

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, 小人脸检测面临着诸多挑战, 如尺度、光照剧烈变化、计算量大和精度较低等问题。针对现有模型中小人脸容易被忽视的问题, 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。首先, 设计一种小人脸增强模块 SFEM, 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, 直接优化预测框与真实框的宽高差异, 提高小人脸检测的精度和稳定性, 相较于原有方法, 所提方法在 WIDER Face 数据集中平均检测准确率提升了 3.1%。实验结果证明了该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.xxxxxx

Small Face Detection Based on Improved YOLOv5s Network

Abstract: In the complex real-world applications, the task of detecting small faces encounters numerous challenges, which include issues related to scale variation, abrupt lighting changes and suboptimal precision. In order to address the concerns of overlooking small facial structures within existing models, this study introduces a novel approach termed SK-YOLOv5s which is based on convolutional kernel attention mechanism. Firstly, we proposal a small face enhancement module to fuse multi-layer features, which can enhance the resolution of small faces and improve its features. Subsequently, we incorporates SKNet attention mechanism into the model to adaptively adjust receptive field sizes across multiple scales so that the detection efficacy of small face can be enhanced. Finally, EIoU is utilized as the loss function, which facilitate a direct optimization of the width and height discrepancies between predicted and actual bounding boxes. This approach enhances the precision and stability of small face detection. The performance on the WIDER Face dataset demonstrate a mean detection accuracy improvement of 3.1% over prevailing methods. The experimental results demonstrate the model's viability for detecting small faces in real-life scenarios.

Key words: Small Face Detection, Attention Mechanism, YOLOv5s

1 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[2]、表情分析[3]和属性识别[4]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和识别小人脸, 及时发现可疑人员或异常情况, 提高监控效率和安全性。目前基于改进 YOLOv5 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[1]设计了更

收稿日期: 2024-03-27; 修回日期: 2024-09-27

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, No.62221005); 重庆市教委科技攻坚计划项目 (No.KJZD-K201900601, No.KJZDK20220606)

Foundation Items: National Natural Science Foundatio of China (No.62276039, No.62221005), Science and Technologh Research Program of Chongqing Municipal Education Commission(No.KJZD-K201900601, No.KJZDK20220606)

有效的网络架构,并且优化了损失函数和使用了新的注意力机制;Qi等[30]在保持高准确率的同时,还能够实现实时的人脸检测,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[35]将小人脸检测优化技术集成到行人检测任务中,该研究也使用了SKNet[28]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。

虽然小人脸检测技术近10年来有了很大突破,但是该任务仍然面临一些问题和挑战,首先是尺度问题:小人脸通常在图像中具有较小尺度,与常规人脸检测相比,小人脸分辨率较低,特征更加微弱,难以被准确的检测和识别;其次是遮挡问题:在一些人群密集的场景中,小人脸往往会被其他人物或物体遮挡,使得部分小人脸无法被完整地观察到,这进一步影响检测和识别地准确性;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,尽管小人脸检测技术在各个领域具有重要的应用价值,但仍然面临着诸多挑战和问题。

针对小人脸容易漏检问题,本文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,该模型首先通过设计一种小人脸增强模块(Small Face Enhancement Module, SFEM),提高小人脸特征图的分辨率,从而增强其特征;其次,引入卷积核注意力机制(Convolution Kernel Attention Mechanism, CKAM),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。

与原始YOLOv5模型相比,本文改进的SK-YOLOv5s模型在WIDER Face数据集中IoU阈值为0.5时,平均检测准确率的平均值提升了7.9%,通过实验证明了该模型在真实场景下进行小人脸检测的可行性。

2 相关工作

目前人脸检测算法可以分为传统方法和基于深度学习的方法。

传统方法有Boosting算法[5]、AdaBoost算法

[6]、FloatBoost算法[7]、以及PA Viola[9]、TAKESHI[8]先后使用的基于多个Haar-like特征的检测算法等,虽然这类方法效率高,但是准确率受到了一定程度的影响。

