RTMPose-MCA:一种改进的瞳孔和眼角点定位模型

张丁玮[1],万亚平[1],邹　刚[1,2],罗　扬[1],张　璇[3]
(1. 南华大学 计算机学院,湖南 衡阳 421001;
2. 湖南中科助英智能科技研究院,湖南 长沙 410000;
3. 中南大学 湘雅医院,湖南 长沙 410000)

摘　要:精确定位瞳孔和眼角点有助于准确测量认知障碍评估中的视觉反应时间这一关键指标,由于形状、光照等图像噪声和干扰问题,实现高精度定位仍存在挑战。为此,提出了一种基于改进 RTMPose 的瞳孔和眼角点定位模型 RTMPose-MCA,以进一步提升定位的准确性。首先,设计了多尺度融合卷积模块 MSFCM 替换原模型的第一个卷积模块,增强了对区域细节信息的提取能力。其次,设计了通道方差融合注意力模块 CVFAM 替换 Backbone 部分模块,增强了对全局和局部特征的捕捉能力,减弱光照等噪声干扰。最后,设计了空洞融合卷积模块 AFCM 替代 Head 中的 7×7 卷积,减少了参数量。实验结果表明,在 BioID 和 GI4E 数据集上,RTMPose-MCA 模型在瞳孔和眼角点的定位精度方面优于其他对比算法,平均像素距离误差分别为 0.82 像素和 1.08 像素,且模型的参数量处于较低水平。这些结果验证了该模型在复杂环境下能够有效定位瞳孔和眼角点。

关键词:瞳孔定位;眼角定位;RTMPose;特征融合;注意力模块
中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1673-629X(2025)04-0164-08
doi:10.20165/j.cnki.ISSN1673-629X.2024.0379

RTMPose-MCA:An Improved Model for Pupil and Eye Corner Localization

ZHANG Ding-wei[1],WAN Ya-ping[1],ZOU Gang[1,2],LUO Yang[1],ZHANG Xuan[3]
(1. School of Computer,University of South China,Hengyang 421001,China;
2. Hunan Zhongke Help Innovation Research Institute,Changsha 410000,China;
3. Xiangya Hospital,Central South University,Changsha 410000,China)

Abstract:Accurate localization of the pupil and eye corners contributes to precisely measuring the visual reaction time,a key indicator in cognitive impairment assessments. However,achieving high-precision localization remains challenging due to shape variations,lighting conditions,and image noise interference. To address these issues,we propose an improved RTMPose-based model,named RTMPose-MCA,for pupil and eye corner localization to enhance accuracy. Firstly,a Multi-Scale Fusion Convolution Module (MSFCM) is designed to replace the first convolution module of the original model,which enhances the ability of extracting regional details. Secondly,the Channel Variance Fusion Attention Module (CVFAM) is designed to replace specific Backbone modules,which strengthens the capture of global and local features and weakens the noise interference such as light. Lastly,the Atrous Fusion Convolution Module (AFCM) substitutes the 7×7 convolution in the Head,which reduces the number of parameters and improves the scalability of the model. Experimental results on the BioID and GI4E datasets show that the RTMPose-MCA model outperforms other comparison algorithms in localization accuracy,achieving mean pixel distance error of 0.82 pixels and 1.08 pixels,respectively,while maintaining a relatively low parameter count. These findings demonstrate the model's effectiveness in accurately localizing the pupil and eye corners in complex environments.

Key words:pupil localization;eye corner localization;RTMPose;feature fusion;attention module

收稿日期:2024-09-29　　**修回日期**:2025-02-06
基金项目:湖南省自然科学基金(2024JJ7428)
作者简介:张丁玮(2000-),男,硕士研究生,CCF 会员(T1317G),研究方向为深度学习;通信作者:万亚平(1973-),男,教授,研究方向为分布式计算、网络存储、信息检索等;通信作者:邹　刚(1969-),男,教授,研究方向为智能信息处理、医工融合、深度数据挖掘等;罗　扬(1962-),男,教授,研究方向为软件工程、图像处理等。

0　引　言

在临床上,眼动参数可以为诊断认知障碍疾病提供重要依据[1]。眼动参数中的视觉反应时间[2],它是目标出现到定向扫视开始之间的反应时间[3]。视觉反应时间是认知评估中最常用的指标之一[4],在预测认知障碍方面具有重要的应用价值[5]。

在自由移动的测试环境中,头部移动会影响通过瞳孔追踪眼球运动的准确性[6],而眼角被视为测量相对瞳孔运动的稳定可靠的参考点[7]。因此,为提高数据可靠性,需在定位瞳孔的同时定位眼角。在连续图像中定位眼角和瞳孔位置,获得时序数据,用以计算视觉反应时间。

目前,定位算法主要分为两大类:传统的定位算法和基于深度学习的定位算法。

传统的瞳孔定位方法基于图像处理和特征提取,如 Cracan 等[8]采用圆形霍夫变换结合自适应二值化和形态学操作实现瞳孔检测。Fuhl 等[9]利用 Haar 特征和统计学习方法实现了瞳孔分割。

