

智能看护场景下的婴幼儿表情识别

王一评, 王改华

(天津科技大学人工智能学院, 天津 300457)

摘要:基于深度学习的表情识别方法逐渐应用于婴幼儿智能看护场景下,但大多数方法都是基于实验环境下的单人、正面图像的识别,对真实场景下多人表情识别效果欠佳;且深度网络模型存在训练耗时过长、数据处理速度慢的问题,导致表情识别的实时性难以得到保障。因此,该文提出一种改进的轻量型婴幼儿表情识别方法。首先,它在YOLOv8的基础上,结合深度可分离卷积和注意力机制理论,设计了轻量化注意力机制网络结构,减少了网络的参数并有效提取出显著特征;其次,搭建知识蒸馏框架,以YOLOv8L为教师网络模型,利用Channel-wise Knowledge Distillation(CWD)知识蒸馏的方式,采用改进的损失函数对学生模型进行蒸馏,提高婴幼儿的表情识别准确率及算法的实时性;最后,通过实验验证,该网络在智能看护场景下的婴幼儿(0-3岁)表情数据集上表现出色,mAP达到了73.4%。参数量仅有2.6 MB,便于部署在边缘设备上,对比其他表情识别算法,在CK+数据集上准确率达到99.1%,参数量也较小。

关键词:婴幼儿;YOLOv8;表情识别;轻量化;注意力机制;知识蒸馏

中图分类号:TP391.41;TP183

文献标识码:A

文章编号:1673-629X(2025)02-0016-08

doi:10.20165/j.cnki.ISSN1673-629X.2024.0307

Infants Expression Recognition in Intelligent Nursing Scene

WANG Yi-ping, WANG Gai-hua

(Institute of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin 300457, China)

Abstract:The expression recognition method based on deep learning is gradually applied to the intelligent nursing scene of infants. However, most methods are based on the recognition of single and front images in the experimental environment, and the recognition of multiple expressions in real scenes is poor. Moreover, the deep network model has long training time and slow processing speed, resulting in real-time guarantee of expression recognition. Therefore, we propose an improved method for lightweight expression recognition in infants. Firstly, based on YOLOv8, we introduce a lightweight attention mechanism network structure by integrating deep separable convolution and attention mechanism. This design effectively reduces network parameters and enhances the extraction of significant features. Then, we establish a knowledge distillation framework utilizing YOLOv8 as the teacher network model. It employs the Channel-wise Knowledge Distillation (CWD) method and an improved loss function to elevate expression recognition accuracy while optimizing real-time performance. The experimental results show that the proposed method performs well on the expression recognition of infants (0-3 years old) in the intelligent care scenario, with the recognition accuracy of 73.4%. The number of parameters is only 2.6 MB, making it easily deployable on edge devices. Besides, accuracy reaches 99.1% on CK+ data set, and the number of parameters is also small.

Key words: infants; YOLOv8; facial expression recognition; light weight; attention mechanism; knowledge distillation

0 引言

随着人工智能逐渐应用到各种领域,以婴幼儿为受益主体的相关产业链逐渐变得智能化^[1]。在婴幼儿智能看护系统中增加自动表情识别这一功能,不但可以协助看护者关注婴幼儿的状态,而且可以辅助看护人员对婴幼儿的突发情况进行及时有效地处理。

传统的人脸表情识别方法通常依赖于手工设计的特征提取器,其性能受到特征表达能力的限制,而深度

学习技术的兴起为解决这一问题提供了有效的途径。

Tang等为研究婴儿微笑收集了婴儿微笑数据集,提出基于上下半脸的双路卷积网络,取得87.16%的分类准确度^[2]。Sun等为研究婴儿舒适相关表情,收集13个月以下婴儿面部表情图像,用DenseNet模型,在自建的婴儿表情库上检测,达到91%的准确率^[3]。Zhang等使用改进的LBP和Sobel边缘检测算子,提出了一种婴儿面部表情识别的多分支融合网络,分别

收稿日期:2024-06-24

修回日期:2024-10-25

基金项目:国家自然科学基金项目(61601176);国家级大学生创新创业训练计划资助项目(202310057233)

作者简介:王一评(2002-),男,研究方向为计算机视觉;通讯作者:王改华(1979-),女,博士,副教授,研究方向为计算机视觉、模式识别、图像处理。

在图像的水平 and 垂直方向提取 3 幅特征图像,输入到多流 CNN 模型中识别婴儿表情^[4]。邓源等提出一种基于 MobileNetV2 与 LBP 双通道特征融合的婴幼儿表情识别算法,在自建的婴幼儿表情数据集中的表情识别准确率可达到 85.71%^[5]。文献[6]提出无参数注意力模块(Parameter-free Attention Module, PFAM)以及 Res Net 改进模型来提升婴幼儿表情识别率。耿等人提出了多尺度信息融合网络,在婴儿面部表情数据集中识别准确率达到 95.46%,在 AUC、召回率和 F1 得分 3 个评价指标上分别达到 99.07%、95.88% 和 95.89%^[7]。