基于深度学习的方法有MTCNN[27]、S3FD[14]等,其中,HU等[15]从尺度不变性、图像分辨率、上下文三个角度思考小人脸检测的问题:针对尺度问题,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;针对分辨率问题,将分辨率低的图片上采样进行检测将结果进行融合;针对上下文问题,该研究通过实验证明,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[16]在人脸检测数据集WIDER FACE[18]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,取得了91.4%的效果,提升了检测器在困难数据集上的小人脸检测性能。ZHANG等[17]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。同时基于距离的回归和基于边框的分类的多任务损失,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处理人脸检测,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。

尽管小人脸检测已经取得了很大进展,但由于真实场景中的尺度变化、光照条件等因素带来的影响,上述方法仍然存在改进空间。

3 网络结构

本文提出的网络结构分为三个部分,即主干网络,小人脸增强网络和卷积核注意力网络。

3.1 主干网络

YOLO[10]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,相较于该系列的其它模型[12-13],其推理速度更快;其次,YOLOv5支持多尺度检测,这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[23-24]取得了比较好的效果;最后,该模型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,本章对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。因此我们选用 YOLOv5s 作为主干网络, 我们改进的网络如图 1 所示, 本模型由输入端、Backbone、Neck、Prediction 四个部分组成。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到相对较低和相对较高的分辨率信息, 随后在 Backbone 部分迭代使用 CBL (Conv Batch Normalization Leaky ReLU, 卷积归一化激活模块) 和 CSP (Cross Stage Partial, 残差模块) 从输入图像中提取特征, Neck 部分通过将不同尺度的特征图进行融合, 以及应用多尺度处理来实现, 最后通过检测头生成检测结果。

3.2 小人脸特征增强模块

在原始 YOLOv5 中, 网络的最后三个卷积层负责将特征图转化为目标检测结果。具体而言, 第一个输出的是每个锚点预测的类别和位置信息, 第二个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。第三个对不同层次的特征图进行融合, 得到最终的检测结果。但是这三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 , 导致对小目标的检测效果较差。因此, 检测结果很可能会漏掉一些小人脸。针对该问题, 本文改进的网络模型中设计了小人脸特征增强模块以增加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, 提高了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用以捕捉小人脸特征。

3.3 融合卷积核注意力机制

我们改进的网络中还使用了 SKNet, 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, 网络可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, 这有助于解决小人脸检测中因尺度变化而漏检的问题。SKNet 结构模型如图 3 所示。

