	5 230	2 871 171
Lamp	24 369	7 725	140 600	59 527 300
M1S2	14 391	5 132	59 986	13 106 160

1)虹膜定位。使用MTCNN(multi-task convolutional neural network)(滕童等,2020)对眼睛所在区域进行定位,并裁剪出虹膜感兴趣区域(region of interest,ROI)区域。

2)虹膜归一化。使用IrisCodeNet(贾丁丁和沈文忠,2022)中的IrisNormalized方法进行虹膜归一化。

另外,由于Distance与M1S2数据集原始图像为包含两只眼睛的脸部图像,因此,原始人脸图像切分为只包含单幅眼睛的图像,再进行上述预处理方法,最终得到尺寸为 128×256 像素的归一化图像作为网络输入,以Distance数据集中的图像为例,预处理过程如图6所示。

2.1.4 评价指标

在评价时,采用等错误率(equal error rate,EER)

与 $FRR@FAR = 0.01\%$ (在错误接受率(false acceptance rate,FAR)为0.000 1时,错误拒绝率(false rejection rate,FRR)的大小)作为主要评价指标。此外本文还采用分离度(discriminating index,DI)来统计类内类间距离分离度,分离度越大,类内分布越紧凑,类间分布越分散,分类效果就越好。分离度DI的计算式为

$$DI = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\delta_1^2 + \delta_2^2}{2}}} \quad (33)$$

式中, μ_1 和 μ_2 分别表示类内和类间相似度分布的均值, δ_1^2 和 δ_2^2 分别表示类内和类间相似度分布的方差。

2.1.5 训练策略

由式(25)可以得知,AAM loss中存在 s 和 m 两个参数, m 影响模型决策边界, s 则影响模型训练过程中的梯度。为了加速并稳定损失函数的优化过程,根据Deng等人(2019)的参数设置,IrisFusionNet训练中取 $s = 30, m = 0.5$ 。训练时使用全部左眼图像,右眼图像则在测试时使用。另外,本文在训练阶段设置网络模型ECA-HGAT中 k 值为8,辅分支输出的全局特征维度为256,相关实验分析将在2.2.1节中给出。