这些相关研究和应用取得了较好的效果,但已有模型大多都是体积大、计算复杂,这些阻碍了网络模型的落地。一些研究者对表情识别设计了轻量化的改进方法。如文献[8]设计了一种轻量级的稠密块,通过将其嵌入 MobileNetV1 并修改网络的宽度,在降低参数数量的同时,识别准确率可达 77.9%,平均帧率为 6 FPS。Zhao 等提出一种轻量级的情感识别(LER)模型,将密集连接的卷积层和模型压缩整合到一个框架,以处理自然条件下的延迟问题^[9]。Anggit Wikanningrum 等采用迁移学习方式,从中大型数据集提高轻量级 CNN 面部表情识别的性能^[10]。

以上的表情识别方法主要针对单个采集的人脸进行分析识别,当出现多人脸场景(如托管中心)时,由于监测设备中多人脸占据较小的区域,使得目标细节不够明显,面部可能会被遮挡;或者当姿势发生较大变化时,网络模型难以获取足够的面部信息,使表情识别变得困难,难以准确地检测和定位。此外,由于婴幼儿尚未具备应对突发情况的能力,在面对危险和身体不适时无法进行相应的反应,识别婴幼儿的面部表情变化对实时性要求更高。Bie 等人提出一种基于特征增强的多人脸表情识别方法,以 YOLOv5 模型进行人脸检测,再使用 Resnet-34_Focal 模型进行表情分类^[11]。但这些表情识别方法对多人脸表情进行识别时,实时性差,不能进行有效的端到端网络训练,导致模型复杂,难以实现高效的模型优化和训练。

因此,该文将以日常看护设备和看护机构的监控系统为出发点进行研究,提出一种基于改进的 YOLOv8 的轻量级、多人脸表情识别网络,在单一框架内整合人脸检测和表情识别,通过结合改进的 YOLOv8 网络结构,能够在实时性和精度之间取得平衡,尤其是在多人脸场景下的应用。

此外,引入知识蒸馏理论,将教师网络大模型的知识迁移到学生模型中,进一步提升表情识别的性能和效率。能够满足实时的多人脸婴幼儿看护场景下表情的有效识别。

1 相关理论

1.1 深度可分离卷积

深度卷积神经网络中的标准卷积操作在提取特征时,同时涉及了通道间的信息和空间信息,导致计算复杂度较高。因此,Xception 和 MobileNet 提出了深度可分离卷积(Depthwise Separable Convolution)方法^[12-14],其背后的灵感源自对标准卷积操作中通道和空间信息耦合的认识。深度可分离卷积将卷积操作分为深度卷积(Depthwise Convolution)和逐点卷积(Pointwise Convolution)两个部分,降低了网络参数。Guo Yunhui 等人提出了一种基于深度可分离卷积的多域学习体系结构,在 ImageNet 等评估,实现只需要 50% 的参数就可达到最高的分数^[15]。Kamal K 等人将深度可分离卷积结构应用于植物病害监测,实现了高精度的精简模型^[16]。文献[17]在三维卷积神经网络(3DCNN)中引入一种注意力机制和深度可分离卷积,构建了 3DCNN-AM-DSC 模型,减少了 91.77% 的训练时间。

1.2 知识蒸馏

知识蒸馏(Knowledge Distillation)是一种深度学习中的模型优化技术,旨在将一个复杂模型的知识(模型的软标签)传递给一个简化的模型,以便提高简化模型的性能。Chen G 等提出用知识蒸馏做目标检测。核心思想是构建三个 loss 函数,分别对主干网络、分类 head 和回归 head 进行蒸馏^[18]。Shu C 等提出通道式蒸馏(Channel-wise Knowledge Distillation, CWD),将每个通道特征图归一化得到软概率图^[19]。通过最小化两个网络的通道概率图之间的 Kullback-Leibler (KL),使蒸馏过程更关注通道显著区域。文献[20]采用通道知识蒸馏方法对 YOLOv5 进行蒸馏,在玉米叶片病害检测数据集中测试,参数比 YOLOv5s 少 15.5%,而 mAP(0.5)和 mAP(0.5:0.95)分别提高 3.8% 和 1.5%。文献[21]提出将全局注意力网络(GIoAN)与 CWD 知识蒸馏相结合,使 GIoAN 的性能得到了进一步提升,MIoU 提高了 4.43%。文献[22]使用通道蒸馏的量化感知训练,以实现在不引入太多额外计算成本的情况下进一步提高模型准确性。文献[23]用 CWD 方法对改进的 Mask RCNN 模型进行知识蒸馏,在测试集上的识别准确率达到 74.13%,每张图像的平均检测时间仅为 62.1 ms。该文利用 CWD 知识蒸馏的方式,以 YOLOv8L 为教师网络模型,对改进的 YOLOv8n 学生模型进行蒸馏。