SKNet 由分裂、融合、选择三个组件组成。假设特征图的大小为 $C \times H \times W$, C 为通道数, H 和

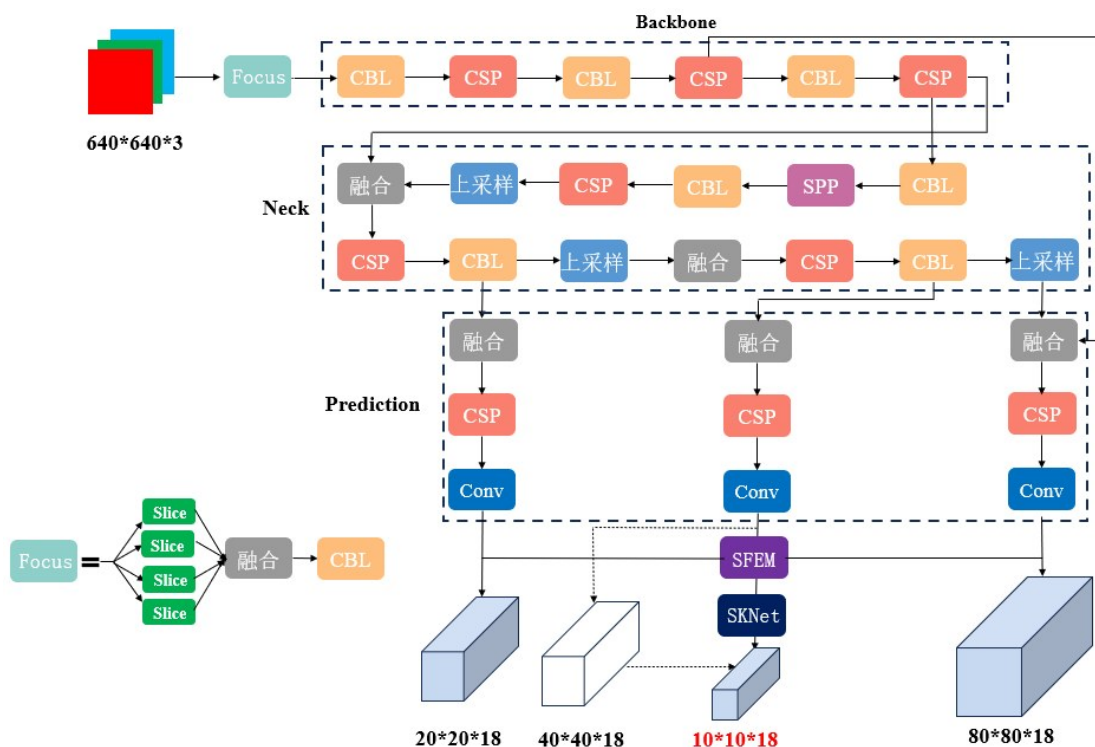


图 1 SE-YOLOv5s 网络模型结构

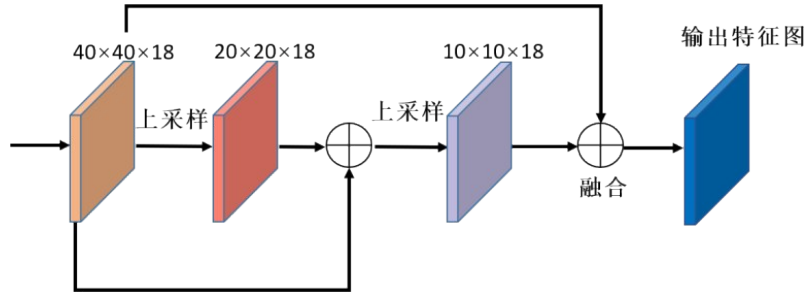


图2 小人脸特征增强模块

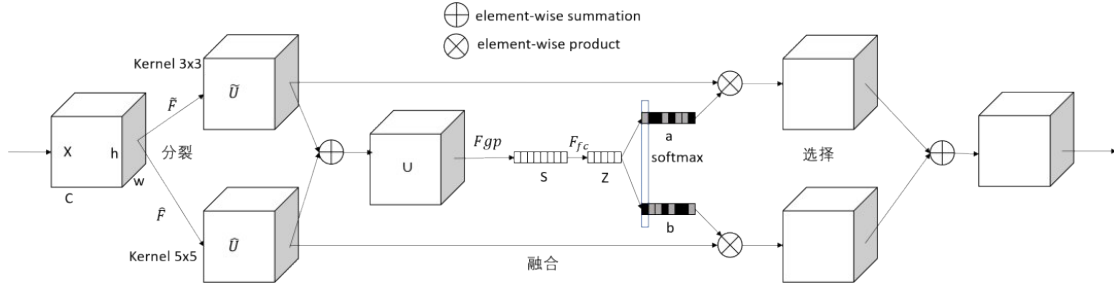


图3 选择性核卷积

W 为特征图的高度和宽度，则 SKNet 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 会应用 Selective Kernel 机制。该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，比如将特征图逐元素相加、在通道维度拼接等融合手段，而这些权重表示了不同卷积核对于捕获不同尺度特征的贡献程度；最后，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。这种加权融合使得模型能够更好地适应不同尺度的特征。通过添加 SKNet，我们对 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。这种机制可以帮助网络更好地捕获不同尺度的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 的实验结果数据，可以看出：相较于原始网络，引入了 SKNet 之后，人脸检测准确率、召回率、IoU 阈值为 0.5 的平均检测准确率的平均值分别提高了 1.8%、0.3%、0.1%。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的应对多尺度问题。

3.4 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三个不同的损失函数，分别用于边界框调整、分类、以及目标定位。

表 1 注意力机制验证实验结果

	准确率	召回率	mAP.5	mAP.5:.95
Baseline	0.875	0.666	0.739	0.399
SKAttention	0.893	0.669	0.74	0.397