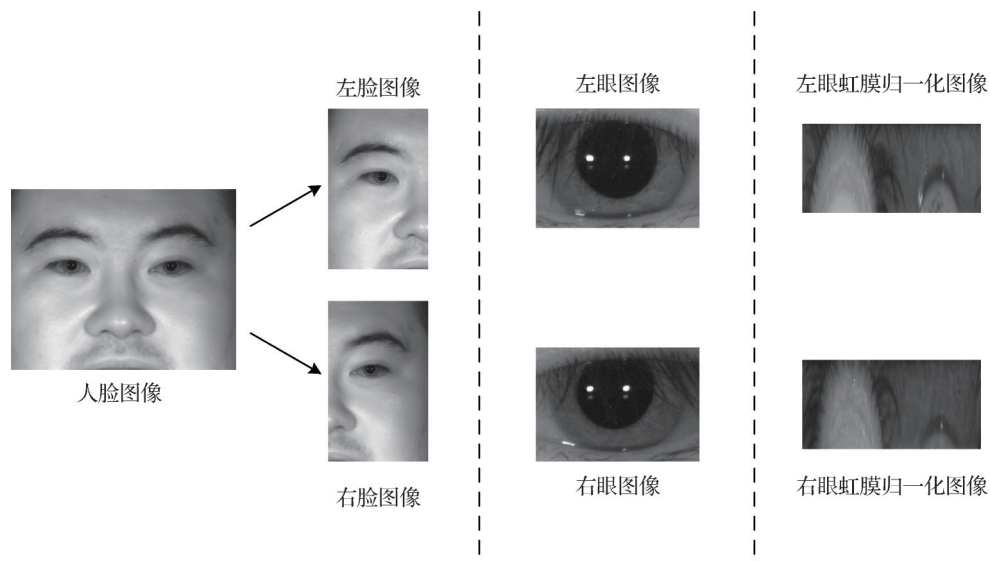


图6 Distance数据集预处理流程

Fig. 6 Preprocessing flow of Distance dataset

2.2 实验结果

2.2.1节通过实验说明引入GNN的有效性;2.2.2节中进行实验以确定最佳参数;2.2.3节中与

一些深度学习中的SOTA方法进行了虹膜识别对比实验,并给出了图结构可视化结果;2.2.4节中给出了消融实验结果。

2.2.1 特征编码和匹配实验

本节在 Distance、Lamp 和 M1S2 三个数据集上,采用相同的 AAM loss,对包含 GNN 分支的 IrisFusionNet(IFN)以及仅由 IFN 辅分支构成的 CNN 网络分别训练,得到特征编码器进行虹膜识别实验,结果如表 3 所示。表 3 中 CNN 表示仅由 IFN 辅分支构成的

常规 CNN 网络,结果显示仅使用 CNN 网络的虹膜特征编码器在低质量样本上识别准确性远远低于引入 GNN 的融合特征编码器,这是因为图结构非拓扑性能够在纹理变化后保持局部特征间联系,很好地抵抗其影响,再经过 GNN 进行节点特征提取,得到图像的关键信息。

表 3 IrisFusionNet 与辅分支 CNN 网络虹膜识别结果
Table 3 Iris recognition results of IrisFusionNet and CNN structure in auxiliary branch

方法	Distance 数据集		Lamp 数据集		M1S2 数据集	
	EER	FRR@FAR = 0.01%	EER	FRR@FAR = 0.01%	EER	FRR@FAR = 0.01%
CNN	3.73	31.32	2.55	17.00	1.42	10.84
IFN	1.06	7.49	0.71	4.21	0.27	0.84

注:加粗字体表示各列最优结果。

2.2.2 参数实验

本节中在 Lamp 数据集上对 IrisFusionNet 中两个重要参数进行实验探究,确定最佳取值。

1)近邻数 k 。ECA-HGAT 中近邻数 k 是影响节点特征聚合效果的重要超参数,实验结果如表 4 所示,当 k 值过小时,在邻居节点信息汇聚过程中丢失了部分重要信息,以至于无法达到较好精度;另一方面,如果 k 值过大,会使与选定节点间无联系或联系较小的邻居节点参与注意力运算,导致汇聚更新后的节点特征中包含一部分分散注意力的信息,从而影响识别效果。因此,本文在模型中取 $k = 8$ 。

2)全局特征维度。在 CNN 架构为基础的虹膜特征编码方法中,特征维度决定了虹膜识别效果。本节设置特征维度分别为 128、256 和 512,实验结果如表 5 所示。结果显示,全局特征维度为 128、512 时模型识别性能均不及全局特征维度为 256 时。经过分析,虹膜特征纹理蕴含丰富信息,故特征维度较小时无法完整表达纹理信息;而另一方面节点和边内已经包含了大部分虹膜特征信息,过大的特征维度会使网络对于节点与边的信息利用弱化,导致识别效果下降。因此,本文采用辅分支全局特征维度为 256 作为最佳特征维度。

2.2.3 虹膜识别实验结果

1)与 CNN 方法的对比实验

本节在 Distance、Lamp 和 M1S2 三个数据集上与基于 CNN 的特征编码方法进行对比实验,包括

表 4 Lamp 数据集上的不同 k 值识别结果
Table 4 Recognition results for different k on Lamp dataset

k	EER	FRR@FAR = 0.01%
4	0.008 0	0.050 1
8	0.007 1	0.042 1
16	0.008 4	0.049 2

注:加粗字体表示各列最优结果。

表 5 Lamp 数据集上的不同全局特征维度识别结果
Table 5 Recognition results for different global feature dimensions on Lamp dataset

特征维度	EER	FRR@FAR = 0.01%
128	0.024 5	0.189 4
256	0.007 1	0.042 1
512	0.015 4	0.044 4

注:加粗字体表示各列最优结果。

ResNet18 (He 等, 2016)、MobileNetV3 (Howard 等, 2019)、EfficientNet (Tan 和 Le, 2019) 和 ConvNext (Liu 等, 2022)、所有 CNN 网络均使用 AAM loss 作为损失函数,训练中 $s = 30, m = 0.5$ 。

实验结果如表 6 和图 7 所示,IFN 特征融合编码方法在 3 个数据集上进行虹膜识别的 EER 与 FRR@FAR = 0.01% 均高于基于 CNN 的方法,展现出极佳泛化性和鲁棒性。