传统的眼角定位方法如孙磊等[10]通过 HAAR 结合区域精化和生长分割实现眼角定位。Laskar[11]采用显式形状回归 ESR 方法实现了眼角定位。传统定位方法对光照、噪声等环境变化的鲁棒性较差,阈值设定困难,算法复杂且效果不佳。

基于深度学习的定位方法在定位精度和鲁棒性上都优于传统的方法,如刘瑞[12]提出了一种基于多任务级联神经网络,先通过 R-Net 定位眼睛区域,再利用 O-Net 定位眼角。孙语等[13]采用基于注意力机制和空洞卷积的 U-Net 结构结合最小二乘法实现瞳孔定位。闵筱萌[14]采用 YOLOv8,GFNet 和 ShuffleNetv2 结合剪枝策略实现轻量化高精度瞳孔定位。王利等[15]采用混合结构的 Vision Transformer 结合 CNN 提取局部特征与全局依赖关系实现瞳孔定位。现有的深度学习定位方法仍存在局限性。如部分卷积神经网络在低分辨率图像上的精度受限,且对不同场景的泛化能力不足。部分网络的复杂结构导致检测速度较慢,或在小误差范围内的定位效果较差。

目前市面上的眼动仪主要依赖近红外光源成像,需要固定头部位置,而且会因为头部运动降低准确度[6]。

考虑到以上问题,该文基于 RTMPose[16]提出了一种改进的瞳孔和眼角点定位模型。首先,通过设计多尺度融合卷积模块(Multi-Scale Fusion Convolution Module,MSFCM),增强了模型对多尺度特征的捕捉能力。同时,设计通道方差融合注意力模块(Channel Variance Fusion Attention Module,CVFAM),增强了模型对局部和全局特征的提取。最后,设计空洞融合卷积模块(Atrous Fusion Convolution Module,AFCM),降低了模型的参数量,提升了模型的可扩展性。

基于上述改进,将该模型命名为 RTMPose-MCA,其中 M 代表 MSFCM,C 代表 CVFAM,而 A 则表示 AFCM。

1　RTMPose 模型

RTMPose 是 2023 年提出的一种高性能姿态估计模型,姿态估计是通过定位关键点的坐标来确定图像中的姿态。该文选择 RTMPose-s 用于瞳孔和眼角点定位,因其在计算资源有限的场景下具备高效性能,能够快速、准确地定位关键点。

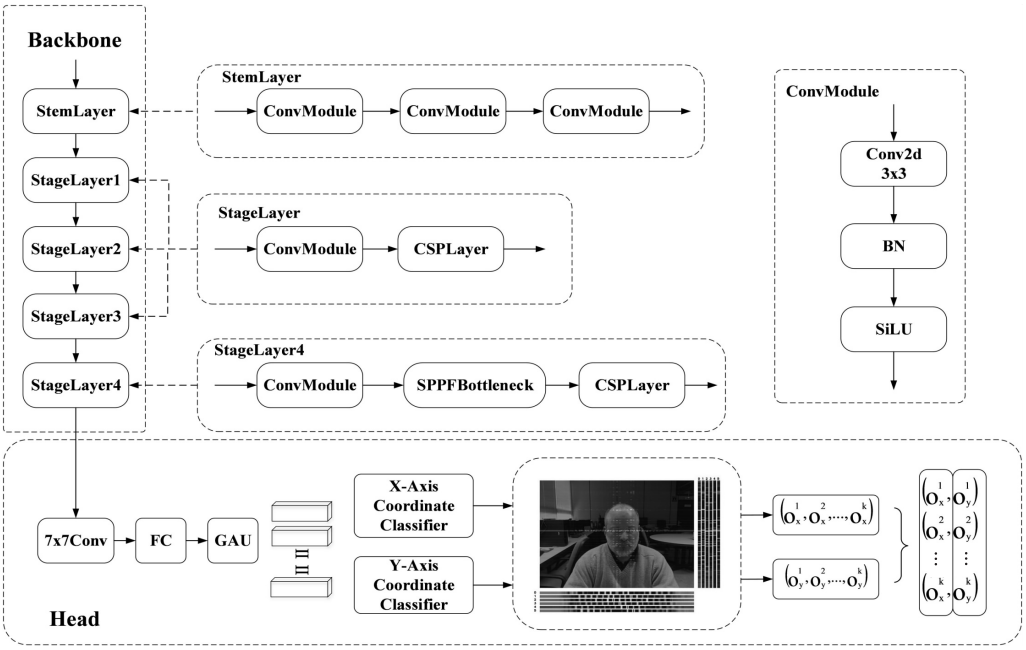

图 1　RTMPose-s 结构

RTMPose – s 的模型结构主要由主干网络(Backbone)和头部网络(Head)两个部分组成,如图 1 所示。主干网络由 StemLayer 和 StageLayer1 到 StageLayer4 组成,其中 StemLayer 用于初步特征提取和下采样,包含三个 ConvModule 模块,而各 StageLayer 进一步提取和强化特征,StageLayer4 包含 SPPFBottleneck 模块以增强多尺度特征表达。ConvModule 模块由 3×3 卷积,批量归一化(BN)和 SiLU 激活函数组成。头部网络部分包含 7×7 卷积,用于压缩通道,通过全连接层(FC)扩展特征维数,利用门控注意力单元(GAU)融合全局和局部信息,最终通过 X 轴和 Y 轴分类器实现关键点的坐标定位。