1.3 YOLOv8 概述

YOLOv8 在保持实时性和高效性的基础上,引入了更为先进的架构设计和优化策略,使其在目标检测任务中表现更加出色。其基本架构由以下关键部分

组成:

(1)骨干网络(Backbone):作为整个模型的基础部分,骨干网络包含 Conv、C2f 和 SPPF 等模块。Conv 模块执行卷积、批量归一化和 SiLU 激活等操作;C2f 模块在轻量化的同时丰富了梯度流信息;SPPF 模块则通过多尺度特征提取和编码,提升了对图像细节的捕捉能力。

(2)颈部网络(Neck):包括特征金字塔网络(Feature Pyramid Network,FPN)和路径聚合网络(Path Aggregation Network,PAN)。FPN 自顶向下传递语义特征,PAN 自底向上强化位置特征,通过 FPN 与 PAN 的结合,实现了不同层次特征图的有效融合。

(3)头部网络(Head):负责最终的目标检测,即预测边界框、目标类别以及置信度。头部网络通常由多层卷积层组成,并在特征图上生成检测结果。

2 文中方法

该文采用的整体网络模型如图 1 所示。以改进后的模型为学生模型,选择 YOLOv8L 作为教师网络模型,使用通道式蒸馏(Channel-wise Knowledge Distillation),在学生模型中集成教师模型的通道信息,确保了模型在学习过程中更好地捕捉特定特征,从而提高了检测的精度。

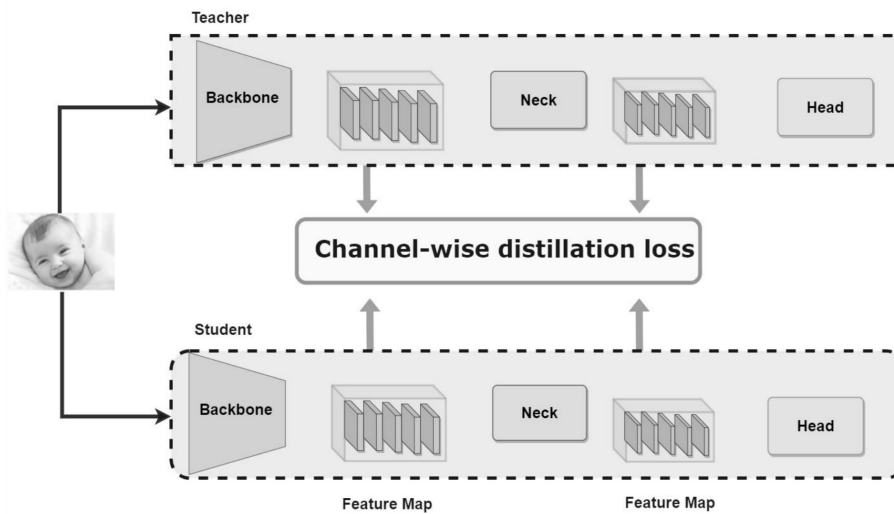


图 1 改进的 YOLOv8 蒸馏模型

2.1 YOLOv8 模型改进

2.1.1 CDSC 模块

在 YOLOv8 的 C2f 模块中堆叠了更多的 Bottleneck 结构,使得不同尺度的特征得到更好地利用,但也不可避免地导致了通道信息的过度冗余以及更大的模型计算量。为此,该文设计了更轻量化的 CDSC 结构,如图 2 所示。输入经过 CBS 模块再经过 Split 层会将其划分成两组通道特征图,第一组使用 n 个 DSC(Depthwise Separable Convolution)Block 模块,之后与第二组按通道拼接。再通过 CBS 块压缩输出。

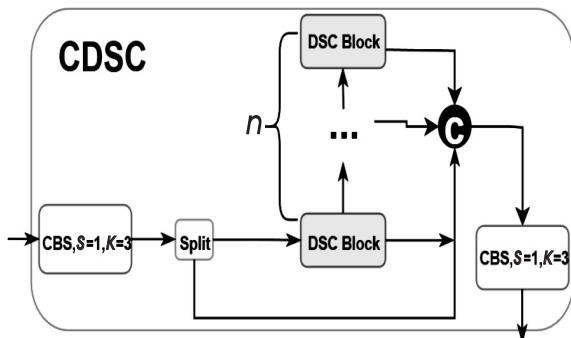


图 2 CDSC 结构

2.1.2 DSC Block

参考 Xception 等模型中的深度可分离卷积模块,该文构建了轻量化的 DSC Block,如图 3(b)所示。其中 CBS 是一个由卷积、批归一化和激活函数组成的基本卷积块。 X 和 X' 分别表示该结构的输入和输出 feature Map。经过 DSConv 模块,首先由 Depthwise Convolution 对输入 X 的每个通道分别使用一个卷积核,然后将所有卷积核的输出再进行拼接得到它的最终输出。然后经过 Pointwise Convolution,让 CDSC 能够自由改变输出通道的数量并对 Depthwise Convolution 输出的 feature map 进行通道融合。