为了更加直观地说明使用 IrisFusionNet 进行虹

膜识别效果卓越的原因,图 8 中展示了 Lamp 数据集中 3 幅虹膜图像的图结构可视化结果。从图中可以看到,节点位置避开了眼睑等干扰因素,并且类内图

像生成的图结构十分相似,类间图像生成的图结构则存在很大区别,虹膜关键特征提取更精确,识别准确性更高。

表 6 在 Distance、Lamp 和 M1S2 数据集虹膜识别的实验结果
Table 6 Results of iris recognition experiments on Distance, Lamp and M1S2 datasets

方法	Distance 数据集		Lamp 数据集		M1S2 数据集	
	EER	FRR@FAR = 0.01%	EER	FRR@FAR = 0.01%	EER	FRR@FAR = 0.01%
ResNet18	4.35	30.44	4.29	21.82	4.99	24.98
MobileNetV3	1.95	10.78	2.26	19.41	2.39	21.03
EfficientNet	2.04	10.45	1.86	7.77	0.68	4.09
ConvNext	2.02	9.81	2.18	8.02	1.51	7.55
IrisFusionNet	1.06	7.49	0.71	4.21	0.27	0.84

注:加粗字体表示各列最优结果。

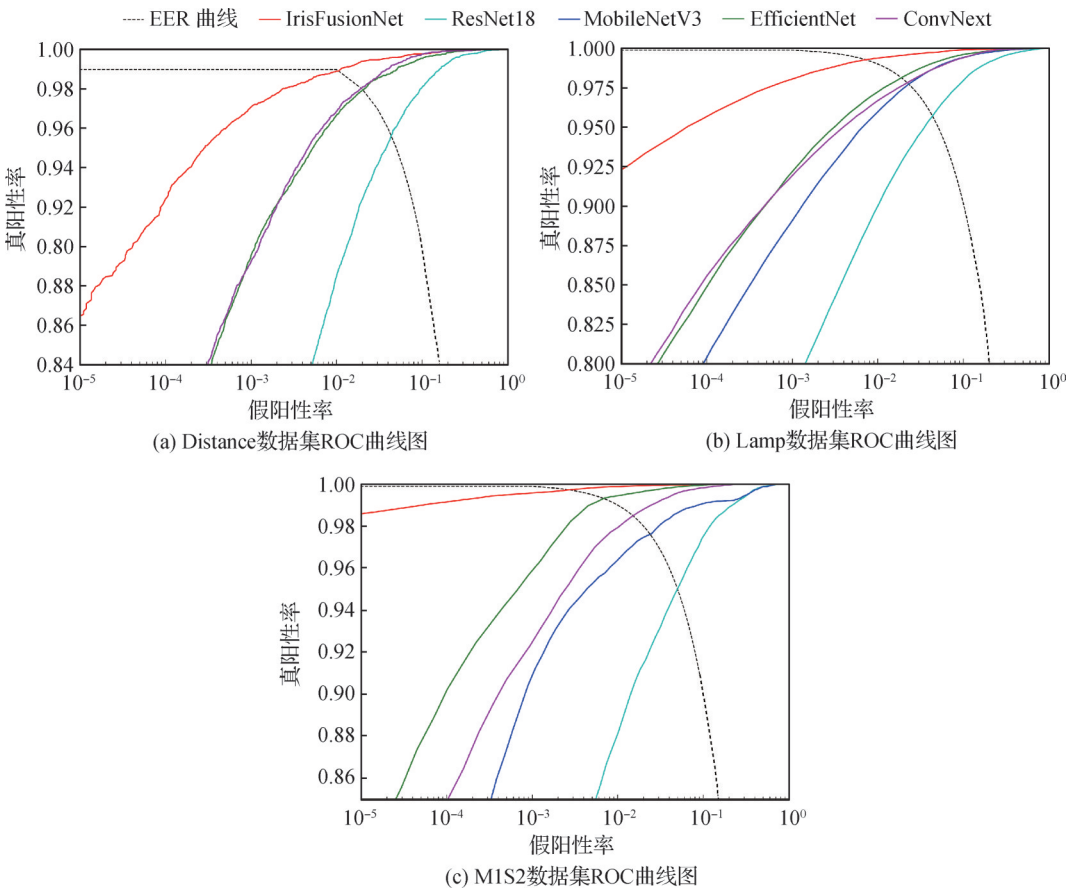


图 7 虹膜识别实验 ROC 曲线图

Fig. 7 ROC curves of iris recognition experiment ((a) ROC curves of iris recognition experiments for Distance dataset;
(b) ROC curves of iris recognition experiments for Lamp dataset;
(c) ROC curves of iris recognition experiments for M1S2 dataset)

