

基于改进YOLOv5s的小人脸检测

周丽芳, 胡振, 刘波

(重庆邮电大学软件工程学院, 重庆 400065)

摘要: 目前在复杂的真实应用场景中, 小人脸检测任务面临诸多挑战, 如人脸尺度小、光照变化剧烈和精度较低等。针对现有模型容易忽视小人脸检测的问题, 提出了一种基于卷积核注意力机制的小人脸检测模型 SK-YOLOv5s。首先, 设计一种小人脸增强模块, 将浅层特征与深层特征进行融合与上采样, 提高了小人脸特征图的分辨率, 从而增强其特征; 其次, 在模型中引入 SKNet 注意力机制, 能够多尺度自适应调节感受野大小, 提高小人脸的检测率; 最后, 采用 EIoU 作为目标损失函数、FReLU 作为激活函数, 直接减小预测与真实边界框之间的宽度和高度差异, 增强特征图的非线性表达能力, 提高小人脸检测的精度和稳定性, 相较于 YOLOv5s, 所改进模型在 WIDER FACE 数据集上的 mAP 提升了 0.079。实验结果证明了所改进模型在真实场景下进行小人脸检测的可行性。

关键词: 小人脸检测; 注意力机制; YOLOv5s

中图分类号: TP311.5

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202438

Small face detection based on improved YOLOv5s

ZHOU Lifang, HU Zhen, LIU Bo

School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Abstract: At present, in the complex real-world application scenarios, the task of small face detection encounters numerous challenges, which include small face scale, abrupt lighting changes and low accuracy. In order to address the concerns of overlooking small face detection within existing models, this study introduced a novel small face detection model termed SK-YOLOv5s which was based on convolutional kernel attention mechanism. Firstly, we proposed a small face enhancement module to fuse and upsample multi-layer features, which enhances the resolution of small face feature maps and strengthens their distinctiveness. Subsequently, we incorporated the SKNet attention mechanism into the model, which can adaptively adjust receptive field sizes across multiple scales and enhance the detection efficacy of small face. Finally, EIoU was utilized as the loss function, which directly reduced the width and height discrepancies between predicted and actual bounding boxes, and FReLU was utilized as the activation function, which could enhance the nonlinear expressiveness of feature maps to improve the precision and stability of small face detection. The performance of the enhanced model on the WIDER FACE dataset demonstrated mean average precision improvement of 7.9% over YOLOv5s. The experimental results demonstrate the viability of the enhanced model for small face detection in real-world scenarios.

Key words: small face detection, attention mechanism, YOLOv5s

收稿日期: 2024-03-27; 修回日期: 2024-09-24

通信作者: 周丽芳, zhoulf@cqupt.edu.cn

基金项目: 国家自然科学基金项目 (NO.62276039); 贵州大学开放基金 (NO.KY-2022 (376))

Foundation Items: National Natural Science Foundation of China (NO.62276039), Guizhou University Open Foundation (NO.KY-2022(376))

0 引言

人脸检测技术在许多现实场景中都具有广泛的应用,包括人脸识别、目标跟踪^[1]、表情分析^[2]和属性识别^[3]等。在车站、机场、商场等高密度人群场景下,人脸尺度较小,安防系统需要通过检测小人脸,及时发现可疑人员或异常情况,提高监控效率。目前基于改进YOLOv5s的小人脸检测研究有对算法的优化、对性能的提升以及对更多应用的兼容。如邱天衡等^[4]设计了更有效的小人脸检测网络架构,并且优化了损失函数,使用了新的注意力机制。Qi等^[5]在保持高检测准确率的同时,还能够实现实时的小人脸检测,这对于需要快速响应的应用场景(如安全监控、人脸认证等)非常重要。高强等^[6]将小人脸检测优化技术集成到行人检测任务中,使用了选择性内核网络(selective kernel network, SKNet)^[7]注意力机制,动态融合不同感受野的特征图,提高对不同行人特征的利用率。

虽然近十年来小人脸检测技术有了很大突破,但是小人脸检测任务仍然面临一些问题和挑战。首先是人脸尺度问题,小人脸通常在图像中具有较小尺度,与常规人脸检测相比,小人脸检测分辨率较低,特征更加微弱,难以被准确地检测;然后是遮挡问题,在一些人群密集的场景中,小人脸往往会被人或物体遮挡,使部分小人脸无法被完整地观察到,这进一步影响检测的准确性;最后是光照和角度变化问题,光照条件和拍摄角度的变化会影响小人脸在图像中的表现形式,可能导致部分小人脸的特征被丢失或变形,给小人脸检测带来困难。综上所述,小人脸检测技术在多个领域具有重要的应用价值,并面临很多挑战和问题。

针对小人脸容易被忽视检测的问题,提出了基于卷积核注意力机制的小人脸检测模型SK-YOLOv5s。首先,该模型通过设计一种小人脸增强模块(small face enhancement module, SFEM),提高小人脸的分辨率,从而增强其特征;其次,引入卷积核注意力机制,通过多尺度自适应调节感受野大小,提高小人脸的检测率;再次,使用EIou作为目标损失函数,提升了对小人脸的检测准确性和稳定性;最后,改用FReLU作为激活函数,增强特征图的非线性表达能力。

