基于改进 YOLOv7-tiny 飞行员表情识别算法研究

杨一豪[1]
YANG Yihao

摘 要 针对飞行员情绪识别中存在的精度不足和效率受限问题，文章提出了一种基于改进深度学习模型的情绪检测算法。首先，在现有模型中引入了 CBAM 注意力机制，以增强模型对关键特征的关注，抑制非关键区域的干扰；其次，采用 Ghost Conv 模块进行轻量化设计，在显著降低模型参数量和计算复杂度的同时保持高效的特征提取能力。通过实验对比分析了单独引入 CBAM 模块、Ghost Conv 模块以及两者结合使用的效果。实验结果表明，改进后的算法在情绪识别的检测精度上提升了 3.1%，计算复杂度（GFLOPs）降低了 21%，模型参数量减少了 10%，在准确性和实时性方面均取得了显著提升。该算法为飞行员情绪监测提供了一种高效、准确的解决方案。

关键词 YOLOv7；情绪识别；目标检测；CBAM；Ghost Conv

doi：10.3969/j.issn.1672-9528.2025.03.017

0 引言

近年来，随着航空业的快速发展，飞行员在高压环境下的疲劳与情绪波动问题日益严重，成为飞行安全的重要隐患。飞行员的状态直接影响飞行安全，因此实时监测和评估飞行员的情绪水平变得尤为重要。有效监测系统能够及时发现潜在的安全风险，从而提升飞行安全性和飞行员的身心健康[1-3]。

传统的目标检测方法主要依赖基于特征的算法，如 Haar[4] 特征级联分类器和 HOG（方向梯度直方图）特征检测，这些方法通过提取图像中的特征点来进行目标识别。然而，这些方法在处理复杂背景或多样化目标时，常常面临精度不足和速度缓慢的问题。

现代目标检测技术通过深度学习，取得了显著的进展。目标检测方法主要分为两阶段和一阶段两种类型。两阶段方法，如区域卷积神经网络（R-CNN）、快速区域卷积神经网络（Fast R-CNN[5]）和掩码区域卷积神经网络（Mask R-CNN[6]），首先生成候选区域，然后对这些区域进行分类和边界框调整。这些方法通常能够提供更高的检测精度，但处理速度相对较慢。

相比之下，一阶段目标检测方法，如 YOLO（you only look once）系列，将整个图像直接输入网络，并同时进行目标检测和定位。这种方法在速度上表现优越，能够迅速完成检测任务，但通常精度较低。为了弥补这一不足，本文使用 YOLOv7-tiny，引入了改进的网络架构和优化技术，进一步提高了检测精度，同时保持其较高的处理速度。

1. 西安石油大学 陕西西安 710065

（a）引入 Ghost Conv[7] 模块，它通过轻量级卷积操作显著减少了计算量和参数量。这一改进使得网络在保持高效性能的同时，满足了实时性的要求。

（b）引入 CBAM[8]（convolutional block attention module），它通过结合通道注意力机制和空间注意力机制，增强了特征表示能力。CBAM 能够有效地突出重要特征并抑制无关信息，从而提升了检测精度和网络的整体性能。

1 YOLOv7-tiny[9] 算法

YOLOv7 是 YOLO 系列的一个关键版本，相比于前代模型，在网络架构和算法优化上进行了显著改进。YOLOv7-tiny 是 YOLOv7 的一个轻量化版本，在保持高精度的同时也降低了参数和计算量，适用于边缘设备上。在 YOLOv7 中，创新性采用了 ELAN 结构作为骨干与特征提取网络，通过更密集的跳连接和并行下采样（结合 Max pooling 与步长为 2×2 的卷积）增强了特征表达能力与计算效率。同时，引入带有 CSP 结构的 SPP 模块，在不显著增加计算量的前提下扩大感受野，优化特征提取，实现了检测精度与速度的显著提升。

2 YOLOv7-tiny 目标检测网络改进

2.1 引入 Ghost Conv（幻影卷积）

在实际的应用场景中，神经网络模型的计算能力和运算速度常常受到硬件环境以及设备空间布局设计的限制。为了应对这些挑战，引入专门为移动设备或资源受限环境设计的网络结构，如 Ghost Conv 网络结构。其核心思想是利用线性操作生成丰富的特征图，同时减少冗余信息的产生，从而降

低模型对硬件存储空间的需求并提升运算速度。使其更加适合在实际部署场景中运行。

与传统的卷积操作相比，Ghost Conv 网络结构通过深度可分离卷积和线性操作生成特征图，显著减少了冗余信息的产生，从而降低了模型的参数量和计算量。这种轻量化设计使得 Ghost Conv 在嵌入式设备或资源受限的环境中具有更高的效率和实时性，尤其适用于对速度和存储要求严苛的应用场景。