2 改进的 RTMPose – s 模型: RTMPose – MCA

RTMPose–MCA 模型在原模型 RTMPose–s 的基础上进行了多方面改进,以提高瞳孔和眼角点的定位精度,同时保持模型的轻量化。RTMPose–MCA 的模型结构如图 2 所示。首先,为解决原模型在多尺度特征捕捉上的不足,设计了多尺度融合卷积模块(MSFCM)并替换了 StemLayer 中的第一个卷积模块,增强了模型对复杂场景中多尺度特征的处理能力。其次,设计了通道方差融合注意力模块(CVFAM),用于替换 StemLayer 和 StageLayer1 中的部分模块,通过全局和方差池化加强了模型对局部和全局特征的捕捉。最后,针对原模型 Head 部分的计算开销问题,设计了空洞融合卷积模块(AFCM),替代 7×7 卷积,有效减少了参数量。这些改进模块有效提升了模型的性能和效率。

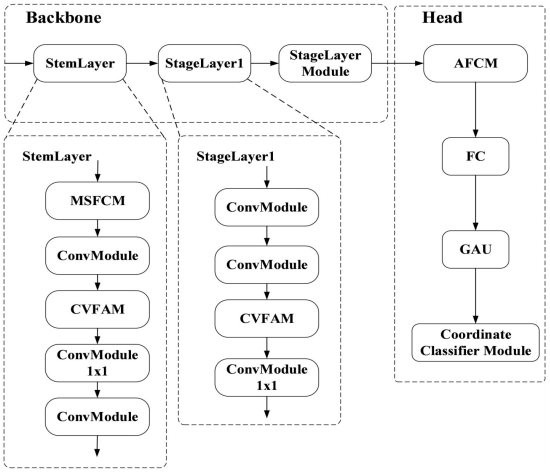

图 2 RTMPose–MCA 结构

2.1 MSFCM 模块

原模型 RTMPose–s 的第一层用于初步特征提取和下采样,这一阶段的卷积操作在处理输入图像的低级特征时起着关键作用,并影响后续层次的特征提取质量。然而,单一尺度的卷积核无法有效地提取复杂图像中这些特征点的多尺度信息,可能导致部分细节信息的丢失。在眼动追踪的认知评估中,瞳孔和眼角点的精确定位对于评估的准确性至关重要。

为了解决这一问题,该文设计了多尺度融合卷积模块 MSFCM,替换原模型 StemLayer 层的第一个卷积模块 ConvModule。MSFCM 模块结构如图 3 所示。该模块采用了并行的 3×3 和 1×1 卷积核结构,旨在提取图像中的多尺度特征信息。3×3 卷积主要用于提取局部空间特征,捕捉图像中的细节,并在卷积操作中保留空间关系,而 1×1 卷积则通过对每个像素点的所有通道值进行线性组合,实现了特征混合。这种设计增强了模型在处理复杂场景时捕捉多尺度信息的能力,将 3×3 和 1×1 卷积的输出逐元素相加进行特征融合,进一步融合了不同尺度的特征信息,提升了模型的表现能力。与原模块的单一尺度卷积模块相比,这种多尺度融合的方法可以提高模型在复杂场景中对关键点定位任务中的表现能力和鲁棒性。

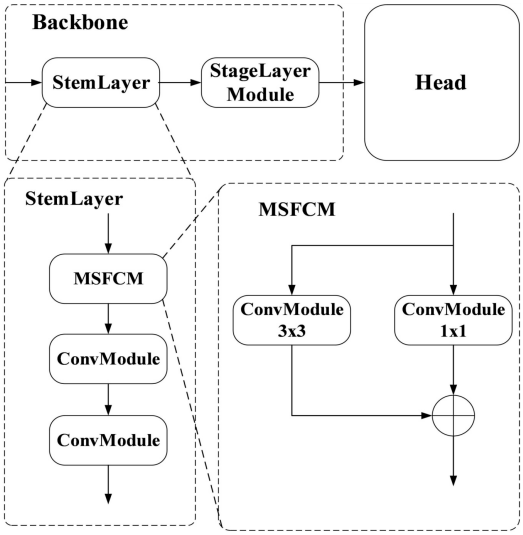

图 3 MSFCM 模块结构

2.2 CVFAM 模块

原模型中的卷积模块在处理复杂特征和噪声时,缺乏对全局信息和通道间差异的关注,难以全面捕捉特征图中的关键信息。为了进一步提升瞳孔和眼角点的定位精度,该文设计了一种融合通道注意力[17]和方差注意力[18]的通道方差融合注意力模块 CVFAM。CVFAM 模块结构如图 4 所示。

