改进 YOLOv8s 的轻量化人脸识别算法

龙子晗[1] 肖小玲[1]
LONG Zihan XIAO Xiaoling

摘 要 针对人脸检测任务中检测模型参数量大、检测精度低的问题，文章基于 YOLOv8 算法和 MobileNetV3 网络，提出一种轻量化人脸目标检测算法。首先，使用基于 MobileNetv3 网络替换 YOLOv8s 的骨干网络，减少参数量和计算量，提高检测速度；在此基础上，在骨干网络中添加额外的 C2f 模块，以此提高模型性能，增加其特征提取的能力；最后，在颈部网络添加额外注意力机制-多尺度空洞结构来进一步提高模型性能。实验结果显示，相较于基准模型，改进算法平均检测精度提高了 2 个百分点，参数量降低了约 50%，验证了其有效性。与其它主流模型相比较，这一算法也有良好的表现。

关键词 轻量化；YOLOv8；MobileNetV3；人脸识别；目标检测

doi：10.3969/j.issn.1672-9528.2024.12.045

0 引言

随着时代的发展，人脸检测技术融入日常生活，从遍布各地的门禁系统，到刷脸支付等新兴交易模式，该技术无处不在，在给生活缔造极大便捷性的同时，更深刻重塑与优化了现代生活的诸多环节与体验。但随着技术的发展，普通的人脸检测技术弊端也逐渐显露。欺诈者利用图片或手机中的图片竟能成功突破人脸识别检测，这一漏洞对用户的生命与财产安全构成了严峻威胁。鉴于此，如何有效区分手机照片、普通图片与真人便成为了本文的核心研究重点[1]。

随着深度卷积神经网络不断取得进展，在此背景下，目标检测根据模型框架的不同，分为单阶段和双阶段算法检测两大类别。其中，双阶段算法需先进行预选框的选取，再通过对于特征的提取，完成分类。以 Faster R-CNN[2] 为代表，虽然准确度较高但检测效率较低；单阶段的算法则是一次性完成分类和回归的任务，具有检测速度快，模型相对较小的特点。而 YOLO 系列算法正是单阶段检测算法的代表算法，虽然检测精度有所下降但是检测速度明显加快。以上算法在 COCO 等数据集上的测试中拥有较高的精度，但是仍然存在模型过于复杂、精准度不算高等问题，仍有优化空间。

由于本文需要完成轻量化人脸识别的任务，因此选择单阶段的目标检测算法 YOLOv8。为了让算法在人脸识别任务中有更好的表现，本文对原算法做出了相应改进。在 Marcelo 等人[3] 的论证方法中，DSConv 模块可以很容易地替换到标准神经网络架构中，并实现更低的内存使用和更高的计算速度。

而对于模型过于复杂的问题。Han 等人[4] 提出的 GhostConv 模块，该模块将特征图分成了两部分，在其中一半的通道上执行深度可分卷积后，再和另一部分的原始特征通道拼接，大大减少参数数量并提高冗余特征利用率。

上述方法虽然在精准度和轻量化上都有改进，其中，精准度高的模型，参数的复杂度也随之增大。而轻量化的模型，尽管参数量减少，却在一定程度上降低了精准度。因此，本文基于 YOLOv8 网络提出了一种轻量化的人脸识别算法。首先用 MobileNetV3[5-6] 代替了 YOLOv8 的骨干网络，使得模型轻量化。与此同时，添加更多 C2f 模块，提高模型特征提取的能力，并且添加额外注意力机制多尺度空洞[7]，从而提高模型精准度。通过这样的方式，在不牺牲模型性能的同时，完成了模型的轻量化改造，使得该模型在人脸识别上达到了更完美的效果。