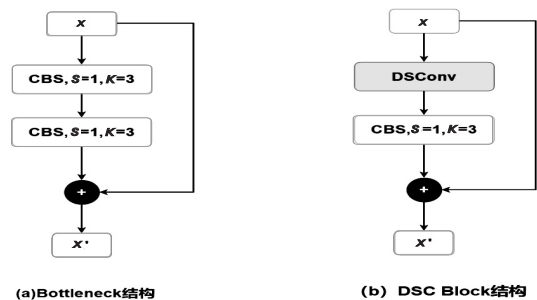


图 3 Bottleneck 与 DSC Block 结构对比

2.2 学生模型

学生网络模型的 Backbone、Neck 和 Head 部分如图 4 所示。用构建的轻量化的 CDSC(CBS-Depthwise Separable Convolution)结构替换了 YOLOV8 的 C2f 模块。其次,在 SPPF 后加入 CA(Coordinate Attention)注意力机制,以捕捉特定位置的空间关系,最后将 Detect 中的 CIoU 损失函数替换为 MPDIoU。网络中的 Conv 为由卷积、批归一化和激活函数组成的 CBS 卷积块。

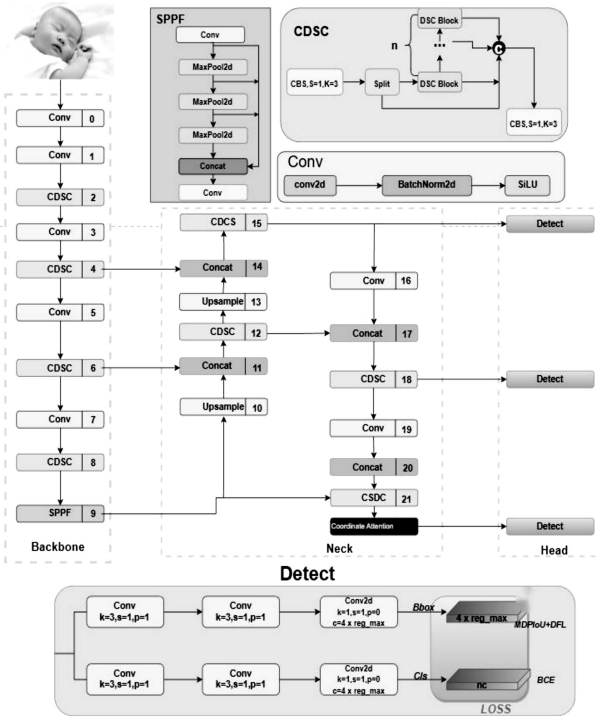


图 4 学生网络模型

2.3 CA 注意力机制模块

CA(Coordinate Attention)注意力机制的核心思想是引入坐标信息,以便模型可以更好地理解不同位置之间的关系,其过程如图 5 所示。

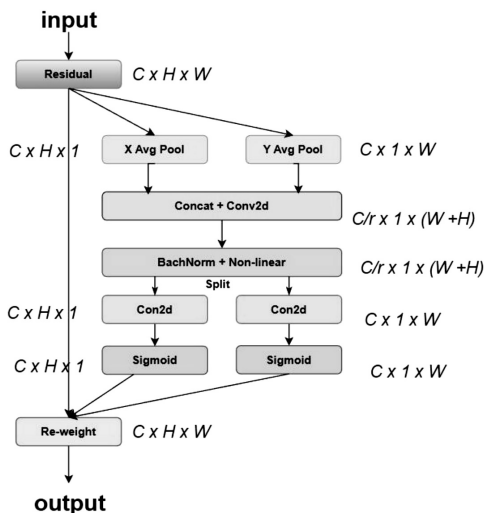


图 5 CA 注意力机制模块

先对输入沿着水平和垂直方向平均池化(X/Y , Avg Pool)得到特征映射形状 $[C, H, 1]$,再将两个特征 Concat 合并,对合并后的特征层进行 1×1 卷积操作(Conv2d),以捕捉水平和垂直维度之间的关系。应用标准化(BachNorm)和激活函数进一步处理特征,得到一个更加丰富的特征表示。将特征表示进行分组,分别代表水平、垂直方向的特征,对两个特征应用 1×1 卷积,以调整通道数,使其适应注意力计算。使用 Sigmoid 激活函数,得到两个维度上的注意力分数,并与原始输入特征图从宽度和高度方向相乘,得到 CA 注意力机制的输出。