该模块首先通过全局平均池化提取输入特征图的全局信息,如公式 1 所示。

$$X_{\text{avg}} = \text{GlobalAvgPool}(X) \tag{1}$$

其中,X 表示输入特征图,X_{avg} 表示经过全局平均池化后的特征图。CVFAM 模块还引入了方差池化,用于计算输入特征图的通道方差信息,如公式 2 所示。

$$X_{\text{var}} = \text{VariancePool}(X) \tag{2}$$

其中,X_{var} 表示经过方差池化后的特征图。接下来,将

这两种特征分别通过 1×1 卷积,如公式 3 和公式 4 所示。

$$F_{avg} = \sigma(\text{Conv2d}(X_{avg})) \tag{3}$$

$$F_{var} = \sigma(\text{Conv2d}(X_{var})) \tag{4}$$

其中,σ 为激活函数 Hardsigmoid。接着,将得到的特征 F_{avg} 和 F_{var} 与原始输入特征进行加权相乘,如公式 5 和公式 6 所示。

$$Y_{avg} = X \cdot F_{avg} \tag{5}$$

$$Y_{var} = X \cdot F_{var} \tag{6}$$

最终,将两个加权后的特征 Y_{avg} 和 Y_{var} 相加,输出最终的特征结果 Y,如公式 7 所示。

$$Y = Y_{avg} + Y_{var} \tag{7}$$

全局平均池化有助于捕捉输入图像的整体信息,而方差池化则能够有效增强模型对局部特征变化的敏感性。通过这种设计,CVFAM 模块能够有效关注重要特征,同时抑制不必要的噪声和干扰。

原模型中的具体改进如图 4 所示。在原模型的第一层 StemLayer 和第二层 StageLayer1 中,将原结构的第二个 ConvModule 和 CSPLayer 模块替换为“ConvModule -> CVFAM -> ConvModule (1×1)”的结构。首先,ConvModule 用于提取局部空间特征,而后通过 CVFAM 模块增强特征图中的通道差异信息,进一步提高模型对局部细节和全局特征的理解能力。最后,ConvModule (1×1)起到了特征融合的作用,在保持计算效率的同时,进一步强化特征表达。这种结构能够提高模型特征提取的精度,并有效控制模型的参数量大小。

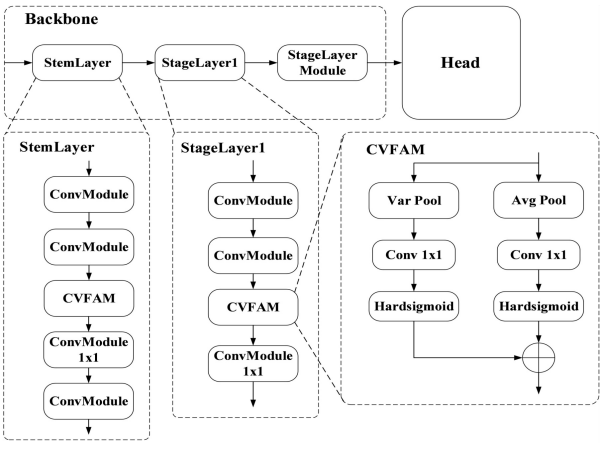

图 4　CVFAM 模块结构

2.3　AFCM 模块

原模型在 Head 部分使用了 7×7 卷积,这种大卷积核能够有效捕捉较大感受野的全局信息,但同时也增加了模型的计算开销和参数量。对于瞳孔和眼角点的细粒度定位任务,7×7 卷积可能带来过多的参数冗余,影响计算效率。为了解决这一问题,该文设计了一种基于空洞卷积[19]的空洞融合卷积模块 AFCM,替换原模型 Head 部分中的 7×7 卷积。AFCM 模块结构如图 5 所示。

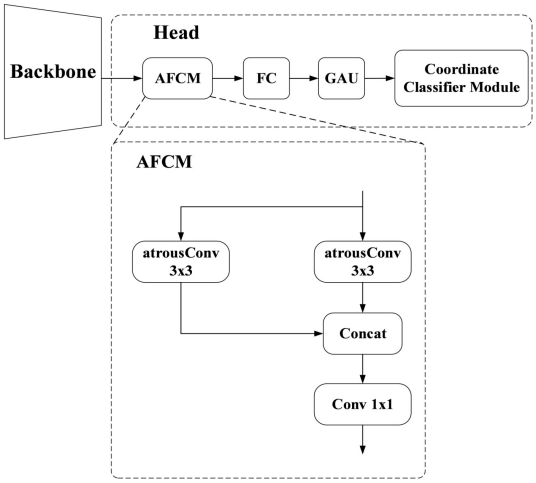

图 5　AFCM 模块结构

AFCM 模块包含两个 3×3 的空洞卷积,膨胀率分别为 1 和 2。接着将这两种不同膨胀率的卷积的输出进行通道维度的拼接,有效扩展了感受野。最后,使用 1×1 卷积将拼接后的特征进行融合,降低通道维度。相比原模型的 7×7 卷积,AFCM 的参数量显著减少。7×7卷积的参数量计算如公式 8 所示。