1 YOLOv8 网络结构

YOLOv8 网络在目标检测的精度和速度两方面都有不错的表现，是目前广泛使用的目标检测网络之一。该模型分为 YOLOv8n、YOLOv8s、YOLOv8m、YOLOv8l 和 YOLOv8x 五个版本，这五个模型网络结构相同，只有网络深度和网络宽度略有不同。本文选用较为轻量化的 YOLOv8s 版本，其整体的网络结构如图 1 所示。YOLOv8s 的网络结构由数据输入（Input）、主干网络（Backbone）、颈部网络（Neck）和检测头（Head）构成。其中主干网络 Backbone 是由 Conv、C2f、SPPF 三个模块构成，它的作用是用来提取目标特征。Neck 部分则是由 C2f、上采样模块（Upample）和相加模块（Concat）构成，其作用则是将输入的不同尺度的特征图进行特征融合，得到更加详细的特征信息。最后将这些特征图

1. 长江大学 湖北荆州 434100

输入 Head 中。Head 部分由几个 Conv 构成，其作用是进行目标分类和预测。

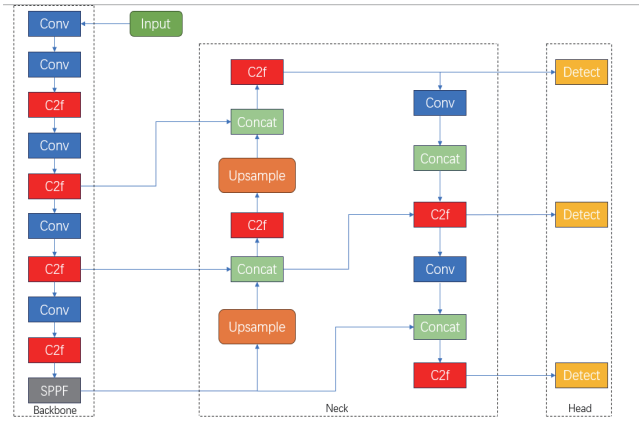

图 1 YOLOv8s 结构

其中，Conv 模块由卷积层（Conv2d）、批归一化层（batch normalization，BN）和激活函数（SiLU）组成。

C2f 模块参考了 YOLOv5 的 C3 模块以及 YOLOv7 的 ELAN 的思想进行的设计。

在 Backbone 的最后采用了快速空间金字塔池化（spatial pyramid pooling fast，SPPF），将不同感受野的特征图融合，以此提高特征图的表达能力。

2 YOLOv8s 的改进

2.1 轻量化网络 MobileNet

MobileNetV3 是由谷歌推出的轻量级网络，MobileNetV1 使用了深度可分离卷积，MobileNetV2 则在此基础上又新增了瓶颈残差模块，MobileNetV3 网络一方面吸收了 MobileNetV2 的优点，另一方面增加了注意力机制 SE（Squeeze and Excitation）模块以及激活函数 h-swish 的应用，在提升效率和可靠性的同时也显著减少了参数量。因此，本文将用 MobileNetV3 网络代替原来的骨干网络。

通道分离卷积是 MobileNet 系列的主要特点，也是其能成为轻量级算法的主要因素，它能够更加有效地处理复杂问题，降低计算量和参数量，加快运行速度，提高模型效率。通道可分离卷积由 channel 方向通道可分离卷积和正常的 1×1 逐点卷积两部分组成，通过使用滤波器和线性组合技术使得参数的复杂度大规模下降，模型效率提高。其算法流程如图 2 所示。

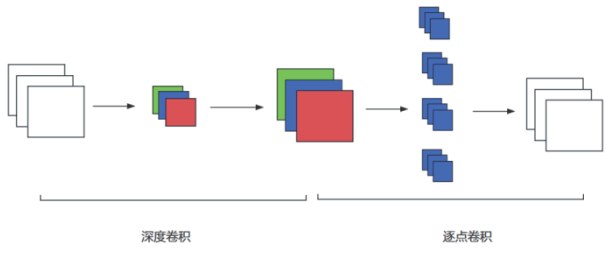

图 2 MobileNetv3 流程结构

MobileNetV3 算法是该系列较为优秀的算法，突出体现了该系列算法的尺寸小、参数量少、精度高等特点，在各种方面如图像分类、检测等领域都有不俗表现。作为 MobileNetV3 算法的基础结构，Bneck 模块采用 1×1 卷积技术，可以提高维度从高维空间获取信息，得到更加丰富的图像特征，配合上后面的 SE 注意力机制模块对各种维度中的信息进行调节，提高算法的性能，最后加上残差链接对输入、输出都进行残差处理更加增强了算法的可靠性。其结构如图 3 所示。