2)与SOTA方法的对比实验

DGR (dynamic graph representation) (Ren 等,

2020)是首个将 GNN 应用于生物特征识别的工作,在虹膜识别任务中取得了 SOTA 效果。本节

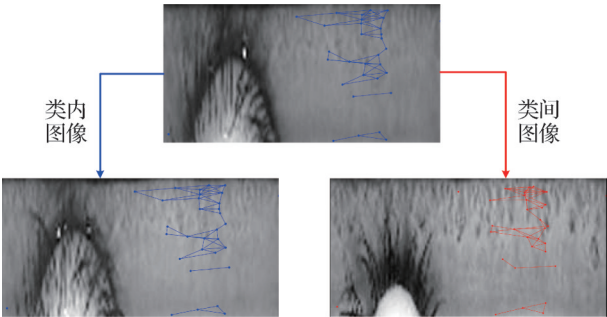


图8 Lamp数据集虹膜样本的Graph可视化
Fig. 8 Graph visualization of iris sample in Lamp dataset

设置 DGR 作为基准算法 (baseline), 分别在 Distance、Lamp 和 M1S2 三个数据集上进行了对比实验, 实验结果如表 7 所示, IrisFusionNet (IFN) 相比于基准算法, 准确性指标大幅提升, 达到了 SOTA 效果。

除准确性指标外, 本节统计了两种方法识别结果的类内类间分布。图 9 展示了基准算法与本文方法的相似度分布直方图, IrisFusionNet 在 3 个数据集上分离度 DI 均大于基准算法。

表 7 SOTA 方法虹膜识别实验结果对比

Table 7 Comparison of the SOTA methods for iris recognition experiments

方法	Distance 数据集			Lamp 数据集			M1S2 数据集		
	EER	FRR@FAR=0.01%	DI	EER	FRR@FAR=0.01%	DI	EER	FRR@FAR=0.01%	DI
DGR	1.41%	11.70%	4.106	1.05%	6.36%	3.428	0.59%	2.21%	6.471
IFN	1.06%	7.49%	6.102	0.71%	4.21%	6.574	0.27%	0.84%	8.451