$$P_{7\times7} = 7 \times 7 \times C_{in} \times C_{out} \tag{8}$$

其中,C_{in} 表示输入通道数,C_{out} 表示输出通道数。AFCM 的参数量则由两个 3×3 空洞卷积和一个 1×1 卷积组成,参数量计算如公式 9 ~ 11 所示。

$$P_{3\times3} = 2 \times (3 \times 3 \times C_{in} \times C_{out}) \tag{9}$$

$$P_{1\times1} = 1 \times 1 \times 2C_{out} \times C_{out} \tag{10}$$

$$P_{AFCM} = P_{3\times3} + P_{1\times1} \tag{11}$$

在该模型中,C_{in} 取值为 512,C_{out} 的大小与模型需要定位的关键点个数有关。由于该文的目标是定位瞳孔和眼角点,模型需要定位六个关键点,设定 C_{out} 取值为 6,最后计算结果表明,AFCM 模块的参数量是原 7×7 卷积的三分之一,使用 AFCM 模块有效减少了模型的参数量。

3　实验结果与分析

3.1　公共数据集

该文选择 GI4E[20] 和 BioID[21] 数据集来测试和评估所提出的瞳孔和眼角点定位模型,所选数据集在多样性和标注精度方面为该文的目标提供了支持,帮助测试模型在不同条件下的稳定性和准确性。

GI4E 数据集包含了 1 236 张 800×600 分辨率的彩色图像,这些图像来自 103 名参与者。GI4E 数据集图像的彩色特性有助于在多种视觉环境下对模型性能进行评估。

BioID 数据集提供了 1 521 张 384×286 像素的灰

度图像,涵盖了 23 名参与者。BioID 数据集捕捉了多种光照,背景和面部变化,可以用于评估模型在复杂场景中的稳定性。

在 GI4E 和 BioID 数据集上均以 3∶1 的比例随机选取图像,分别用于训练集和测试集。

3.2 实验设置和评估指标

所有的实验基于深度学习框架 Pytorch 平台的 Nvidia GTX 3080 Ti GPU 进行,操作系统为 Windows11。模型训练采用 AdamW 优化器,并通过 LinearLR 和 CosineAnnealingLR 两阶段调度学习率,第一阶段线性衰减,第二阶段采用余弦退火。数据处理通过 Albumentation 库进行图像增强。通过固定随机种子确保了实验的可重复性。

瞳孔定位对比实验采用了 Gou 等人[22]使用的归一化误差评估指标 e。归一化误差评估指标如公式 12 所示。

$$e = \frac{\max(d_L, d_R)}{\mathrm{PCD}} \tag{12}$$

其中,d_L 和 d_R 分别是检测出的左瞳孔中心和右瞳孔中心坐标与真实坐标之间的欧几里得距离,而 PCD 是两瞳孔中心真实坐标之间的欧几里得距离。

这一指标通过双眼瞳孔定位中较大的误差与真实双眼瞳孔坐标欧几里得距离的比值,衡量瞳孔定位的精度。常用的归一化误差阈值包括:e <0.025(2.5%)通常对应瞳孔中心的精度,e <0.05(5%)通常对应瞳孔的收缩大小的精度,而 e <0.1(10%)则大致对应瞳孔区域。

瞳孔和眼角定位的对比实验采用了平均像素距离误差和参数量的评估指标。像素距离误差是预测点与真实点之间的欧几里得距离。在数据集中的每张图片上,计算所有预测点与真实点的距离误差,随后对所有图片中的距离误差进行平均,最终得出平均像素距离误差(Mean Pixel Distance Error,MPDE)。平均像素距离误差评估指标如公式 13 所示。

$$\mathrm{MPDE} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{PDE}_i \tag{13}$$

其中,PDE_i 是第 i 张图片的像素距离误差,N 是数据集中的图片总数。

3.3 瞳孔眼角定位对比实验

为了验证改进算法在瞳孔和眼角点定位任务中的有效性,在 BioID 数据集上进行了实验,并将其与现有主流关键点定位算法进行了对比。所有对比算法在相同的实验环境下进行。评价指标采用平均像素距离误差和模型参数量。实验结果如表 1 所示。

RTMPose-MCA 在平均像素距离误差和模型参数量上均表现出较为优异的性能。与经典的 ResNet 相比,RTMPose-MCA 的平均像素距离误差从 0.87 降低到 0.82,且参数量从 12.35 M 显著减少至 5.48 M,展现出更高的效率和精度。与轻量级网络 ShufflenetV1 和 MobilenetV2 相比,RTMPose-MCA 的精度进一步提升,分别比它们的 0.89 和 1.52 平均像素距离误差取得了明显的优势,同时保持了与这类轻量级模型相近的参数量。相比 PVT 和 SimCC 等模型,RTMPose-MCA 在保持较低参数量的同时,平均像素距离误差更低。

表 1 对比实验结果(BioID 数据集)

方法	MPDE	参数量/M
ResNet[23]	0.87	12.35
ShufflenetV1[24]	0.89	2.23
MobilenetV2[25]	1.52	2.10
PVT[26]	1.41	12.03
YOLO_Pose[27]	0.95	8.24
SimCC[28]	1.20	7.18
RTMPose-s	0.84	5.56
RTMPose-MCA	0.82	5.48