2.4 损失函数

YOLOv8 检测头中的损失函数包括分类损失(CLS Loss)、回归损失 CIoU Loss+DFL Loss。将原有的回归损失中的 CIoU (Complete Intersection over Union) 损失替换为 MPDIoU (Modified Partial Differentiable IoU) 损失。

$$MPDIoU = IoU - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (1)$$

其中, IoU 为交并比, w 与 h 分别表示图片的宽度和高度, d_1 、 d_2 表示预测边界框与实际标注边界框之间的左上角和右下角点距离。

最终的边界框回归损失函数 \mathcal{L}_{MPDIoU} 为:

$$\mathcal{L}_{MPDIoU} = 1 - MPDIoU \quad (2)$$

其中, \mathcal{L}_{MPDIoU} 损失函数与 MPDIoU 的相似度呈反比关系,即 MPDIoU 越高, \mathcal{L}_{MPDIoU} 损失越低,这推动模型预测的边界框更加接近真实框。

最终训练学生模型时总损失为:

$$\mathcal{L} = \mathcal{L}_{MPDIoU} + \mathcal{L}_{DFK} + \mathcal{L}_{CLS} \quad (3)$$

在图 4 中,该文利用 1×1 卷积对改进的 YOLOV8 学生模型进行上采样,实现与教师网络模型 YOLOv8L 的通道进行对齐,分别软化概率分布,使用 KL 散度量两个概率分布之间的差异,从而利用教师网络的知识去引导学生网络,实现了更好的性能。Channel-wise distillation Loss 为:

$$\varphi(\theta(y^T), \theta(y^S)) = \varphi(\theta(y_c^T), \theta(y_c^S)) \quad (4)$$

其中, y^T 表示从 teacher 模型需要蒸馏的特征层中获取的 activation map; y^S 表示从 student 模型特征层中获取的 activation map; φ 表示计算 KL 散度; θ 表示将 activation value 转换成概率分布。

3 实验与结果

3.1 数据集

目前并没有应用较为广泛的、专门的婴幼儿表情数据集。该文从互联网上收集家庭日常活动、托儿所和户外活动等看护场景下 7 408 张的单脸婴幼儿 (0-

3 岁)和 500 张包含多脸的图片,建立自己的婴幼儿面部表情数据集 IEDB,其中多脸照片仅用作推理测试,样本示例如图 6 所示。然后将图像裁剪至 256×256 像素,并进行数据清洗。最后进行人工标记,使用数据标记工具 LabelImg^[24]在图像每个人脸区域绘制一个框,并标注出每张图片中的婴幼儿表情。婴幼儿的大脑和心理的发育远不如成人成熟,只能将其少量的心理变化通过简单的面部表情表现出来,不像成人那样可以展现比较复杂的面部表情。因此,该文所构建的婴幼儿面部图像数据库的表情仅区分为快乐、平静、哭泣、睡觉四种表情,单脸数据集划分结果如表 1 所示。

表 1 数据集划分情况

类别	样本数量			
	训练集	验证集	测试集	总计
快乐	1 249	356	178	1 783
平静	1 334	381	190	1 905
哭泣	1 515	432	216	2 163
睡觉	1 091	311	155	1 557
总计	5 189	1 480	739	7 048



图 6 自建的 IEDB 婴幼儿表情数据集

CK+数据集是在实验室环境下采集得到的,包含 123 名测试者的 593 个视频序列,选择其中表情表达比较明显的图像进行实验,共有 981 张图像,包括 7 种基本面部表情类别^[25]。数据集示例如图 7 所示。



图 7 CK+数据集示例

3.2 实验设备及评价指标

模型的开发语言是 Python 3.8,使用深度学习框架 PyTorch 作为网络框架。模型的硬件测试环境 CPU 选用海光 7285, 32C @ 2.0 GHz, GPU 选用 DCU Z100L, 32G 显存。

该文以准确率 (Accuracy)、平均精度均值 (mean Average Precision, mAP)、参数量 (number of parameters, Params)、浮点运算每秒 (GFLOPs)、每秒钟

检测到的帧数 (Frames Per Second, FPS) 作为评价指标。GFLOPs 用于度量模型或算法的复杂性,Params 表示模型的大小。一般来说,Params 和 GFLOPs 越小,表示模型所需的计算能力就越小,对硬件的性能要求就越低,FPS 越大,检测速度越快。

3.3 在 IEDB 图像数据集上的实验结果

将文中网络模型在自建的婴幼儿数据集上进行训练和测试,批量大小为 32,迭代次数设为 200 个 epoch,当平均精度没有显著提高时,程序会自动停止训练。其结果的混淆矩阵和 PR 曲线如图 8、图 9 所示。