注: 加粗字体表示各列最优结果。

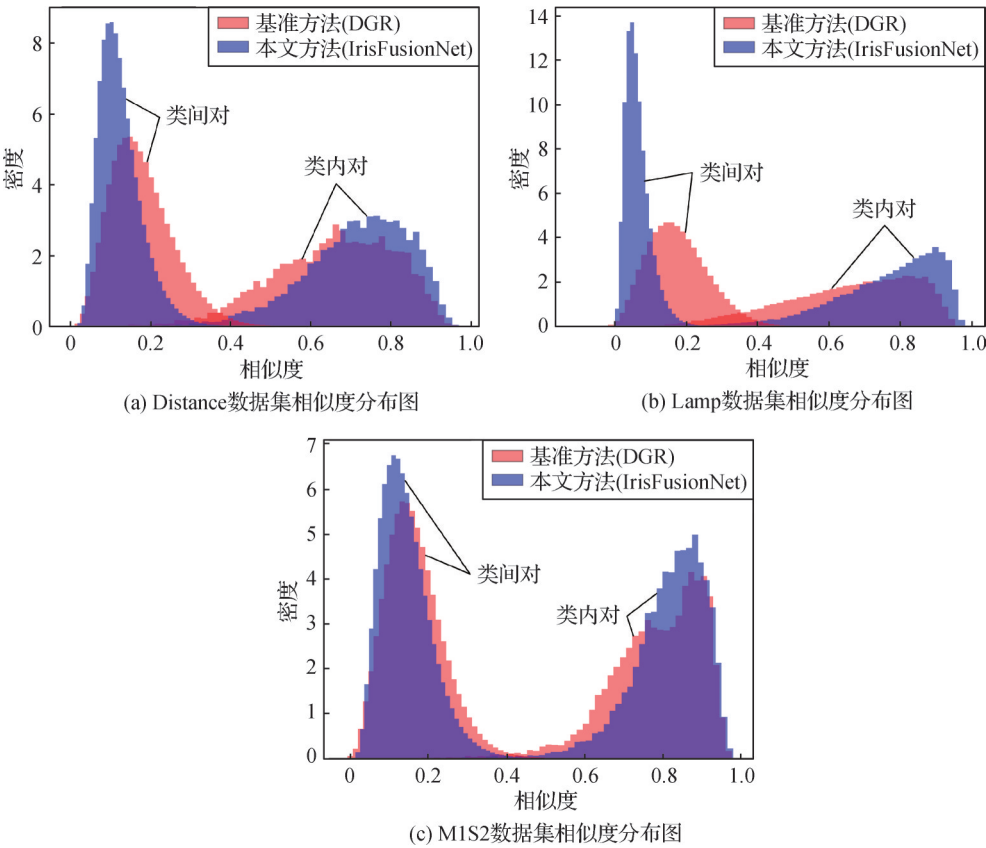


图9 虹膜识别实验相似度分布图

Fig. 9 Similarity distribution of iris recognition experiments ((a) similarity distribution on the Distance dataset; (b) similarity distribution on the Lamp dataset; (c) similarity distribution on the M1S2 dataset)

2. 2. 4 消融实验

为进一步验证本文提出算法的有效性和可行

性, 本节在 Distance 与 M1S2 数据集上进行消融实验测试。基准算法中设置相同的卷积模块, 主分支中

采用图注意力网络(GAT)作为GNN骨干,采用相同的图三元组损失;辅分支则采用softmax作为损失函数,在此基础上进行消融实验。表8给出了消融实验结果,表中EH表示ECA-HGAT模块,ROC(receiver operating characteristic curve)曲线如图10所示。

首先,组别1为基准方法,相比于组别1,组别2中主分支仍使用Graph triplet loss,将辅分支的softmax替换为AAM loss后,在Distance与MIS2数据集上EER和FRR@FAR = 0.01%,均获得小幅度提升。

组别3将GAT模块替换为ECA-HGAT模块,ECA层通过非降维局部跨通道注意力机制对输入特征进行通道特征加强,HGAT层则通过只选择最重

要的 k 个节点邻居节点选择进行图结构硬注意力操作,相比于GAT中的软注意力操作,重点关注包含重要信息的虹膜纹理特征区域,从而学习到更重要的节点特征。实验结果也证明,使用ECA-HGAT模块取得的识别效果更佳。

组别4在网络初始阶段加入PLE模块,在Distance与MIS2两个数据集上再次将EER指标分别提升0.11%和0.15%,同时FRR@FAR = 0.01%指标也取得0.75%和0.73%大幅度提升。这是因为PLE模块通过对局部特征和全局特征增强,极大消除了由于虹膜采集过程中无法预知因素而引入的噪声不确定性,从而达到提升识别精度和类内样本聚集的效果。

表 8 Distance 和 MIS2 数据集消融实验结果
Table 8 Ablation experimental results of Distance and MIS2 datasets

实验组别	网络结构	辅分支损失函数	/%			
			EER	FRR@FAR = 0.01%	EER	FRR@FAR = 0.01%
1	baseline	softmax	1.41	11.70	0.59	2.12
2	baseline	AAM loss	1.24	10.88	0.46	2.00
3	baseline (EH)	AAM loss	1.17	8.24	0.42	1.57
4	baseline (EH)+PLE	AAM loss	1.06	7.49	0.27	0.84