为了进一步验证改进算法,在 GI4E 数据集上进行了实验。实验结果如表 2 所示。RTMPose-MCA 在 GI4E 数据集上的表现优于其他对比算法,展示了良好的定位精度和参数效率。相比 ResNet,RTMPose-MCA 的平均像素距离误差从 1.27 降低至 1.08,且参数量对比有明显减少。与 ShufflenetV1 和 MobilenetV2 轻量级模型相比,RTMPose-MCA 的平均像素距离误差更低,分别比它们的 1.20 和 2.19 有明显改善,同时参数量保持在相似的水平。PVT 和 SimCC 等模型在 GI4E 数据集上的平均像素距离误差较高,而 RTMPose-MCA 显著降低了误差,进一步说明了其在精度上的优势。相比原模型 RTMPose-s,RTMPose-MCA 将平均像素距离误差从 1.28 降低至 1.08,且参数量略有减少,这表明改进后的模型在 GI4E 数据集上具有更高的定位精度和较好的模型复杂度。

表 2 对比实验结果(GI4E 数据集)

方法	MPDE	参数量/M
ResNet[23]	1.27	12.35
ShufflenetV1[24]	1.20	2.23
MobilenetV2[25]	2.19	2.10
PVT[26]	1.98	12.03
YOLO_Pose[27]	1.66	8.24
SimCC[28]	2.05	7.18
RTMPose-s	1.28	5.56
RTMPose-MCA	1.08	5.48

3.4　消融实验

为了验证提出的改进策略对模型性能的有效性,设计并开展了一系列消融实验。实验以 RTMPose-s 为基础模型,选择 GI4E 数据集进行测试,并保持其他参数一致。通过逐步移除或添加各个改进模块,对比各模块对模型性能的影响。实验以平均像素距离误差和模型参数量为评估指标,结果如表 3 所示。其中,“√”表示使用了相应的模块。

表 3　RTMPose-MCA 消融实验结果(GI4E 数据集)

MSFCM	CVFAM	AFCM	MPDE	参数量/M
			1.28	5.56
√			1.31	5.56
	√		1.38	5.58
		√	1.37	5.46
√	√		1.11	5.58
	√	√	1.17	5.48
√		√	1.33	5.46
√	√	√	1.08	5.48

通过消融实验的结果可以看出,各个模块对模型的精度和参数量有着不同的影响。首先,在没有使用任何模块的基础模型中,平均像素距离误差为 1.28,参数量为 5.56 M。当仅使用 MSFCM 模块时,模型的平均像素距离误差略微上升到 1.31,表明 MSFCM 在单独使用时对瞳孔和眼角点的定位精度贡献有限。其参数量保持不变,说明该模块的引入并未增加模型的复杂度。只使用 CVFAM 模块时,平均像素距离误差略微增加到 1.38,参数量略有上升至 5.58 M。但在结合 AFCM 模块时,平均像素距离误差降低到 1.17,表明这两个模块的协同作用对模型精度的提升起到了积极作用,同时参数量降低到 5.48 M,显示出两者结合具有较好的效果。

单独使用 AFCM 模块时,平均像素距离误差为 1.37,参数量减少到 5.46 M,这表明该模块起到了优化参数量的作用。相比之下,当 MSFCM 模块与 AFCM 模块组合时,模型的平均像素距离误差为 1.33,尽管误差没有大幅降低,但参数量维持在较低水平,说明这种组合能够在保持模型轻量化的同时提供一定的定位性能提升。

进一步分析发现,当 MSFCM 模块与 CVFAM 模块结合使用时,模型的平均像素距离误差下降到 1.11,显示出这两个模块的结合对模型精度有较强的提升作用,虽然参数量略有增加。这一组合证明了多尺度特征融合和通道方差融合注意力模块在联合应用时能有效提升瞳孔和眼角点的定位精度。

在所有模块都使用的情况下,模型达到了最佳表现,平均像素距离误差为 1.08,参数量为 5.48 M。相比基础模型,同时降低了误差和模型参数量。这说明,多尺度特征融合,通道方差融合注意力模块与空洞融合卷积模块的结合有效增强了模型对细节的捕捉能力,进一步验证了提出的改进策略在提高定位瞳孔和眼角点精度方面的有效性。

3.5　瞳孔眼角定位效果比较

为了更直观地对比改进后的模型 RTMPose-MCA 在瞳孔和眼角点定位任务中的效果,在 BioID 和 GI4E 数据集中通过随机选取样本图像并采用多种关键点定位模型进行定位。定位效果对比如图 6 和图 7 所示。