边界框损失是用于衡量预测框的位置与真实框的位置之间的差异，该损失分为中心坐标误差、宽高坐标误差，如公式 1 所示：

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

中心坐标误差如公式 2 所示：

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

宽高坐标误差如公式 3 所示：

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

其中， N 是正样本的数量； $(x_{t_i}, y_{t_i}, w_{t_i}, h_{t_i})$ 是真实框的中心坐标和宽高； (x_i, y_i, w_i, h_i) 是预测框中心坐标、宽高。中心坐标和宽高坐标的误差越小，说明预测框的位置越接近真实框的位置。

分类损失用于衡量预测框的类别与真实框的类别之间的差异，只计算正样本（即存在目标的预测框）的分类损失，采用的是二元交叉熵损失，具体如公式 4 所示。

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

其中 N 是样本数量， C 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。

目标损失用于衡量预测框的置信度与真实框的置信度之间的差异, 置信度是指预测框与真实框之间的重叠程度, 用 CIoU (Complete Intersection over Union, 完整交并比) 来衡量。如公式 5 所示:

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

其中, IoU (Intersection over Union, 交并比) 是两个框的交并比, c 是两个框的中心距离, 如公式 6 所示:

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

而 α 是用于调整完整交并比和中心距离之间的权重, u 和 v 是边界框 w 和 s 的对角线长度, 计算方式如公式 7 和公式 8 所示:

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

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

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框和真实框之间的距离和形状, 并且有效地测量预测框和真实框之间的重叠面积、中心点和长宽比的差异, 解决 DIoU [25] (Distance Intersection over Union, 距离交并比) 在计算中心点距离时忽略了预测框和真实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的问题。同时, CIoU 在计算中心点距离时使用欧式距离, 这可能会忽略了目标在图像中的位置信息。因此, 针对 CIoU 存在的问题, 本文改进的模型将使用 EIoU [26] (Embedding Intersection over Union, 嵌入交并比)。首先, 它在 IoU 的基础上增加了边长的惩罚项, 使得预测框更加符合真实框的尺度。其次, EIoU 也可以有效地测量预测框和真实框之间的重叠面积、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的问题。损失函数分为三个部分: L_{IoU} 、 L_{dis} 、 L_{asp} 分别表示 IoU 损失、距离损失和方面损失。如公式 9 所示:

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

具体如公式 10 所示:

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

其中 $\rho(\cdot)$ 表示欧几里得距离; w^c 和 h^c 是覆盖预测框、真实框的最小封闭框的宽度和高度; w 和 h 是预测框的宽和高; w^{gt} 和 h^{gt} 是真实框的宽和高。相比于 CIoU 损失, EIoU 能够更直接地优化两个框的宽高差异, 从而提高目标检测的精度和稳定性。

那么总损失函数如公式 11 所示:

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

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

最后, 考虑到原 YOLOv5s 使用的 SiLU 激活函数存在计算量大, 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问题。我们使用 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。如表 2 所示, 改进的激活函数在小人脸检测任务中, 在召回率、阈值为 0.5 的平均检测准确率、阈值为 0.5-0.95 的平均阈值检测准确率分别提高了 0.3%、0.3%、0.4%。实验结果表明, 我们在改进的模型中使用 FReLU 激活函数有助于提高模型性能。

表 2 激活函数验证实验结果

	准确率	召回率	mAP.5	mAP.5:.95
SiLU	0.875	0.666	0.739	0.399
FReLU	0.873	0.669	0.742	0.403

4 实验结果分析

4.1 实验设备

本文实验工作站的 CPU 为 Intel i9-9900K @ 3.60GHz×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。

4.2 数据集与评价指标

4.2.1 WIDER Face 数据集

实验使用了 WIDER Face 数据集数据集，在 WIDER Face 数据集中挑选了 32203 张图像，并对其中的人脸进行了精细标注，共标注了 393703 个人脸。这些人脸具有高度可变性，包括比例、姿势和遮挡等方面。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 数据集对不同尺度目标的划分标准，对 WIDER Face 中的不同尺度的人脸进行划分，如表 3 所示。