图 1 为传统的卷积，传统卷积在特征提取过程中会产生大量的特征图，这些特征图虽然有助于捕捉图像中的复杂信息，但同时也带来了计算量大、参数量多以及硬件存储空间占用高等问题。图 2 为 Ghost Conv，其中 Conv2d 的参数分别为输入维度，输出维度，卷积核大小，stride 和 group。通过图 2 可知，输入的数据首先通过 1×1 的卷积核来调整维度，再通过一个深度可分离卷积来提取特征，最终将二者连接起来就是 Ghost Conv 的整体结构。

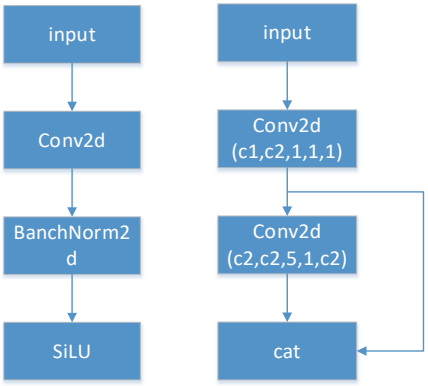

图 1 传统卷积

图 2 Ghost 卷积

普通卷积和 Ghost Conv 的计算量和参数量也存在巨大差异，假设输入数据的 shape 是 $c*h*w$，分别是输入的 channel、特征图的高和宽。经过第一次卷积后输出的 shape 为 $c*h'*w'$，分别是输出的 channel、输出特征图的高和宽。常规卷积核大小为 d，经过 s 次变换，普通卷积的额计算量和 Ghost Conv 的计算量对比：

$$r_s=\frac{n*h'*w'*c*k*k}{\frac{n}{s}*h'*w'*c*k*k+(s-1)*\frac{n}{s}*h'*w'*d*d}\approx s \quad (1)$$

普通卷积和 Ghost Conv 的计算量对比：

$$r_c=\frac{n*c*k*k}{\frac{n}{s}*c*k*k+(s-1)*\frac{n}{s}*d*d}\approx s \quad (2)$$

2.2 引入 CBAM 注意力机制

CBAM（convolutional block attention module）注意力机制的优势有多个方面。之前的注意力机制往往只关注通道注意力或空间注意力中的一个方面，这在一定程度上限制了它们对特征信息的全面捕捉和利用。而 CBAM 则巧妙地结合了通道注意力和空间注意力两种机制，使得模型能够同时关注特征图的通道和空间维度，从而更全面地捕捉和利用特征信息。这种双重注意力的结合不仅增强了模型对特征的感知能力，还提高了模型在不同任务上的适应性和鲁棒性。此外，CBAM 的设计简洁而高效，能够在不显著增加计算成本的前提下，有效提升模型的性能。因此，CBAM 注意力机制相比之前的注意力机制，在特征信息的全面捕捉、模型性能的提升以及计算效率的优化等方面都表现出了更好的性能。

CBAM 模块的核心思想是对输入特征图进行两阶段的精炼：首先通过通道注意力模块关注于“哪些通道是重要的”，然后通过空间注意力模块关注于“在哪里”是一个有信息的部分。这种双重注意力机制使 CBAM 能够全面捕获特征中的关键信息。

图 3 为 CBAM 的总体流程，主干网络的特征图为 $F\in\mathbb{R}^{C\times H\times R}$，由 CBAM 产生的 1D 通道注意力特征图为 $M_c\in\mathbb{R}^{C\times 1\times 1}$，2D 空间注意力特征图为 $M_s\in\mathbb{R}^{1\times W\times H}$。

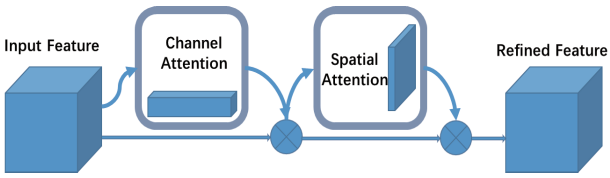

图 3 CBAM 注意力机制

（1）通道注意力机制

在神经网络中，特征图的每一个通道都可以被视为一个特征检测器，在训练过程中应该给重要的特征层加大的权重，这样才能有效提升校测效率。为了更高效地计算通道注意力特征采用了平均池化和最大池化两种方法，将两种池化的结果送入多层感知机（MLP）中提取最终的通道注意力特征图 $M_c\in\mathbb{R}^{C\times 1\times 1}$，如图 4 所示。