与YOLOv5s相比,改进的SK-YOLOv5s在WIDER FACE数据集^[8]中IoU阈值为0.5时,平均

精度均值提升了0.079,通过实验证明了所改进模型在真实场景中进行小人脸检测的可行性。

1 相关工作

目前人脸检测算法可以分为传统算法和基于深度学习的算法。

传统算法有Boosting^[9]、AdaBoost^[10]、Float-Boost^[11],以及基于多个Haar-like特征的检测算法^[12-13]等,虽然这类算法检测效率高,但是准确率受到一定程度的限制。

基于深度学习的算法有MTCNN(multi-task cascaded convolutional network)^[14]、单次尺度不变人脸检测器^[15]等。HU等^[16]从尺度不变性、图像分辨率、上下文3个角度思考小人脸检测的问题。针对尺度不变性问题,利用在单个特征层次结构定义的特征,以多任务方式训练和运行特定尺度的检测器。针对图像分辨率问题,对分辨率低的图片进行上采样并与其他同维特征图融合。针对上下文问题, HU等^[16]通过实验证明,有效地利用上下文信息将有助于提升小人脸检测的性能。Deng等^[17]在WIDER FACE数据集上标注了人脸的5个关键点,使用人脸分类、边界框回归、5个关键点以及3D人脸模型预测方法,该方法的准确率提高到91.4%,提升了检测器在困难数据集上的小人脸检测性能。Zhang等^[18]通过改进的差分架构搜索,实现了高效的多尺度特征融合和背景增强。同时基于距离回归和基于边界框分类的多任务损失,设计了新的损失函数来预测准确的边界框,并在WIDER FACE困难数据集上实现了92.1%的检测准确率。ZHU等^[19]使用TinaFace进行人脸检测,将ResNet50^[20]作为骨干网络,在WIDER FACE困难数据集上实现了92.4%的检测准确率。

尽管小人脸检测技术已经取得了很大进展,但由于检测准确率受真实场景中的尺度变化、光照条件等因素影响,上述人脸检测算法仍然存在改进空间。

2 网络结构

本文提出的网络结构分为4个部分,即主干网络、SFEM、卷积核注意力机制与改进的置信度损失函数与激活函数。

2.1 主干网络

YOLO^[21]是针对目标检测任务而提出的一系列

模型, 且YOLOv5在该任务上具有较好的表现。首先, YOLOv5的设计注重实时性能, 相较于该系列的其他模型, 其推理速度更快^[22-23]; 其次, YOLOv5支持多尺度检测, 有助于在不同尺寸的目标上实现更好的性能, 在小目标检测领域^[24-25]取得了比较好的效果; 再次, YOLOv5的设计考虑了不同任务和场景, 可以通过调整参数配置进行灵活定制, 以完成各种真实场景下的检测任务。因此, 笔者对YOLOv5进行了改进, 构建了一个针对小人脸检测任务的模型。

在YOLOv5系列中, YOLOv5s更加轻量化, 可以通过减少网络深度和参数量, 在不影响计算精度的同时, 提高推理速度, 减少训练用时。并且YOLOv5s专注于轻量级设计, 有助于捕捉和保留小目标的特征, 在小尺度目标的检测上表现更好。因此选用YOLOv5s作为主干网络, 改进后的SK-YOLOv5s的结构如图1所示。本模型由输入端、Backbone、Neck、Prediction组成。模型首先通过Focus模块将输入特征图进行切片和卷积, 这种操作有助于模型在较早的层级上同时捕获相对较低和相对较高的分辨率信息, 随后在Backbone模块迭代使用卷积归一化激活模块和残差模块, 从输入图像中提取特征, Neck模块负责融合不同尺度的特征图, 迭代使用卷积、上采样/下采样操作处理不同尺度的信息, 将它们调整为相同尺度, 使用跨阶段连接的方式进行融合, 并做卷积处理, 随后

将融合后的特征输入Prediction模块。Prediction模块接收来自Focus和Neck模块提供的3组多尺度特征图, 使用小人脸增强模块和卷积核注意力机制处理后, 输出预测结果。

2.2 SFEM

在YOLOv5s中, 网络的最后3个卷积层负责将特征图转化为目标检测结果。具体而言, 第一个卷积层输出的是每个锚点预测的类别和位置信息。第二个卷积层用于将较小的特征图上采样至较大的特征图, 以便与较大的特征图进行融合, 提高检测精度。第三个卷积层对不同层次卷积层的特征图进行融合, 得到最终的检测结果。但是这3个卷积层的输入特征图分辨率太低, 分别是 80×80 、 40×40 和 20×20 , 导致对小人脸的检测效果较差。因此, 检测时很可能会漏掉一些小人脸。针对该问题, 所改进模型设计了SFEM, 以提高小人脸特征图的分辨率。如图2所示, 在原始主干网络的第24层后, 对特征图采用最近邻插值的方法进行上采样处理以扩大特征图, 将其与第27层和第30层的特征图进行融合, 提高浅层特征在网络信息处理过程中的权重, 这样就得到了更大的不同尺度的特征图, 以捕捉小人脸特征。