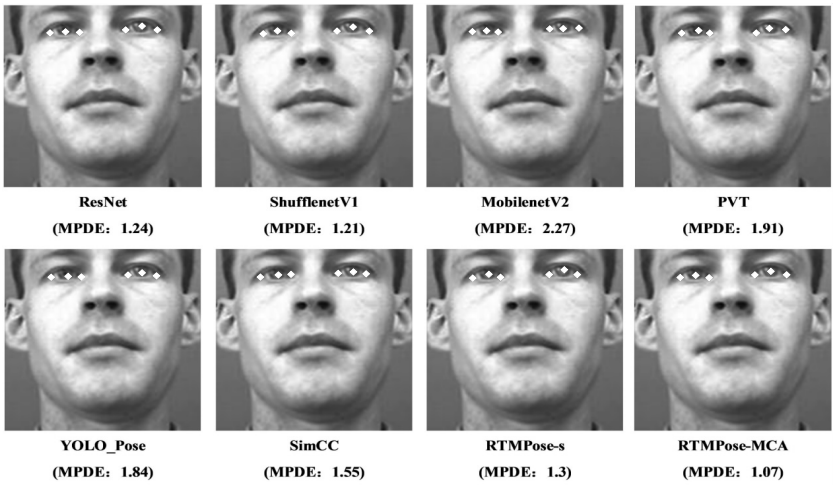

图 6　BioID 数据集上各模型定位效果对比

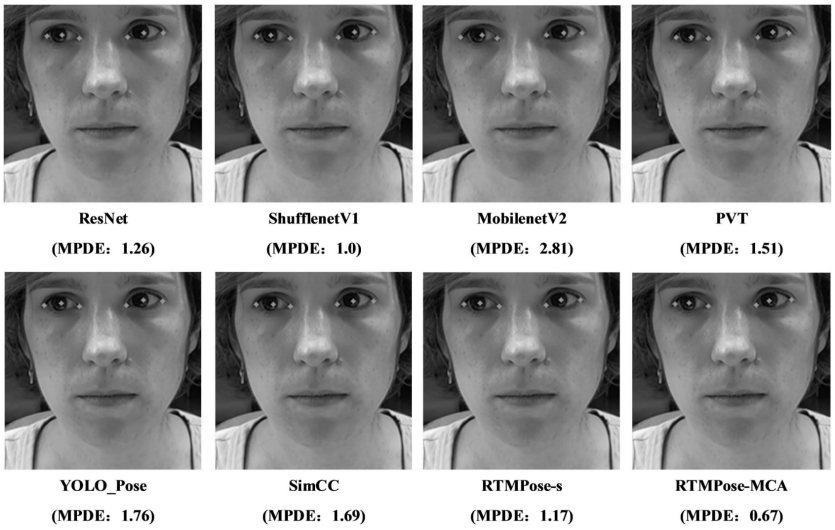

图 7 GI4E 数据集上各模型定位效果对比

BioID 数据集上各模型的瞳孔和眼角点定位效果如图 6 所示,图片中标注了每个模型的平均像素距离误差值。从图中可以看出,原模型 RTMPose-s 的平均像素距离误差为 1.3,而 RTMPose-MCA 的平均像素距离误差降低到 1.07,表现出更高的定位精度。其他模型中,ResNet 和 ShufflenetV1 的误差较小,分别为 1.24 和 1.21,但相较 RTMPose-MCA 仍有一定差距。而 MobileNetV2 和 PVT 的误差较大,分别为 2.27 和1.91。YOLO_Pose 和 SimCC 的表现居中,但与 RTMPose-MCA 相比,误差较为偏大。

在 GI4E 数据集上各模型的瞳孔和眼角点定位效果如图 7 所示。从图中可以看出,RTMPose-MCA 表现出最低的平均像素距离误差,仅为 0.67,显著优于其他模型。相比之下,原模型 RTMPose-s 的误差为 1.17,与 RTMPose-MCA 相比仍有提升空间。ShufflenetV1 在该图像上的误差为 1.0,表现相对较好,但相比 RTMPose-MCA,效果仍有不足。MobileNetV2 的误差达到 2.81,表现较差。其他模型如 ResNet 和 PVT 的误差分别为 1.26 和 1.51,精度表现为中等。

通过对比在两个数据集下的定位效果,改进后的模型 RTMPose-MCA 能够更精确地定位瞳孔和眼角点,还在不同场景下保持了较高的稳定性,特别是在 GI4E 数据集中的表现尤为突出,验证了该模型在瞳孔和眼角点定位任务中的优越性。

4 结束语

针对瞳孔和眼角点的精确定位问题,提出了一种改进 RTMPose-s 的模型 RTMPose-MCA,以提升定位的精度。相较于原模型 RTMPose-s,RTMPose-MCA 有以下改进:设计了多尺度融合卷积模块 MSFCM,用于提升模型特征提取能力。设计了通道方差融合注意力模块 CVFAM,用于关注重要特征,抑制噪声干扰,提升瞳孔和眼角点的定位性能。设计了空洞融合卷积模块 AFCM,用于减少模型的参数量。通过在 BioID 和 GI4E 数据集上的实验验证,RTMPose-MCA 在对比其他模型中表现出了更高的定位精度,平均像素距离误差分别为 0.82 像素和 1.08 像素,同时相比大部分模型具有较小的参数量,其参数量大小为 5.48 M。未来的研究将聚焦于优化模型结构和参数配置,在降低计算开销的同时,进一步提升模型对瞳孔和眼角点的定位精度。瞳孔和眼角点的准确定位为眼动参数分析提供了可靠的基础数据,从而为诊断和评估认知障碍疾病提供支持。

参考文献:

[1] OPWONYA J, DOAN D N T, KIM S G, et al. Saccadic eye movement in mild cognitive impairment and Alzheimer's disease: a systematic review and meta-analysis[J]. Neuropsychology Review, 2022, 32(2): 193-227.