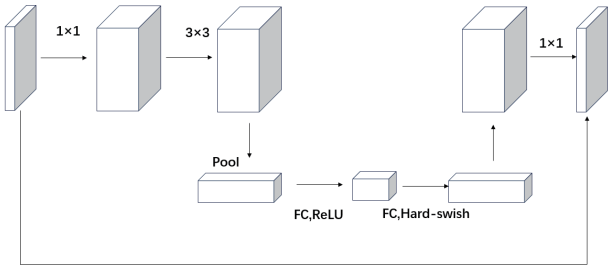

图 3 Bneck 结构

2.2 多尺度空洞注意力（MSDA）

注意力机制可以使感兴趣的区域更加准确地被捕获，帮助网络更快地识别目标，能够有效在复杂场景中更快找到显著区域。

注意力机制 MSDA 能够模拟小范围内的局部和稀疏的图像块交互，从而减少全局注意力模块中存在大量的冗余。在浅层次上，注意力矩阵具有局部性和稀疏性两个关键属性，这表明在浅层次的语义建模中，远离查询块的块大部分无关，全局注意力模块中存在大量的冗余。

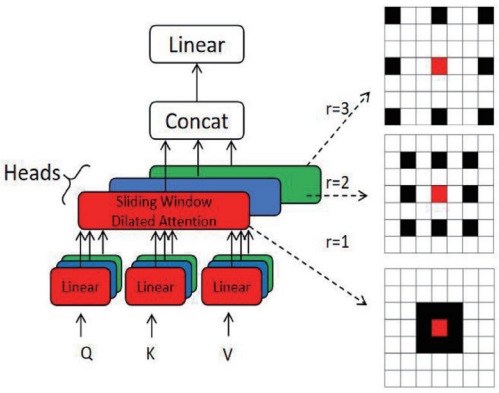

图 4 MSDA 工作原理

图 4 展示了多尺度扩张注意力（MSDA）的工作原理。在 MSDA 中，特征图的通道首先被分割成不同的头部，然后每个头部内部使用不同的扩张率 r（dilation rates）来执行自注意力操作。这些操作在围绕红色查询块的窗口内的彩色块之间进行。

图 4 中事例展示了三种不同的扩张率（r=1,2,3），它们

分别对应不同的感受野大小（3×3，5×5，7×7）。每个头部的自注意力操作针对的是其对应的扩张率和感受野。这样，模型能够在不同的尺度上捕捉图像特征，这些特征随后被连接在一起，并送入一个线性层进行特征聚合。

这种设计允许模型在不同的尺度上理解图像，从而提高对图像内容的整体理解。通过这种方法，MSDA 不仅可以捕捉局部细节，也能够感知到更广泛区域的上下文信息，增强了模型的表现力。

2.3 改进后 YOLOv8s 网络

基于上述信息对于原始的YOLOv8s的网络进行了改进，在骨干网络中利用 MobileNetV3-small 网络替换了原来的网络结构，为提高模型的性能在 MobileNetV3 网络结构的基础上添加了额外的 C2f 模块。

另外，为增强对细节的感知，在 Neck 部分添加了 MSDA，并将其和C2f相结合，想要让模型更加关注某些细节，提高模型性能。整体改进后的YOLOv8s网络结构如图5所示。

其中，CBH 模块即一个普通卷积 Conv 和一个 BN 模块以及 MobileNetv3 新提供的一个激活函数 h-swish 拼接而成。而 Bneck 模块也是原 MobileNetv3 中多个 Bneck 模块串联形成。