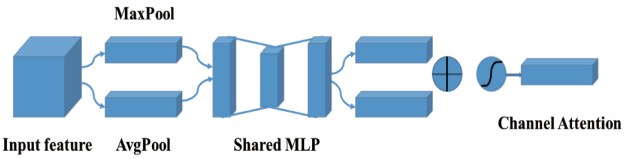

图 4 通道注意力机制

（2）空间注意力机制

空间注意力主要关注的是在特征图中哪个区域更为重要。空间注意力机制的输入是通道注意力和原始输入，通过压缩特征图的通道得到 $H\times W\times 1$ 的两个特征图，得到这两个特征图后再通过一个卷积核大小为 7×7 的卷积运算得到一个 $H\times W\times 1$ 特征图。最后经过激活层即可得到最终的空间注意力特征，如图 5 所示。

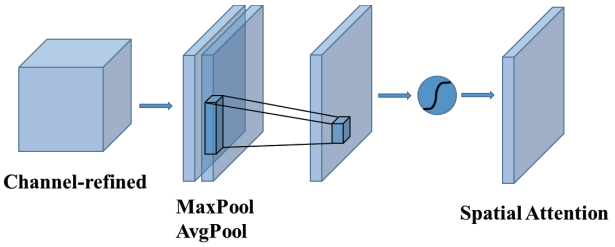

图 5 空间注意力机制

3 实验及结果分析

3.1 构建人脸情绪数据集

为了全面评估人脸表情识别算法的性能，本文从多种公开资源中搜集并整合了一个包含约 4 000 张图片的人脸表情数据集。数据集中涵盖了“愤怒（Angry）”“厌恶（Disgusted）”“高兴（Happy）”“中立（Neutral）”“焦虑（Anxious）”“悲伤（Sad）”“害怕（Scared）”和“惊讶（Surprised）”8 种表情类别，充分保证了表情的多样性与代表性。在数据划分上，按照 7:2:1 的比例分为训练集、验证集和测试集，分别用于模型的训练、参数调优和性能评估。这种划分方式既保证了模型训练的充分性，又增强了实验结果的科学性和可靠性，为评估表情识别算法的实际效果奠定了基础。

3.2 实验环境

本实验使用的深度学习框架为 PyTorch，版本为 2.1.2 的 GPU 版本；编译语言为 Python3.11.9；CUDA 版本为 12.3；操作系统为 Win10 专业版；CPU 为 12th Gen Intel(R) Core(TM) i5-12490F 3.00 GHz；运行内存 32 GB；GPU 为 NVIDIA GeForce RTX3060 12 GB。

3.3 评价指标

在本实验中，为全面评估 YOLOv7 的目标检测性能，采用了 mAP（mean average precision）作为主要评价指标。mAP 是目标检测领域广泛应用的评价标准，它综合考虑了模型的精度和召回率，提供了对目标检测效果的全面评估。mAP 的计算过程包括以下几个步骤：

（1）精度 - 召回曲线：对于每个类别，绘制精度 - 召回曲线（precision-recall curve），该曲线显示了在不同置信度阈值下模型的精度和召回率。

（2）平均精度（AP）：通过计算精度 - 召回曲线下的面积，得到每个类别的平均精度（AP）。常见方法包括对精度召回曲线进行插值，以确保更稳健的 AP 计算。

（3）mAP 计算：对所有类别的 AP 值取均值，得到 mAP。mAP 值越高，表示模型在所有类别上的整体检测性能越优。

3.4 消融实验结果与分析

本文以 YOLOv7-tiny 为基准，依次组合添加改进措施，进行一系列消融实验，实验结果如表 1 所示。

表 1 消融实验结果

方法	CBAM	Ghost Conv	*P*/%	*R*/%	mAP50/%	Params/10^6	FLOPS/10^9
base			58.7	78	0.72	6.0	13.2
A	√		61.2	81.3	0.756	6.3	13.7
B		√	59.1	78.4	0.731	5.1	9.9
C	√	√	61.8	82.3	0.759	5.4	10.4

方法 A 引入 CBAM 模块后，YOLOv7-tiny 模型的精度有所提升，*P*、*R* 和 mAP50 分别提高了 2.5%、3.3% 和 3.6%，尽管参数量和计算量略有增加，但整体性能得到了显著提升；方法 B 通过引入 Ghost Conv 模块，使得模型的 *P*、*R* 和 mAP50 分别提升了 0.4%、0.4% 和 1.1%，同时减少了计算量；方法 C 结合了 CBAM 和 Ghost Conv 模块，使得精度进一步提高，*P*、*R* 和 mAP50 分别提升了 3.1%、4.3% 和 3.9%，虽然参数和计算量有所增加，但整体性能在精度和效率上达到了最佳平衡。