2.3 卷积核注意力机制

所改进模型还融合了SKNet注意力机制, 这是一种轻量化的注意力机制, 旨在解决传统的卷积神经网络在特定任务中卷积核大小固定的问题, 其可

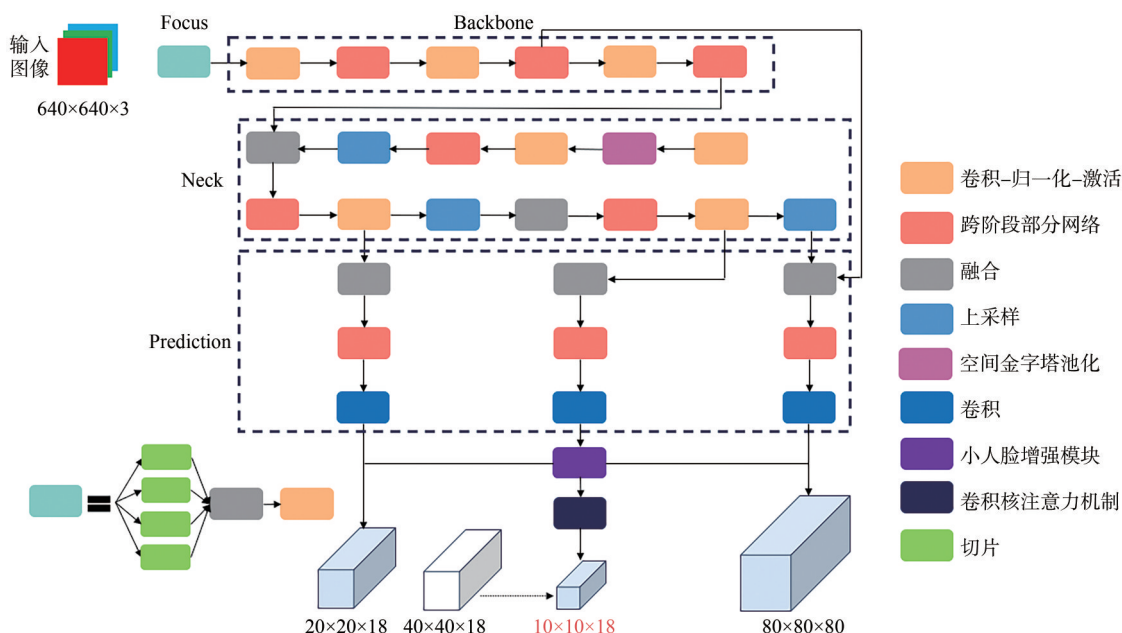


图1 SK-YOLOv5s的结构

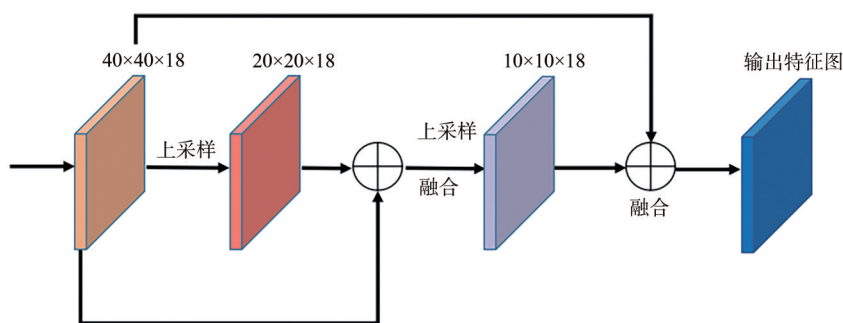


图2 小人脸特征增强模块

以学习关注不同的接受域,能够根据输入的尺度自适应地调节感受野,有助于解决小人脸检测任务中因尺度变化而忽视检测小人脸的问题。SKNet注意力机制的结构如图3所示。

SKNet注意力机制由分裂、融合、选择3个部分组成。假设特征图的大小为 $C \times H \times W$, C 为通道数, H 和 W 分别为特征图的高度和宽度,则SKNet注意力机制会通过以下步骤来处理特征图。首先,将特征图分成不同的组,每个组都会采用不同的卷积核进行卷积操作,例如 1×1 、 3×3 和 5×5 。然后,对于每个分组的卷积结果,SKNet注意力机制通过计算每个卷积核在不同尺度上的重要性权重,决定了如何融合不同尺度的特征,然后根据计算得到的权重,将不同分组的卷积结果进行加权融合。这种加权融合使模型能够更好地适应不同尺度的特征。通过添加SKNet注意力机制,对SFEM中得到的特征图进行处理,从而实现了多尺度特征的有效融

合。这种机制可以帮助所改进模型更好地捕获小人脸的纹理和形状信息,提升了模型的性能和泛化能力。

根据表1的实验结果可以看出,相较于YOLOv5s,引入了SKNet注意力机制之后,人脸检测准确率、召回率、IoU阈值为0.5时的平均精度均值以及IoU阈值为0.5~0.95时的平均精度均值分别提高到了0.893、0.669、0.740和0.399。实验结果表明,引入融合卷积核注意力机制改进的模型可以更有效地解决多尺度人脸检测问题。