[2] LENG Q, DENG B, JU Y. Application and progress of advanced eye movement examinations in cognitive impairment[J]. Frontiers in Aging Neuroscience, 2024, 16: 1377406.

[3] WOLF A, TRIPANPITAK K, UMEDA S, et al. Eye-tracking paradigms for the assessment of mild cognitive impairment: a systematic review[J]. Frontiers in Psychology, 2023, 14: 1197567.

[4] KIM J, FRANCISCO E, HOLDEN J, et al. Visual vs. tactile reaction testing demonstrates problems with online cognitive testing[J]. The Journal of Science and Medicine, 2020, 2(2): 1-10.

[5] TENG J, MCKENNA M R, GBADEYAN O, et al. Linking the neural signature of response time variability to Alzheimer's disease pathology and cognitive functioning[J]. Network Neuroscience, 2024, 8(3): 697-713.

[6] EHINGER B V, GROß K, IBS I, et al. A new comprehensive eye-tracking test battery concurrently evaluating the pupil

labs glasses and the EyeLink 1000 [J]. PeerJ, 2019, 7: e7086.

[7] LAZARUS M Z, GUPTA S, PANDA N. A literature survey on eye corner detection techniques in real-life scenarios[C]// International conference on advances in computing and data sciences. Singapore: Springer, 2019: 597-608.

[8] BONTEANU P, BOZOMITU R G, CRACAN A, et al. A new pupil detection algorithm based on circular Hough transform approaches[C]//2019 IEEE 25th international symposium for design and technology in electronic packaging (SIITME). Cluj-Napoca: IEEE, 2019: 260-263.

[9] FUHL W, SCHNEIDER J, KASNECI E. 1000 pupil segmentations in a second using haar like features and statistical learning[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal: IEEE, 2021: 3466-3476.

[10] 孙　磊,陈树越,戚亚明. 室外环境下红外热图像内眼角定位[J]. 红外技术,2022,44(10):1103-1111.

[11] AHMED M, LASKAR R H. Evaluation of accurate iris center and eye corner localization method in a facial image for gaze estimation[J]. Multimedia Systems, 2021, 27(3): 429-448.

[12] 刘　瑞. 基于神经网络的学习者在线学习状态研究[D]. 长春:长春理工大学,2022.

[13] 孙　语,刘文龙,蒋茂松. 基于注意力机制和空洞卷积的瞳孔定位算法[J]. 电子测量技术,2024,46(15):126-132.

[14] 闵筱萌. 基于深度学习的轻量化瞳孔定位与跟踪方法研究[D]. 太原:中北大学,2024.

[15] 王　利,王长元. Vision Transformer 的瞳孔定位方法[J]. 西安工业大学学报,2023,43(6):561-567.

[16] JIANG T, LU P, ZHANG L, et al. Rtmpose: real-time multi-person pose estimation based on mmpose[J]. arXiv: 2303.07399, 2023.

[17] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 7132-7141.

[18] BEHJATI P, RODRIGUEZ P, FERNÁNDEZ C, et al. Single image super-resolution based on directional variance attention network[J]. Pattern Recognition, 2023, 133: 108997.

[19] YU F. Multi-scale context aggregation by dilated convolutions[J]. arXiv: 1511.07122, 2015.

[20] VILLANUEVA A, PONZ V, SESMA-SANCHEZ L, et al. Hybrid method based on topography for robust detection of iris center and eye corners[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2013, 9(4): 1-20.

[21] JESORSKY O, KIRCHBERG K J, FRISCHHOLZ R W, et al. The BioID face database[EB/OL]. http://www.bioid.com/downloads/facedb/index.php.

[22] GOU C, ZHOU Y, XIAO Y, et al. Cascade learning for driver facial monitoring[J]. IEEE Transactions on Intelligent Vehicles, 2022, 8(1): 404-412.

[23] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 770-778.

[24] ZHANG X, ZHOU X, LIN M, et al. Shufflenet: an extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 6848-6856.

[25] SANDLER M, HOWARD A, ZHU M, et al. Mobilenetv2: inverted residuals and linear bottlenecks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 4510-4520.

[26] WANG W, XIE E, LI X, et al. Pyramid vision transformer: a versatile backbone for dense prediction without convolutions[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal: IEEE, 2021: 568-578.

[27] MAJI D, NAGORI S, MATHEW M, et al. Yolo-pose: enhancing yolo for multi person pose estimation using object keypoint similarity loss[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans: IEEE, 2022: 2637-2646.

[28] LI Y, YANG S, LIU P, et al. Simcc: a simple coordinate classification perspective for human pose estimation[C]//European conference on computer vision. Cham: Springer, 2022: 89-106.