注:加粗字体表示各列最优结果。

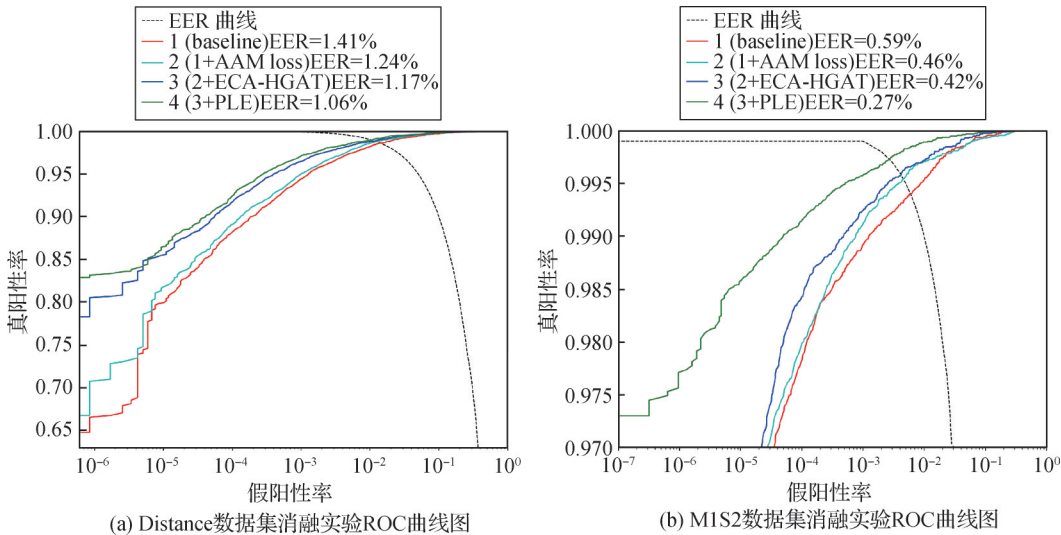


图 10 消融实验 ROC 曲线图
Fig. 10 ROC curves of ablation experiments

((a) ROC curves of ablation experiments for Distance dataset; (b) ROC curves of ablation experiments for MIS2 dataset)

3 结 论

本文针对低质量困难虹膜图像的识别任务,提

出了一种图神经网络与 CNN 融合的虹膜特征编码网络 IrisFusionNet: 首先对于输入的归一化虹膜图像,提出一个 PLE 模块以消除未知采集因素带来的不确定性影响;骨干网络分为两个分支,主分支通过

图转换器从图像中构建图结构,并使用ECA-HGAT选择最具关联的 k 个邻居节点进行节点特征聚合更新;辅分支则通过一个常规CNN网络提取全局特征。训练阶段利用双分支提取融合特征,并通过联合损失函数GTAU-Loss优化网络参数,提升网络进行虹膜识别的准确性与类聚性;推理阶段则使用融合特征匹配策略,其中基于门控的方法充分利用图非规则结构来进行相似节点匹配。

为了验证本文提出方法在虹膜识别任务中的有效性,在Distance、Lamp和MIS2公开的困难虹膜数据集上分别取得了1.06%、0.71%和0.27%的EER与7.49%、4.21%和0.84%的FRR@FAR=0.01%,均远优于目前已知的SOTA方法;同时,IrisFusionNet特征编码器在上述3个数据集测试集中分别取得了6.102、6.574和8.451的DI,相比于目前已知的SOTA方法提升50%以上,在虹膜识别任务中具有极高的准确性和类聚性。

尽管IrisFusionNet在低质量虹膜图像识别任务中取得优秀结果,但是虹膜识别任务中仍存在斜视样本识别、跨光谱识别等挑战。未来的研究中将对模型进行轻量化,并加入其他困难虹膜样本的识别研究。

参考文献(References)