2.4 改进的置信度损失函数与激活函数

目标检测任务中一般有3个不同的损失函数,分别用于边界框调整、分类,以及目标定位。

边界框损失函数用于衡量预测与真实边界框位置之间的差异,该损失函数分为中心坐标误差、宽高坐标误差,如式(1)所示。

$$L_{\text{box}} = L_{\text{center}} + L_{\text{wh}} \quad (1)$$

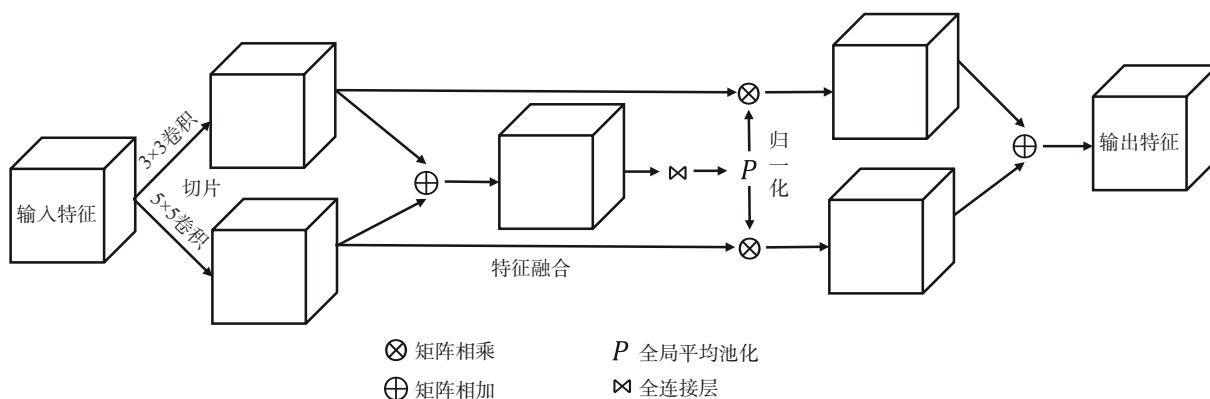


图3 SKNet注意力机制的结构

表1 注意力机制验证实验结果

模型	准确率	召回率	IoU阈值为0.5时的平均精度均值	IoU阈值为0.5~0.95时的平均精度均值
YOLOv5s	0.875	0.666	0.739	0.397
SK-YOLOv5s	0.893	0.669	0.740	0.399

中心坐标误差为:

$$L_{\text{center}} = \frac{1}{N} \sum_{i=1}^N \left((x_i - x_{t_i})^2 + (y_i - y_{t_i})^2 \right) \quad (2)$$

宽高坐标误差为:

$$L_{\text{wh}} = \frac{1}{N} \sum_{i=1}^N \left((w_i - w_{t_i})^2 + (h_i - h_{t_i})^2 \right) \quad (3)$$

其中, N 是所有样本的数量; $(x_{t_i}, y_{t_i}, w_{t_i}, h_{t_i})$ 是真实边界框的中心坐标和宽高坐标; (x_i, y_i, w_i, h_i) 是预测边界框的中心坐标和宽高坐标。预测边界框和真实边界框之间的中心坐标和宽高坐标的误差越小, 说明预测边界框的位置越接近真实边界框的位置。

分类损失函数用于衡量预测与真实边界框类别之间的差异, 只计算正样本(即存在目标的预测框)的分类损失函数, 采用的是二元交叉熵损失函数, 具体如式(4)所示。

$$L_{\text{cls}} = -\frac{1}{N^+} \sum_{i=1}^{N^+} \sum_{j=1}^{\text{Cls}} y_{t_i}^j \log(y_i^j) \quad (4)$$

其中, N^+ 是正样本数量, Cls 是类别数量, y_i^j 是模型预测的第 i 个边界框属于类别 j 的概率, $y_{t_i}^j$ 是实际类别分布中类别 j 的真实概率。

目标损失函数用于衡量预测与真实边界框置信度之间的差异, 置信度是指预测与真实边界框之间的重叠程度, 用完整交并比来衡量, 如式(5)所示。

$$\text{CIoU} = \text{IoU} - \frac{(c - \alpha(v, u))^2}{\alpha} \quad (5)$$

其中, IoU 是预测和真实边界框的交并比。 c 是预测和真实边界框的中心距离, 如式(6)所示。 α 是用于调整完整交并比和中心距离之间权重的一个参数, u 和 v 分别是预测和真实边界框的对角线长度, 如式(7)和式(8)所示。

$$c = \sqrt{(x - x_t)^2 + (y - y_t)^2} \quad (6)$$

$$u = \sqrt{w^2 + h^2} \quad (7)$$

$$v = \sqrt{w_t^2 + h_t^2} \quad (8)$$

使用 CIoU 作为模型训练过程的损失函数, 虽然可以有效地测量预测和真实边界框之间的距离和形状, 并且有效地测量预测和真实边界框之间的重叠面积、中心点和长宽比的差异, 解决距离交并比^[26]在计算中心点距离时忽略了预测和真实边界框之间的形状信息的问题, 但是, CIoU 在计算长宽比时