图 8 显示了文中模型的混淆矩阵,描述了其在数据集中四类婴幼儿表情的预测准确性,并说明了预测之间的关系。在图中,行表示真实的标签,列表示预测的类别,对角线元素表示正确的检测率。可以看出,文中模型在每个类别中都达到了较高的精度。

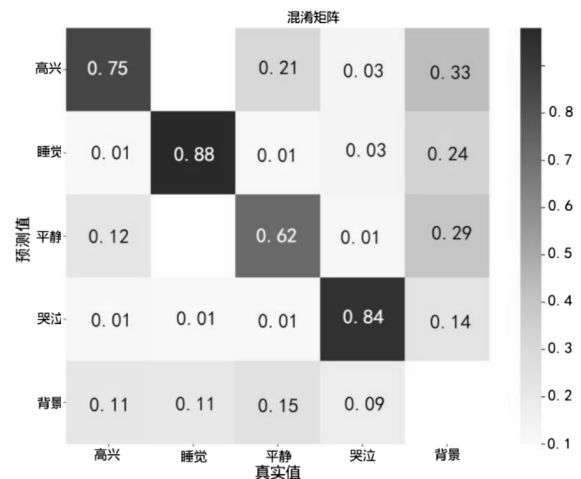


图 8 文中模型的混淆矩阵

图 9 显示了该模型的 PR 曲线。可以观察到,精度的变化率随着召回率的增加而增加。可以明显看出,该模型的 PR 曲线接近右上角,表明该模型有较高的查全率和精度。PR 曲线下的面积相对较大,说明模型的性能较好。此外,PR 曲线也很平滑,表明在文中模型中,召回率和精度之间存在相对稳定的关系。

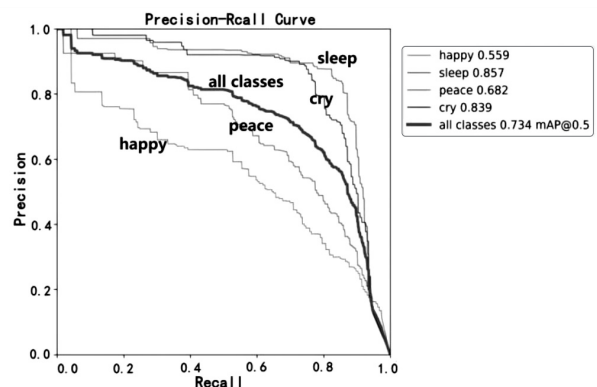


图 9 文中模型的 PR 曲线

检测结果如图 10 所示。与原始的 YOLOv8 模型相比,文中模型的检测结果更加精确,图 10 前两张的

检测框的置信度更高,第三张显示原始结果出现表情误检,文中模型能够准确识别。

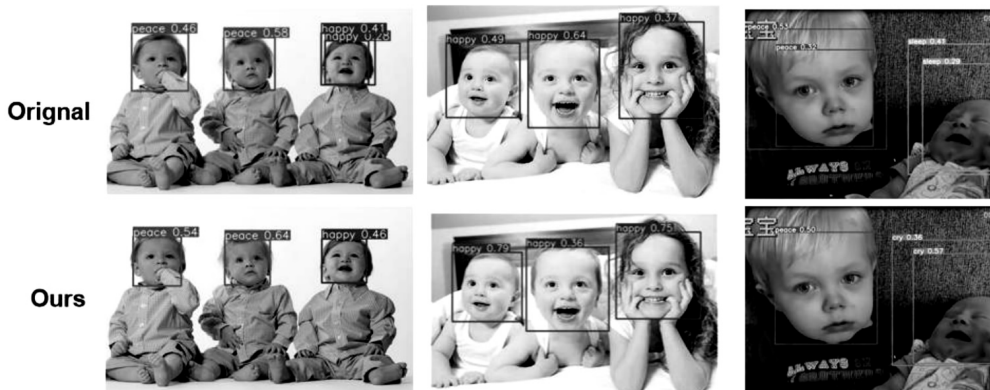


图 10 测试结果

3.4 消融实验对比

为了验证设计的改进的轻量化 YOLOv8 模型的有效性,基于婴幼儿面部表情数据集对其进行消融实验,并做出定量分析。通过控制变量法进行实验,结果如表 2 所示。在使用 CDSC 替换 C2f 之后,mAP 有所下降,但 GFLOPs 下降了 15.8%。在原模型上加入

CA 注意力机制,能在几乎不影响模型复杂度的情况下提升模型性能。加入知识蒸馏后模型的性能均能得到提升,mAP 最高,但 Params 仅有 2.6 MB,计算量 GFLOPs 为 7.1,帧率为 46.6,与原始模型相比性能显著提升,可证明该改进是有效的。