- Chen Y, Wu W Q, Xu L and Guo S B. 2023. Iris-periocular features-fused non-collaborative authentication. *Journal of Image and Graphics*, 28(5): 1462-1476 (陈英, 吴文强, 徐亮, 郭书斌. 2023. 融合虹膜一眼周特征的非协作身份认证. *中国图象图形学报*, 28(5): 1462-1476) [DOI: 10.11834/jig.220649]
- Daugman J. 2009. How iris recognition works//Bovik A, ed. *The Essential Guide to Image Processing*. 2nd ed. Burlington, USA: Academic Press: 715-739 [DOI: 10.1016/B978-0-12-374457-9.00025-1]
- Deng J K, Guo J, Xue N N and Zafeiriou S. 2019. ArcFace: additive angular margin loss for deep face recognition//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA: IEEE: 4685-4694 [DOI: 10.1109/CVPR.2019.00482]
- Gangwar A and Joshi A. 2016. DeepIrisNet: deep iris representation with applications in iris recognition and cross-sensor iris recognition//*Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)*. Phoenix, USA: IEEE: 2301-2305 [DOI: 10.1109/ICIP.2016.7532769]
- Gangwar A, Joshi A, Joshi P and Ramachandra R. 2019. DeepirisNet2: learning deep-IrisCodes from scratch for segmentation-robust visible wavelength and near infrared iris recognition [EB/OL]. [2023-09-21]. <https://arxiv.org/ftp/arxiv/papers/1902/1902.05390.pdf>
- Gao H Y and Ji S W. 2019. Graph representation learning via hard and channel-wise attention networks//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Anchorage, USA: Association for Computing Machinery: 741-749 [DOI: 10.1145/3292500.3330897]
- Hamilton W L, Ying R and Leskovec J. 2018. Inductive representation learning on large graphs [EB/OL]. [2023-09-21]. <https://arxiv.org/pdf/1706.02216.pdf>
- Han K, Wang Y H, Guo J Y, Tang Y H and Wu E H. 2022. Vision GNN: an image is worth graph of nodes [EB/OL]. [2023-09-21]. <https://arxiv.org/pdf/2206.00272.pdf>
- Han Y, Wang P H, Kundu S, Ding Y and Wang Z Y. 2023. Vision HGNN: an image is more than a graph of nodes//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 19821-19831 [DOI: 10.1109/ICCV51070.2023.01820]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hou L, Liu J H, Yu X and Du J W. 2023. Review of graph neural networks [J/OL]. *Computer Science*, 1-25 [2023-09-21]. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJ&filename=JSJA20231008002> (侯磊, 刘金环, 于旭, 杜军威. 2023. 图神经网络研究综述[J/OL]. *计算机科学*, 1-25 [2023-11-02]. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJ&filename=JSJA20231008002>
- Howard A, Sandler M, Chen B, Wang W J, Chen L C, Tan M X, Chu G, Vasudevan V, Zhu Y K, Pang R M, Adam H and Le Q. 2019. Searching for MobileNetV3//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 1314-1324 [DOI: 10.1109/ICCV.2019.00140]
- Institute of Automation, Chinese Academy of Sciences. 2018. CASIA-Iris-Mobile-V1.0 database [DB/OL]. [2023-09-21]. <http://biometrics.idealtest.org> (中国科学院自动化研究所. 2018. CASIA-Iris-Mobile-V1.0 数据库 [DB/OL]. [2023-09-21]. <http://biometrics.idealtest.org>)
- Institute of Automation, Chinese Academy of Sciences. 2020. CASIA-IrisV4 database [DB/OL]. <http://biometrics.idealtest.org> (中国科学院自动化研究所. 2020. CASIA-IrisV4 数据库 [DB/OL]. <http://biometrics.idealtest.org>)
- Jia D D and Shen W Z. 2022. IrisCodeNet: iris feature coding network. *Computer Engineering and Applications*, 58(10): 185-192 (贾丁丁, 沈文忠. 2022. IrisCodeNet: 虹膜特征编码网络. *计算机工程与应用*, 58(10): 185-192) [DOI: 10.3778/j.issn.1002-8331.