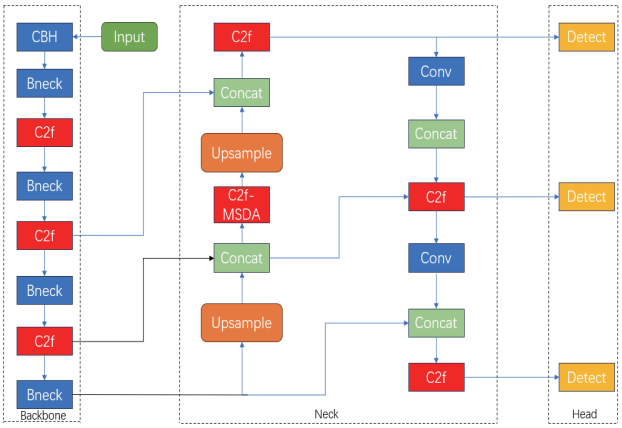

图 5 改进后 YOLOv8 网络结构

3 实验结果和分析

3.1 实验数据集

实验数据集为自建数据集，其中的数据来源于 https://universe.roboflow.com 网站，其内容分为四个种类，真人、相片、手机照片、用图片遮掩的真人，尺寸像素均为 640×640。

3.2 实验环境和设备

实验环境硬件配置：CPU 为 Intel i7 97500H，GPU 为 NVIDIA GTX 1660 Ti，操作系统为 Windows11，编译环境为 Python3.8.18+PyTorch1.8.1+CUDA10.2。在实验中迭代次数设置为 300 次，batch size 设置为 4。

3.3 评估指标

实验过程中不同的算法性能的好坏是通过某些指标进行表示的，如精确度（precision，P）、召回率（recall，R）、平均精度（average precision，AP）和平均精度均值（mean average precision，mAP）等作为评价指标。

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{AP} = \sum_{i=1}^{n-1} \frac{(R_{i+1} + R_i)(P_{i+1} + P_i)}{2} \tag{3}$$

$$\mathrm{mAP} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{AP}_i \tag{4}$$

式中：TP（true postives）为预测正确的正类样本数量；FP（false postives）为预测错误的正类样本数量；FN（false negatives）为预测错误的负类样本数量；n 为数据集中的类别数。设定 AP 的 IOU 检测阈值为 0.45，看在此结果下 mAP 的数值。本文将综合考虑 mAP 和参数量作为模型性能的评估标准。

3.4 结果分析

如图 6 所示，即训练后的一部分结果图。第一行第一张图片，识别为 1，代表这是真人。第一行第二张图片，有 1 和 0 两个框，表示一个是真人，一个是手机照片。第二行第一张图片，显示为 2 表示是用图片遮掩的真人。第二行第二张图片和第三张图片都是 3，显示这是普通照片。其余的几幅图也都成功识别了图上的真人、相片、手机照片、用图片遮掩的真人这四类。

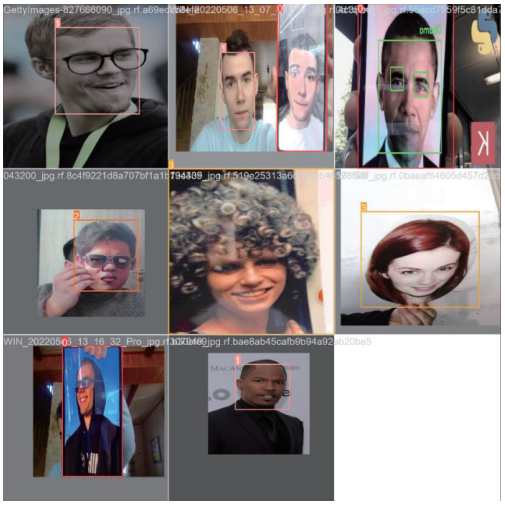

图 6 训练结果

改进算法 1 是在原算法 YOLOv8s 的基础上，将其 Backbone 模块更换成为了 MobileNetV3 模块；改进算法 2 是在改进算法 1 的基础上，在 MobilenetV3 模块中添加额外的