使用对数函数, 这可能导致梯度消失或爆炸。同时, CIoU 在计算中心点距离时使用欧式距离, 这可能忽略了目标在图像中的位置信息。因此, 针对 CIoU 存在的问题, 所改进模型将使用 EIoU ^[27]。首先, 它在 IoU 的基础上增加了边长的惩罚项, 使预测框更加符合真实边界框的尺度。其次, EIoU 也可以有效地测量预测和真实边界框之间的重叠面积、中心点和边长几何因素的差异。最后, EIoU 可以解决 CIoU 在计算长宽比时使用对数函数导致的梯度消失或爆炸问题。损失函数分为 L_{IoU} 、 L_{dis} 、 L_{asp} 共3个部分, 分别表示 IoU 损失、距离损失和方面损失。如式(9)所示。

$$L_{\text{EIoU}} = L_{\text{IoU}} + L_{\text{dis}} + L_{\text{asp}} \quad (9)$$

具体为:

$$L_{\text{EIoU}} = 1 - \text{IoU} + \frac{\rho^2(b, b')}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w')}{(w^c)^2} + \frac{\rho^2(h, h')}{(h^c)^2} \quad (10)$$

其中, $\rho(\cdot)$ 表示欧几里得距离; b 和 b' 分别表示预测边界框中心点和真实边界框中心点; w^c 和 h^c 是覆盖预测边界框、真实边界框的最小封闭边界框的宽度和高度; w 和 h 是预测边界框的宽和高; w' 和 h' 是真实边界框的宽和高。相比于 CIoU 损失, EIoU 能够更直接地减小两个边界框的宽高差异, 从而提高目标检测的精度和稳定性。

那么总损失函数如式(11)所示。

$$L = L_{\text{cls}} + L_{\text{obj}} + L_{\text{box}} \quad (11)$$

其中, L_{cls} 、 L_{obj} 、 L_{box} 分别表示分类损失函数、目标损失函数和定位损失函数。使用改进后的置信度损失函数, 可以提高 SK-YOLOv5s 的数值稳定性, 加快收敛速度。

最后, 考虑到 YOLOv5s 使用的 SiLU 激活函数计算量大, 需要额外的乘法运算, 同时输出不是以0为中心, 会导致梯度下降的速度变慢等, 使用 FReLU 激活函数。该激活函数可以增强特征图的非线性表达能力, 并且在激活函数阶段实现像素级的空间信息建模, 提高模型的空间感知能力, 可以和常规卷积层无缝集成, 不增加额外的参数和计算量, 易于实现和部署。表2所示为改进的激活函数在小人脸检测任务中的实验结果, 召回率达到了0.669, IoU 阈值为0.5时的平均精度均值达到了0.742, IoU 阈值为0.5~0.95时的平均阈值检测准确率也有一定的提高。实验结果表明, 在改进的模型

表2 改进的激活函数验证实验结果

激活函数	准确率	召回率	IoU 阈值为0.5时的平检精度均值	IoU 阈值为0.5~0.95时的平均精度均值
SiLU	0.875	0.666	0.739	0.399
FReLU	0.873	0.669	0.742	0.403

中使用FReLU激活函数有助于提高模型性能。

3 实验结果分析

3.1 实验设备

实验工作站的CPU为Intel i9-9900K@3.60 GHz×16, GPU为GeForce RTX 2070 SUPER×2, 设备系统为Ubuntu 16.04 LTS。该模型是用PyTorch框架实现的, PyTorch版本为1.7.1, CUDA版本为10.1, 驱动版本为430.64。

3.2 数据集与评价指标

(1) WIDER FACE 数据集

实验使用了WIDER FACE数据集, 在WIDER FACE数据集中挑选了32 203张图像, 并对其中的脸进行了精细标注, 共标注了393 703张人脸。这些人脸具有高度可变性, 包括比例、姿势和遮挡等。WIDER FACE数据集由61个事件类别组成, 这些事件类别包括各种场景和情境, 例如室内、室外、白天、夜晚、人群、个人等。WIDER FACE数据集部分样本如图4所示。

依照COCO数据集对不同尺度检测目标的划分标准, 对WIDER FACE数据集中的不同尺度的人脸进行划分, 见表3。

表3 WIDER FACE数据集中人脸尺度划分标准

分辨率/px	人脸尺度
小于或等于322	小尺度
其他	中等尺度
大于962	大尺度



图4 WIDER FACE数据集部分样本

(2) 评价指标

小人脸检测模型的评价指标为平均精度均值 (mean average precision, mAP)。在计算mAP之前, 需要先计算准确率 (precision, P) 和召回率 (recall, R)。具体如式(12)、式(13)所示。

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

其中, TP (true positive) 表示真正例, 即分类器正确地将正例预测为正例的样本数; FP (false positive) 表示假正例, 即预测为负例但实际为正例的样本数; FN (false negative) 表示假负例, 即预测为正例但实际为负例的样本数。

准确率是一种评估分类模型性能的指标, 它表示预测为正类的样本中有多少是真正的正类。式(12)表示真正例占有所有预测为正例的样本的比例。精确度越高, 说明模型准确选出正例的能力越高, 即越不容易将负类误判为正例。召回率是一种评估分类模型性能的指标, 它表示实际为正例的样本中有多少被预测为正。

$$mAP = \sum_{k=1}^N P(k) \Delta R(k) \quad (14)$$

其中, $\Delta R(k)$ 表示第 k 张图像和第 $k-1$ 张图像召回率之差。

3.3 WIDER FACE 数据集处理

(1) 标签格式说明

WIDER FACE数据集给出的标记人脸的详细信息格式为 $(x_1, y_1, w, h, \text{Blur}, \text{Expression},$

Illumination, Invalid, Occlusion, Pose)。其中 x_1 , y_1 , w , h 分别代表人脸边界框的左上角 x 、 y 坐标, 以及宽、高坐标, 而Blue、Expression、Illumination等标记依次代表人脸的模糊程度、表情、光照条件等信息, 见表4。