表 3 WIDER Face 数据集中人脸尺度划分标准

区域	人脸尺度
area<322	small
322<area<962	middle
962<area	large

4.2.2 评价指标

小人脸检测模型的评价指标为 mAP (mean Average Precision, 平均准确率)。在计算 mAP 之前，需要先计算准确率 P (Precision, 准确率) 和召回率 R (Recall, 召回率)。具体公式如 13、14 所示：

$$P = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

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。公式中的 TP 表示真正例 (True Positive)，即分类器正确地将正例预测为正例的样本数；FP 表示假正例

(False Positive)，即预测为负例但实际为正例的样本数。公式的意义是，准确率等于真正例的比例，即真正例占有所有预测为正例的样本的比例。精确度越高，说明模型预测的正类越准确，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例；FN 表示假负例。AP (Average Precision, 平均准确率) 表示人脸类别预测正确的概率平均值，如公式 15 所示：

$$AP = \frac{\sum_{k=1}^m P}{M} \quad (15)$$

其中 M 表示图片的总张数；P 表示预测正确的概率， $\sum$ 表示概率和。mAP 表示对所有类别的 AP 进行平均化，具体如公式 16 所示：

$$mAP = \frac{\sum_{l=1}^F AP}{F} \quad (16)$$

其中，F 表示类别数；AP 表示人脸类别预测正确的概率平均值。

在 WIDER Face 数据集中，只有人脸类别，类别为 1，mAP 就是 AP。AP 越高代表模型的性能越好。对 mAP.5、mAP.5:.95 进行定义：mAP.5 表示人脸类别在 IoU 阈值为 0.5 时的 AP 值，mAP.5:.95 表示人脸类别在 IoU 阈值为 0.5-0.95 时的 AP 值。

4.3 WIDER Face 数据集处理

4.3.1 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 (x1, y1, w, h, blur, expression, illumination, invalid, occlusion, pose)。其中 x1, y1, w, h 分别代表人脸边界框的左上角 x、y 坐标，以及宽、高，使用绝对坐标，而 blur、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，如表 4 所示：



图 4 WIDER Face 数据集部分样本图

表 4 WIDER Face 数据集标签格式

	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	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	-

4.3.2 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，将左上角和右下角坐标转换为边界框的中心坐标和宽度和高度。对于 Wider Face 数据集中的每个边界框，将左上角和右下角坐标分别计算出 x_1 、 y_1 、 x_2 、 y_2 。计算边界框的中心坐标 x 、 y 以及宽度 w 和高度 h ，将边界框的中心坐标和宽度/高度转换为 YOLO 格式的标签。如公式 17 所示：

$$\begin{cases} x = \frac{x_1 + x_2}{2} \\ y = \frac{y_1 + y_2}{2} \\ w = x_2 - x_1 \\ h = y_2 - y_1 \end{cases} \quad (17)$$

然后，为了转换为 YOLO 格式，对于每个边界框，要将它的中心坐标和宽度、高度进行归一化处理，将坐标值归一化到 $[0,1]$ 之间。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 w 、高度 h 。

4.4 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，我们使用的是版本为 6.0 的 YOLOv5 的 YOLOv5s 配置文件，类别设置为 1，模型深度和宽度设置分别为 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 Kmeans 算法进行计算，并且使用初始学习率 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 ，不使用任何数据预处理操作。

4.4.1 损失函数验证实验

将损失函数 CIoU 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 可以有效地衡量预测框和真实框之间的重叠面积、中心点和边长几何因素的差异，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的问题，使得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。