表 2 消融实验对比

Algorithms	CDSC	CA	KD	mAP50	mAP50-95	Params/MB	GFLOPs	FPS
YOLOv8n				0.722	0.329	3.0	8.2	35.6
A	✓			0.697	0.310	2.5	6.9	56.2
B		✓		0.728	0.336	3.0	8.2	44.2
C	✓	✓		0.722	0.321	2.6	7.1	46.1
D	✓		✓	0.717	0.312	2.5	6.9	40.5
E		✓	✓	0.730	0.334	2.6	7.1	45.2
Ours	✓	✓	✓	0.734	0.337	2.6	7.1	46.6

3.5 不同算法对比实验结果及分析

3.5.1 不同目标检测算法对比实验结果

如表 3 所示,对比其他模型发现,YOLOv8l 模型相对于 YOLOv8n,mAP 高出 4.71%,但是参数量大、检测速度相对较低。YOLOv5s 与二阶段的 Faster-

RCNN 相比,检测精度和检测速度都较差。该文提出的改进的 YOLOv8 模型的 mAP(50) 达到 73.4%,相对于原始的 YOLOv8n 有提高,高于其他对比的模型,并且参数量和 FPS 与原始 YOLOv8n 相比基本变化不大,实验效果好。

表 3 不同目标检测算法对比实验结果

Algorithms	mAP50	mAP50-95	Params/M	GFLOPs	FPS
Faster-RCNN	0.679	-----	137.05	370.2	14.3
YOLOv5s	0.625	0.258	7.0	15.8	30.3
YOLOv7	0.729	0.320	37.2	105.2	24.2
YOLOv7-tiny	0.726	0.310	6.0	13.2	34.4
YOLOv8n	0.722	0.302	3.0	8.2	35.6
YOLOv8l	0.746	0.339	43.7	164.8	30.3
Ours	0.734	0.337	2.6	7.1	46.6

3.5.2 与其他表情识别算法对比实验结果

由于大多表情识别均为不涉及位置的检测,可以从 YOLOv8 的评价标准中计算出准确率 (Acc) 与其他

表情识别算法进行对比,结果如表 4、表 5 所示。可以看出,文中改进的模型在 IEDB 数据集上的准确率达 0.754,CK+上准确率达 99.1%,模型参数量也最小。

表 4 在 IEDB 数据集上与其他表情识别算法的实验对比

Algorithms	Acc	Params/M
VGG16	0.611	138.1
ResNet50	0.702	25.6
MobileNet v2	0.711	3.4
MobileNet v3-SSE ^[26]	0.752	20.1
A-MobileNet ^[27]	0.801	3.4
MCLResNet ^[28]	0.737	8.7
Ours	0.754	2.6

表 5 在 CK+数据集上与其他表情识别算法的实验对比

Algorithms	Acc	Params/M
VGG16	0.952	138.2
ResNet50	0.941	25.5
MobileNet v2	0.924	3.4
MobileNet v3	0.935	3.3
EfficientNet ^[29]	0.967	169.6
Ours	0.991	2.6

4 结束语

该文设计了一种基于改进的 YOLOv8 与特征蒸馏相结合的端到端集成的多人脸表情识别网络模型,能够实时准确地检测出监控画面中所有婴幼儿的表情状态。将原始的 YOLOv8n 结合深度可分离卷积,大幅减小了参数量,引入 CA 注意力机制将位置信息嵌入到通道注意力中,使得轻量级网络能够在更大的区域上提取显著特征。使用 Channel-wise Knowledge Distillation 知识蒸馏,对学生模型进行蒸馏。最终的模型尺寸压缩至 2.6 MB,mAP@0.5 达到 73.4%。

然而,该研究仍然存在一些局限性。首先,数据量的不足仍是一个挑战,未来计划在不同的智能看护场景中积累更全面、多样的原始婴幼儿图片,以提高模型的泛化能力。其次,目前模型的功能仅限于表情识别,未来的研究可以考虑进一步扩展应用,如融合姿态信息和语音信息进行更全面的情感计算。

参考文献:

- [1] WEBB R, AYERS S, ENDRESS A. The city infant faces database: a validated set of infant facial expressions[J]. Behavior Research Methods, 2018, 50: 151-159.
- [2] TANG C, ZHENG W, ZONG Y, et al. Automatic smile detection of infants in mother-infant interaction via CNN-based feature learning[C]//Proceedings of the joint work-

shop of the 4th workshop on affective social multimedia computing and first multi-modal affective computing of large-scale multimedia data. Seoul: [s. n.], 2018: 35-40.