C2f 模块；本算法则是在改进算法 2 的基础上添加了额外的注意力机制 MSDA。

根据表 1 的数据显示，改进算法 1 在将骨干网络替换为 MobileNet 之后参数量有明显下降，mAP 也有所下降。改进算法 2 在添加额外的 C2f 模块之后 mAP 有了提高，但和原算法相比还是有些许差距。而本算法，在参数量比改进算法 2 更小的情况下，mAP 有了大幅度提高，甚至高于原算法，并且参数量相比于原算法也降低了约 50%。通过以上对比可知，本算法不仅在 mAP 上有提高，在参数量上也大幅度减小，简化了模型的规模。

表 1 不同改进方法对算法性能的影响

算法	MobileNetV3	C2f	MSDA	参数量 /10^6	mAP/%
原算法	—	—	—	3.00	79.8
改进算法 1	√	—	—	1.18	75.2
改进算法 2	√	√	—	1.47	77.7
本算法	√	√	√	1.45	80.9

将本算法和其它主流算法如原算法 YOLOv8s、YOLOv5n、YOLOv5s、YOLOv3tiny 相比较，其比较结果如表 2 所示。

表 2 不同算法性能对比

算法	参数量 /10^6	mAP/%
原算法	3.00	79.8
YOLOv5n	2.50	77.5
YOLOv5s	9.11	77.9
YOLOv3tiny	8.67	64.4
本算法	1.45	80.9

通过表 2 的对比可知，与原算法 YOLOv8s 相比，本算法的 mAP 提高了 1.1 个百分点，参数量下降了 51.7%；与模型 YOLOv5n 相比，本算法的 mAP 提高了 3.4 个百分点，参数量下降了 42%；与模型 YOLOv5s 相比，本算法的 mAP 提高了 3 个百分点，参数量下降了 84.1%；与模型 YOLOv3tiny 相比，本算法的 mAP 提高了 16.5 个百分点，参数量下降了 83.3%。通过上述算法的对比实验，本文提出的算法在保证 mAP 的同时，大大缩减了参数量，达到了精准度和轻量化的统一。

4 结论

为了实现人脸检测，本文通过选取一个由真人、相片、手机照片、由照片遮掩的真人组成的数据集来实现这一目的，同时考虑到降低部署的成本，以及提高模型的性能。在 YOLOv8s 算法的基础上，通过使用 MobileNetV3 网络结构替换原本的骨干网络，大幅减小了原模型的参数量，并在此基础上增加 C2f 模块，添加注意力机制 MSDA 以此来保证模型的性能。经过这样的改进，在减少了参数量，使模型轻量化的同时，增加了模型性能，使得 mAP 有了不小的提升。通过对比，本文提出的模型和原模型相比，参数量减少了一半，mAP 也有一定的提升。此外，和其它优秀目标检测算法相比，在保持较高 mAP 的同时，将参数量控制在了一个较小的范围。在人脸检测的任务中有较为优异的表现。

参考文献：

[1] 张晋婧 . 融合注意力机制的轻量级人脸识别模型研究 [D]. 太原 : 中北大学 ,2023.

[2] 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.

[3] MARCELO G A, VICTOR P, ROGER F. Dsconv: efficient convolution operator[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 5148-5157.

[4] HAN K, WANG Y H, TIAN Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Confer ence on Computer Vision and Pattern Recognition (CVPR).Piscataway:IEEE, 2020: 1577-1586.

[5] ANDREW G H , ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[DB/OL]. (2017-04-17)[2024-08-26].https://doi.org/10.48550/arXiv.1704.04861.

[6] 刘宏利 , 侣永磊 , 邵磊 , 等 . 基于改进 YOLO v4 的夜间车辆检测算法研究 [J/OL]. 天津理工大学学报 ,1-8[2024-05-29]. http://kns.cnki.net/kcms/detail/12.1374.N.20240515.1448.007.html.

[7] 沈建华 . 复杂交通场景下车辆行人检测算法研究 [D]. 保定 : 河北大学 , 2023.

【作者简介】

龙子晗（2001—），男，湖北襄阳人，硕士研究生，研究方向：目检测与计算机视觉。

肖小玲（1973—），女，湖北荆州人，博士，教授，研究生导师，研究方向：智能信息处理与网络安全。

（收稿日期：2024-09-10）