表4 WIDER FACE数据集标签格式

名称	含义	0	1	2
Blur	模糊程度	清晰	模糊	极端
Expression	表情	正常	极端	—
Illumination	光照条件	正常	极端	—
Invalid	有效性	有效	无效	—
Occlusion	遮挡程度	无	遮挡	极端
Pose	姿态	正常	非正常	—

(2) 标签格式转换

要将WIDER FACE数据集标签格式转换为YOLO训练格式, 需要进行以下步骤。

首先, 遍历WIDER FACE数据集中的每条小人脸标记信息, 得到每个边界框的左上角和右下角坐标, 记作 (x_1, y_1) 、 (x_2, y_2) ; 然后据此计算边界框的中心坐标 (x, y) 以及宽度和高度坐标 w 、 h , 这样就将WIDER FACE转换成了YOLO格式的标签, 如式(15)所示。

$$\begin{cases} x = \frac{x_1 + x_2}{2} \\ y = \frac{y_1 + y_2}{2} \\ w = x_2 - x_1 \\ h = y_2 - y_1 \end{cases} \quad (15)$$

然后, 为了将WIDER FACE数据集标签格式转换为YOLO格式, 对于每个边界框, 要将它的中心坐标和宽度、高度坐标进行归一化处理。最终将其转换为一个包含5个值的向量: 类别ID、中心坐标 x 、中心坐标 y 、宽度坐标 w 、高度坐标 h 。

3.4 实验设置

在数据集的配置文件中, 将训练集、验证集改为转换后的数据集文件夹, 将类别设置为1, 分类

名称设置为Face。在模型配置文件中, 使用的是版本为6.0的YOLOv5s配置文件, 类别设置为1, 模型深度和宽度分别设置为0.33、0.50。

在训练阶段, 将WIDER FACE数据集的训练集图片分辨率设置为640×640, 每个训练批次大小设置为56张图片, 迭代次数设置为400。输入图像通过Mosaic数据增强进行预处理, 对于锚框数量, 使用K-means算法进行计算, 并且使用初始学习率为0.001、衰减率为0.9的Adam优化器进行参数优化。在测试阶段, 将WIDER FACE数据集中验证集图片分辨率设置为640×640, 不进行任何数据预处理操作。

(1) 损失函数验证实验

将CIoU损失函数的小人脸检测器作为测试基准(Baseline)进行实验对比, 验证EIoU损失函数的有效性。该实验在WIDER FACE数据集上进行, 采用的评价指标是AP值, 实验结果如图5所示。

从实验结果可以看出, EIoU损失函数可以更有效地衡量预测和真实边界框之间的重叠面积、中心点和边长几何因素的差异, 解决CIoU损失函数在计算长宽比时使用对数函数导致的梯度消失或爆炸问题, 使损失函数更加自适应, 同时提高人脸检测的准确性, 尤其是对于小目标和密集目标的人脸检测。

(2) 模块验证实验

将YOLOv5s的小人脸检测器作为测试基准(Baseline)进行实验对比, 验证不同模块的有效性, 实验结果见表5。相比YOLOv5s, 分别加入了SFEM和卷积核注意力机制后, 性能都有不同程度的提升, 且SK-YOLOv5s在小人脸检测的准确率上提升了0.025, 召回率提升了0.071, 在IoU阈值为0.5的平均精度均值提升了0.079, 在IoU阈值为0.5~0.95的平均精度均值提升了0.068。因此证明了所改进模型对小人脸检测的有效性。

3.5 其他检测模型对比结果

为了验证模型的有效性, 在公开数据WIDER

表5 不同改进策略验证实验结果

模型	P	R	IoU阈值为0.5的平均精度均值	IoU阈值为0.5~0.95的平均精度均值
Baseline	0.875	0.666	0.739	0.399
SFEM-YOLOv5s	0.871	0.698	0.770	0.428
SKAttention	0.893	0.669	0.740	0.397
SK-YOLOv5s	0.900	0.737	0.818	0.467