- 2011-0466]
- Kipf T N and Welling M. 2016. Semi-supervised classification with graph convolutional networks [EB/OL]. [2023-09-21]. <https://arxiv.org/pdf/1609.02907.pdf>
- Li G H, Müller M, Qian G C, Delgadillo I C, Abualshour A, Thabet A and Ghanem B. 2023. DeepGCNs: making GCNs go as deep as CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(6): 6923-6939 [DOI: 10.1109/TPAMI.2021.3074057]
- Liu Y C, Shao Z R, Teng Y Y and Hoffmann N. 2021. NAM: normalization-based attention module [EB/OL]. [2023-09-21]. <https://arxiv.org/pdf/2111.12419.pdf>
- Liu Z, Mao H Z, Wu C Y, Feichtenhofer C, Darrell T and Xie S N. 2022. A ConvNet for the 2020s//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, USA: IEEE: 11966-11976 [DOI: 10.1109/CVPR52688.2022.01167]
- Miyazawa K, Ito K, Aoki T, Kobayashi K and Nakajima H. 2008. An effective approach for iris recognition using phase-based image matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10): 1741-1756 [DOI: 10.1109/TPAMI.2007.70833]
- Monro D M, Rakshit S and Zhang D X. 2007. DCT-based iris recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4): 586-595 [DOI: 10.1109/TPAMI.2007.1002]
- Ren M, Wang Y L, Sun Z N and Tan T N. 2020. Dynamic graph representation for occlusion handling in biometrics//*Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York, USA: AAAI: 11940-11947 [DOI: 10.1609/aaai.v34i07.6869]
- Schroff F, Kalenichenko D and Philbin J. 2015. FaceNet: a unified embedding for face recognition and clustering//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA: IEEE: 815-823 [DOI: 10.1109/CVPR.2015.7298682]
- Shankar S, Garg S and Sarawagi S. 2018. Surprisingly easy hard-attention for sequence to sequence learning//*Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing*. Brussels, Belgium: Association for Computational Linguistics: 640-645 [DOI: 10.18653/v1/D18-1065]
- Sun Z N and Tan T N. 2009. Ordinal measures for iris recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12): 2211-2226 [DOI: 10.1109/TPAMI.2008.240]
- Tan M X and Le Q V. 2019. EfficientNet: rethinking model scaling for convolutional neural networks//*Proceedings of the 36th International Conference on Machine Learning*. Long Beach, USA: PMLR: 6105-6114
- Teng T, Shen W Z and Mao Y F. 2020. Multi-task iris fast location method based on cascaded neural network. *Computer Engineering and Applications*, 56(12): 118-124 (滕童, 沈文忠, 毛云丰. 2020. 基于级联神经网络的多任务虹膜快速定位方法. *计算机工程与应用*, 56(12): 118-124) [DOI: 10.3778/j.issn.1002-8331.1903-0145]
- Veličković P, Cucurull G, Casanova A, Romero A, Liò P and Bengio Y. 2018. Graph attention networks [EB/OL]. [2023-09-21]. <https://arxiv.org/pdf/1710.10903.pdf>
- Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020. ECA-Net: efficient channel attention for deep convolutional neural networks//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 11531-11539 [DOI: 10.1109/CVPR42600.2020.01155]
- Wang Y H, Zhu Y and Tan T N. 2002. Biometrics personal identification based on iris pattern. *Acta Automatica Sinica*, 28(1): 1-10 (王蕴红, 朱勇, 谭铁牛. 2002. 基于虹膜识别的身份鉴别. *自动化学报*, 28(1): 1-10)
- Xiao L H, Sun Z N, He R and Tan T N. 2013. Coupled feature selection for cross-sensor iris recognition//*Proceedings of the 6th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*. Arlington, USA: IEEE: 1-6 [DOI: 10.1109/BTAS.2013.6712752]
- Yu B, Yin H T and Zhu Z X. 2021. ST-UNet: a spatio-temporal U-network for graph-structured time series modeling [EB/OL]. [2023-09-21]. <https://arxiv.org/pdf/1903.05631.pdf>
- Yue H L, Wang H Y and Peng R Y. 2022. The research progress of graph neural network//*Proceedings of the 17th Chinese Conference on Stereology and Image Analysis*. Chinese Society of stereology: 99-102 (岳含霖, 王浩宇, 彭瑞云. 2022. 图神经网络研究进展//*第十七届中国体视学与图像分析学术会议论文集*. 中国体视学学会: 99-102) [DOI: 10.26914/c.cnkihy.2022.051480]
- Zhao Z J and Kumar A. 2017. Towards more accurate iris recognition using deeply learned spatially corresponding features//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 3829-3838 [DOI: 10.1109/ICCV.2017.411]

作者简介

孙金通,男,硕士研究生,主要研究方向为图神经网络与虹膜识别. E-mail: sjt1998@mail.shiep.edu.cn

沈文忠,通信作者,男,副教授,主要研究方向为机器视觉与智能控制、生物特征识别(虹膜识别)和嵌入式系统开发。

E-mail: 7841426@qq.com