3.5 对比实验分析

为验证本文算法相比当前算法的优越性，将本文算法与 Faster R-CNN、SSD、YOLOv3-tiny[10]、YOLOv4s[11]、YOLOv5s[12]、YOLOv6s[13] 进行比较，实验结果如表 2 所示。

表 2 对比实验结果

方法	*P*/%	*R*/%	mAP50/%	Params/10^6	FLOPS/10^9
Faster R-CNN	60.6	81.3	73.2	41.2	206.7
SSD	55.4	76.8	66.1	24.0	117.1
YOLOv3-tiny	57.1	79.3	69.9	8.8	17.6
YOLOv4s	58.1	80.2	71.5	6.3	14.5
YOLOv5s	59.1	80.8	72.1	7.2	16.5
YOLOv6s	60.2	81.5	73.0	5.8	12.2
Ours	61.8	82.3	75.9	5.4	10.4

从表 2 实验结果可以看出，本文算法在精度和计算效率上均显著优于其他对比方法。在性能方面，本文算法的精度 *P*、召回率 *R* 以及 mAP50 分别达到了 61.8%、82.3% 和 75.9%，相比主流的 Faster R-CNN、YOLOv5s 和 YOLOv6s 均有不同程度的提升，尤其在 mAP50 上表现尤为突出。同时，在计算效率方面，本文算法的参数量仅为 5.4×10^6，FLOPS 为 10.4×10^9，显著低于 Faster R-CNN 和 SSD 等模型，与轻量化模型 YOLOv5s 和 YOLOv6s 相比也展现出了更优的效率优势。

4 检测效果

为了更直观展示本文改进算法的检测效果，图 6 为原始 YOLOv7-tiny 模型的检测结果，图 7 为改进版 YOLOv7-tiny 模型的检测效果。从图像中可以明显看出，改进后的模型在同一幅图片中的检测结果更加精确。具体而言，改进后的模型不仅能更好地识别目标的位置，还显著提高了对物体的检测准确性。这样的改进有效地提升了模型在实际应用中的可靠性，尤其是在复杂环境下的物体检测任务中，表现得尤为突出。

图 6 YOLOv7-tiny

图 7 改进版 YOLOv7-tiny

5 结语

飞行员情绪识别对于提高航空安全和飞行员健康管理至关重要。为了提高情绪识别的准确性和实时性，本文在现有算法基础上进行了优化。本文改进了特征提取网络，使模型在复杂情绪识别任务中表现更好。同时，采用了轻量化设计，保证了模型在保持高精度的同时，具有较快的处理速度。实验结果表明，改进后的算法在准确性和实时性上都有显著提升，能够满足实际应用的需求。尽管如此，算法在某些复杂情绪的识别精度上仍有改进空间，未来将继续优化并关注其实际应用效果。

参考文献：

[1] 孙可新，鲁慧民 . 基于深度学习的人脸识别仿真技术研究与实现 [J]. 计算机仿真 , 2024, 41(9): 189-193.

[2] 邓科豪 . 基于深度时空网络的面部表情识别 [D]. 西安 : 西安理工大学 , 2024.

[3] 李帅超 . 注意力机制与特征加权融合策略的微表情识别研究 [D]. 北京 : 中国人民公安大学 , 2024.

[4] 林齐发，吴晨曦，邹鑫 . 基于 Haar-like 特征的人脸检测算法研究与应用 [J]. 现代计算机 , 2024,30(17):13-17.

[5] REN S Q, HE K M, SUN J, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[6] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 42: 386-397.

[7] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1577-15876.

[8] WOO S, PARK J C, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision-ECCV 2018: 15th European Conference. NewYork: ACM, 2018: 3-19.

[9] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway: IEEE, 2023: 7464-7475.

[10] REDMON J, FARHADI A. YOLOv3: an incremental improvement[DB/OL].(2018-04-08)[2024-03-19].https://doi.org/10.48550/arXiv.1804.02767.

[11] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[DB/OL].(2020-04-23)[2024-05-13].https://doi.org/10.48550/arXiv.2004.10934.

[12] ZHANG Y, GUO Z Y, WU J Q, et al. Real-time vehicle detection based on improved YOLOv5[J]. Sustainability, 2022, 14(19): 12274.

[13] LI C Y, LI L L, ZHANG B, et al. YOLOv6: a single-stage object detection framework for industrial applications[DB/OL]. (2023-09-20)[2024-06-14]. https://doi.org/10.48550/arXiv.2209.02976.

【作者简介】

杨一豪（1999—），男，陕西咸阳人，硕士研究生，研究方向：图像处理。

（收稿日期：2024-11-25）