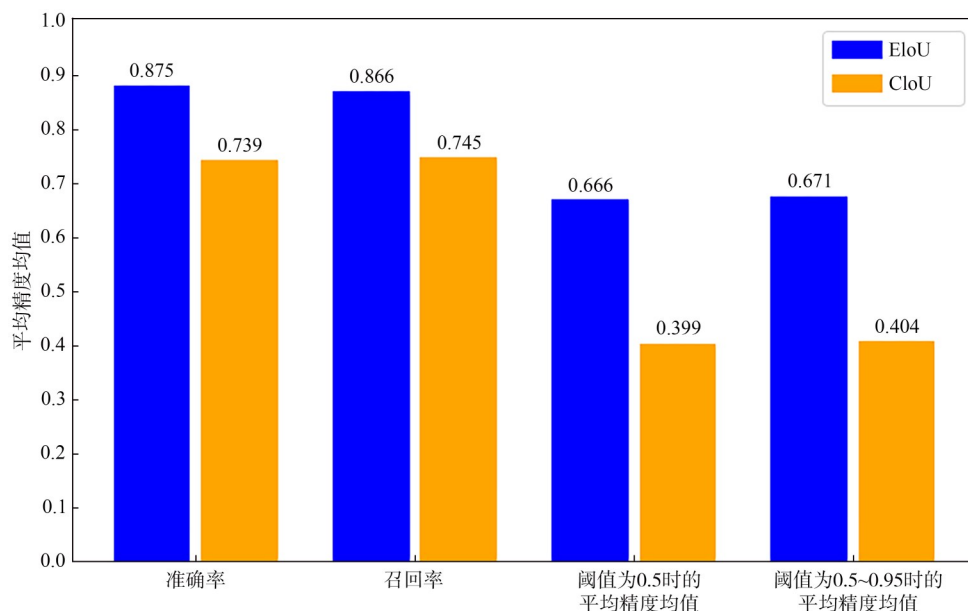


图5 损失函数验证实验结果

FACE上与近年来主流的人脸检测模型的检测结果进行对比。从表6的实验结果可以看出,相比于Baseline,所改进模型在Easy数据集的平均精度均值提升到了0.952、在Medium数据集上提升到了0.928、在Hard数据集中提升到了0.856,但是与较先进的人脸检测模型Retinaface、Tinaface等相比还有一定差距,需要继续对小人脸特征进行优化,提升SK-YOLOv5s的检测性能。

3.6 检测结果分析

(1) 性能测试

从验证集中选择小人脸图像进行检测,然后对比YOLOv5s的检测结果和SK-YOLOv5s的检测结果,如图6所示。

针对小人脸图像,YOLOv5s检测出8张人脸,SK-YOLOv5s检测出9张人脸。分析图6(a)、图6(b),人与人之间在不同姿态、遮挡的问题,在YOLOv5s检测结果中,没有将右后方戴着长筒圣诞帽的人脸检测出来,而SK-YOLOv5s能够将

表6 不同人脸检测模型在各数据集下检测人脸的平均精度均值

网络模型	Easy数据集	Medium数据集	Hard数据集
Baseline	0.894	0.871	0.808
SCRFD-34Gr ^[28]	—	0.949	0.853
MogFace ^[29]	—	0.936	0.874
CenterFace ^[30]	—	0.921	0.873
Img2pose ^[31]	—	0.890	0.839
Poly-NL ^[32]	—	0.957	0.921
EXTD ^[33]	—	0.903	0.850
FDNet ^[34]	0.959	0.935	0.879
FAN ^[35]	0.953	0.942	0.864
Retinaface ^[17]	0.969	0.962	0.919
Tinaface ^[19]	0.963	0.952	0.924
SK-YOLOv5s	0.952	0.928	0.856

其正确的人脸检测出来,可以看出,SK-YOLOv5s对小人脸的正确检测率优于YOLOv5s。

(2) 效率测试

YOLOv5s是轻量化的模型,SK-YOLOv5s在



图6 YOLOv5和SK-YOLOv5s人脸检测结果对比

注重提升性能的基础上,也同时注重图像的检测效率,对YOLOv5s和SK-YOLOv5s进行小人脸检测效率的对比。采用微软发布的一项世界吉尼斯纪录全球自拍照人数最多的一张图进行测试,该图中一共包含约1 151张人脸,SK-YOLOv5s的检测结果见表7。可以看出,所改进模型检测出的小人脸数量增加了21张,且计算速度仍在0.1 s级以内。

表7 不同人脸检测模型的效率对比实验结果

网络模型	检测出的人脸数量/张	用时/s
Baseline	361	0.051
SK-YOLOv5s	382	0.075

4 结束语

本文提出了一种基于卷积核注意力机制的小人脸检测模型SK-YOLOv5s。该模型针对小人脸检测技术存在的问题设计了小人脸特征增强模块,将浅层特征、深层特征进行融合,提高小人脸特征的分辨率,从而增强其特征。同时引入选择性核注意力机制,多尺度自适应调节感受野大小,并且采用EIou作为目标损失函数,采用FReLU作为激活函数,提升特征图的非线性表达能力。实验结果证明,相较于YOLOv5s,所改进模型在真实场景下进行小人脸检测更加可行和有效。

参考文献:

- [1] 郭凡,卢铨宇,李嘉怡,等.基于深度神经网络的目标跟踪算法综述[J].航空兵器,2024,31(1):1-12.
- [2] GUO F, LU X Y, LI J Y, et al. A review of target tracking algorithm based on deep neural network[J]. Aero Weaponry, 2024, 31(1): 1-12.
- [3] PANTIC M, ROTHKRANTZ L J M. Automatic analysis of facial expressions: the state of the art[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1424-1445.
- [4] KUMAR N, BERG A C, BELHUMEUR P N, et al. Attribute and simile classifiers for face verification[C]//Proceedings of the 2009 IEEE 12th International Conference on Computer Vision. Piscataway: IEEE Press, 2009: 365-372.
- [5] 邱天衡,王玲,王鹏,等.基于改进YOLOv5的目标检测算法研究[J].计算机工程与应用,2022,58(13):63-73.
- [6] QIU T H, WANG L, WANG P, et al. Research on object detection algorithm based on improved YOLOv5[J]. Computer Engineering and Applications, 2022, 58(13): 63-73.
- [7] QI D L, TAN W J, YAO Q, et al. YOLO5Face: why reinventing a face detector[M]. Cham: Springer Nature Switzerland, 2023: 228-244.
- [8] 高强,唐福兴,李栋,等.基于改进YOLOv5的密集场景行人检测方法研究[J].国外电子测量技术,2023,42(4):125-130.
- [9] GAO Q, TANG F X, LI D, et al. Research on pedestrian detection method in dense scene based on improved YOLOv5[J]. Foreign Electronic Measurement Technology, 2023, 42(4): 125-130.
- [10] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 510-519.
- [11] YANG S, LUO P, LOY C C, et al. WIDER FACE: a face detection benchmark[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 5525-5533.
- [12] SCHAPIRE R E. The strength of weak learnability[J]. Machine Learning, 1990, 5(2): 197-227.
- [13] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of online learning and an application to boosting[M]. Heidelberg: Springer, 1995: 23-37.
- [14] LIENHART R, MAYDT J. An extended set of Haar-like features for rapid object detection[C]//Proceedings of the International Conference on Image Processing. Piscataway: IEEE Press, 2002.
- [15] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2001.
- [16] MITA T, KANEKO T, HORI O. Joint haar-like features for face detection[C]//Proceedings of the Tenth IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2005: 1619-1626.
- [17] ZHANG K P, ZHANG Z P, LI Z F, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503.
- [18] ZHANG S F, ZHU X Y, LEI Z, et al. S3FD: single shot scale-invariant face detector[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 192-201.
- [19] HU P Y, RAMANAN D. Finding tiny faces[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 1522-1530.
- [20] DENG J K, GUO J, ZHOU Y X, et al. Retinaface: single-stage dense face localisation in the wild[EB]. arXiv preprint, 2019, arXiv: 1905.00641.
- [21] ZHANG B, LI J, WANG Y B, et al. ASF: automatic and scalable face detector[EB]. arXiv preprint, 2003, arXiv: 2003.11228.
- [22] ZHU Y J, CAI H X, ZHANG S H, et al. Tinaface: strong but simple baseline for face detection[EB]. arXiv preprint, 2020, arXiv: 2011.13183.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [24] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 6517-6525.
- [25] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: learning what you want to learn using programmable gradient information[M]. Cham: Springer Nature Switzerland, 2024: 1-21.
- [26] 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]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 7464-7475.
- [27] 田庆,胡蓉,李佐勇,等.基于SE-YOLOv5s的绝缘子检测[J].智能科学与技术学报,2021,3(3):312-321.
- [28] TIAN Q, HU R, LI Z Y, et al. Insulator detection based on SE-YOLOv5s[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(3): 312-321.

- [25] 谢椿辉, 吴金明, 徐怀宇. 改进YOLOv5的无人机影像小目标检测算法[J]. 计算机工程与应用, 2023, 59(9): 198-206.
- XIE C H, WU J M, XU H Y. Small object detection algorithm based on improved YOLOv5 in UAV image[J]. Computer Engineering and Applications, 2023, 59(9): 198-206.
- [26] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[EB]. arXiv preprint, 2019, arXiv: 1911.08287.
- [27] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [28] GUO J, DENG J K, LATTAS A, et al. Sample and computation redistribution for efficient face detection[EB]. arXiv preprint, 2021, arXiv: 2105.04714.
- [29] LIU Y, WANG F, DENG J K, et al. MogFace: towards a deeper appreciation on face detection[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 4083-4092.
- [30] XU Y Y, YAN W, YANG G K, et al. CenterFace: joint face detection and alignment using face as point[J]. Scientific Programming, 2020, 2020: 7845384.
- [31] ALBIERO V, CHEN X Y, YIN X, et al. img2pose: face alignment and detection via 6DoF, face pose estimation[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2021: 7613-7623.
- [32] BABILONI F, MARRAS I, KOKKINOS F, et al. Poly-NL: linear complexity non-local layers with polynomials[EB]. arXiv preprint, 2021, arXiv: 2107.02859.
- [33] YOO Y, HAN D, YUN S. EXTD: extremely tiny face detector via iterative filter reuse[EB]. arXiv preprint, 2019, arXiv: 1906.06579.
- [34] SUN X D, WU P C, HOI S C H. Face detection using deep learning: an improved faster RCNN approach[J]. Neurocomputing, 2018, 299:

42-50.

- [35] WANG J F, YUAN Y, YU G. Face attention network: an effective face detector for the occluded faces[EB]. arXiv preprint, 2017, arXiv: 1711.07246.

[作者简介]



周丽芳 (1974-), 女, 博士, 重庆邮电大学软件工程学院教授, 主要研究方向为图像特征提取与识别、目标检测与跟踪。



胡振 (1999-), 男, 重庆邮电大学软件工程学院硕士生, 主要研究方向为深度学习、计算机视觉。



刘波 (1998-), 男, 重庆邮电大学软件工程学院硕士生, 主要研究方向为深度学习、计算机视觉。