- [3] SUN Y, SHAN C, TAN T, et al. Detecting discomfort in infants through facial expressions[J]. Physiological Measurement, 2019, 40(11): 115006.
- [4] ZHANG L, XU C, LI S. Facial expression recognition of infants based on multi-stream CNN fusion network[C]//2020 IEEE 5th international conference on signal and image processing (ICSIP). Nanjing: IEEE, 2020: 37-41.
- [5] 邓源, 施一萍, 江悦莹, 等. 基于 MobileNetV2 与 LBP 特征融合的婴幼儿表情识别算法[J]. 电子科技, 2022, 35(8): 47-52.
- [6] 李聪聪. 基于深度学习架构的婴幼儿表情识别方法研究[D]. 武汉: 武汉工程大学, 2023.
- [7] 耿磊, 齐婷婷, 张芳, 等. 基于 MIFNet 的婴儿面部表情识别[J]. 计算机系统应用, 2023, 32(8): 42-53.
- [8] 黄嘉伟. 用于医疗看护的人脸表情识别技术研究[D]. 南京: 南京理工大学, 2023.
- [9] ZHAO G, YANG H, YU M. Expression recognition method based on a lightweight convolutional neural network[J]. IEEE Access, 2020, 8: 38528-38537.
- [10] WIKANNINGRUM A, RACHMADI R F, OGATA K. Improving lightweight convolutional neural network for facial expression recognition via transfer learning[C]//2019 international conference on computer engineering, network, and intelligent multimedia (CENIM). Surabaya: [s. n.], 2019:

- 1-6.
- [11] BIE M, LIU Q, XU H, et al. FEMFER: feature enhancement for multi-faces expression recognition in classroom images [J]. *Multimedia Tools and Applications*, 2024, 83 (2): 6183-6203.
- [12] CHOLLET F. Xception: deep learning with depthwise separable convolutions [C]//*Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*. Honolulu: IEEE, 2017: 1251-1258.
- [13] HOWARD A. Mobilenets: efficient convolutional neural networks for mobile vision applications [J]. *arXiv*: 1704.04861, 2017.
- [14] SIFRE L, MALLAT S. Rigid-motion scattering for texture classification [J]. *arXiv*: 1403.1687, 2014.
- [15] GUO Y, LI Y, WANG L, et al. Depthwise convolution is all you need for learning multiple visual domains [C]//*Proceedings of the AAAI conference on artificial intelligence*. Honolulu: AAAI, 2019: 8368-8375.
- [16] KAMAL K, YIN Z, WU M, et al. Depthwise separable convolution architectures for plant disease classification [J]. *Computers and Electronics in Agriculture*, 2019, 165: 104948.
- [17] LI W, CHEN H, LIU Q, et al. Attention mechanism and depthwise separable convolution aided 3DCNN for hyperspectral remote sensing image classification [J]. *Remote Sensing*, 2022, 14 (9): 2215.
- [18] CHEN G, CHOI W, YU X, et al. Learning efficient object detection models with knowledge distillation [J]. *Advances in Neural Information Processing Systems*, 2017, 30: 742-751.
- [19] SHU C, LIU Y, GAO J, et al. Channel-wise knowledge distillation for dense prediction [C]//*Proceedings of the IEEE/CVF international conference on computer vision*. Montreal: IEEE, 2021: 5311-5320.
- [20] HU Y, LIU G, CHEN Z, et al. Lightweight one-stage maize leaf disease detection model with knowledge distillation [J]. *Agriculture*, 2023, 13 (9): 1664.
- [21] KANG G, WANG J, ZENG F, et al. Lightweight detection system with global attention network (gloan) for rice lodging [J]. *Plants*, 2023, 12 (8): 1595.
- [22] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications [J]. *arXiv*: 2209.02976, 2022.
- [23] LV J, XU H, XU L, et al. An image rendering-based identification method for apples with different growth forms [J]. *Computers and Electronics in Agriculture*, 2023, 211: 108040.
- [24] ALJABRI M, ALAMIR M, ALGHAMDI M, et al. Towards a better understanding of annotation tools for medical imaging: a survey [J]. *Multimedia Tools and Applications*, 2022, 81 (18): 25877-25911.
- [25] LUCEY P, COHN J F, KANADE T, et al. The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression [C]//*2010 IEEE computer society conference on computer vision and pattern recognition-workshops*. San Francisco: IEEE, 2010: 94-101.
- [26] 雷晓鹏. 基于改进轻量卷积神经网络 MobileNetV3 的人脸表情识别 [J]. *现代计算机*, 2024, 30 (10): 29-34.
- [27] NAN Y, JU J, HUA Q, et al. A-MobileNet: an approach of facial expression recognition [J]. *Alexandria Engineering Journal*, 2022, 61 (6): 4435-4444.
- [28] 赵晓, 杨晨, 王若男, 等. 基于注意力机制 ResNet 轻量网络的面部表情识别 [J]. *液晶与显示*, 2023, 38 (11): 1503-1510.
- [29] KOONCE B. Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization [M]. Berkeley: Apress, 2021: 109-123.