4.4.2 模块验证实验

将 YOLOv5s 的小人脸检测器作为测试基准进行实验对比，验证不同模块的有效性，实验结果如表 5 所示。相比原模型，分别加入了 SFEM 和

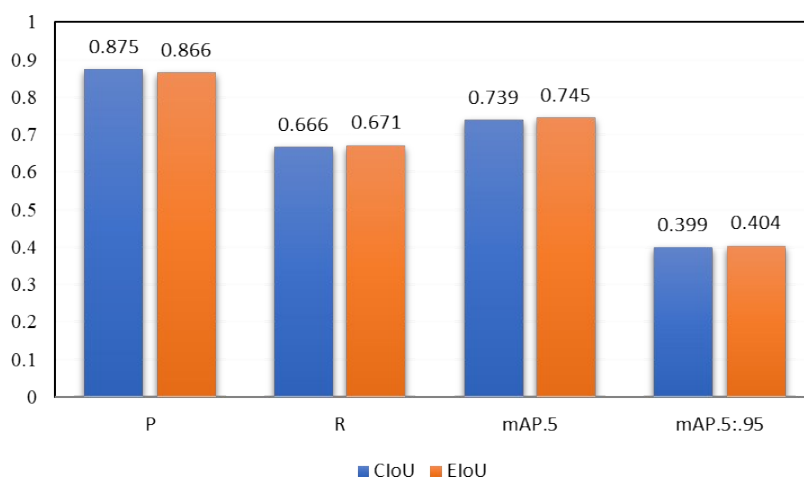


图 5 损失函数验证实验结果

SKAttention 模块后, 性能都有不同程度的提升, 并且同时使用两个模块的 SK-YOLOv5s 模型在小人脸检测的准确率上提升了 2.5%, 召回率提升了 7.1%, 在 IoU 阈值为 0.5 的平均检测准确率提升了 7.9%, 在 IoU 阈值为 0.5-0.95 的平均检测准确率提升了 6.8%。因此证明了本文所提模型对小人脸检测的有效性。

表 5 不同改进策略验证实验结果

	P	R	mAP.5	mAP.5:.95
Baseline	0.875	0.666	0.739	0.399
SFEM-YOLOv5s	0.871	0.698	0.77	0.428
SKAttention	0.893	0.669	0.74	0.397
SK-YOLOv5s	0.9	0.737	0.818	0.467

4.5 其他检测模型对比结果

为了验证模型的有效性, 在公开数据集 Wider Face 上与近年来主流的人脸检测模型的结果进行对比。从表 5 实验结果可以看出, 所设计模型在 Easy 数据集的平均准确率相较于基准网络提升了 6.1%, 在 Medium 数据集比基准网络提升了 5.7%, 在 Hard 数据集中比基准网络提升了 4.8%。本章所设计的 SK-YOLOv5s 模型与早期的人脸检测模型 M-Cascade、CMS-RCNN、SSH 相比在 Easy、Medium、Hard 数据集均有不同程度的提升, 但是与最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。

4.6 检测结果分析

4.6.1 性能测试

从验证集中选择小人脸图片进行检测, 然后对比 YOLOv5s 模型的检测结果和本章设计的 SK-YOLOv5s 检测模型的结果, 如图 6 所示。

针对小人脸图片, YOLOv5s 检测出 8 个人脸,

表 6 不同人脸检测模型的实验结果

网络模型	Easy(map%)	Medium (map%)	Hard(map%)
Baseline	89.4	87.1	80.8
SCRFD-34GF[31]	-	94.9	85.3
MogFace[32]	-	93.6	87.4
CenterFace[33]	-	92.1	87.3
Img2pose[29]	-	89.0	83.9
Poly-NL[11]	-	95.7	92.1
EXTD[34]	-	90.3	85.0
FDNet [22]	95.9	93.5	87.9
FAN[21]	95.3	94.2	86.4
Retinaface[16]	96.9	96.2	91.9
Tinaface[19]	96.3	95.2	92.4
SK-YOLOv5s	95.2	92.8	85.6

SK-YOLOv5 检测出 9 个人脸。分析图 6(a)、图 6(b), 人与人之间存在不同姿态、遮挡的问题, 在 YOLOv5s 检测结果中, 没有将右后方带着红色长筒圣诞帽的人脸检测出来, 而 SK-YOLOv5s 将其正确的检测出来。在小人脸图像检测中, SK-YOLOv5s 的小人脸正确检测率优于原始模型, 并且没有将其他物品检测为人脸。

4.6.2 效率测试

YOLOv5s 是轻量化的模型, 我们改进的 SK-YOLOv5s 在注重提升性能的基础上, 也同样注重图像的检测效率, 我们对原始的 YOLOv5 和改进的 SK-YOLOv5s 进行小人脸检测效率的对比, 检测结果的对比展示采用微软发布一项世界吉尼斯记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 1151 张人脸, 如图 7 所示为 SK-YOLOv5s 的检测效果图, 该图片的检测效果如表 7 所示, 可以看出, 我们改进的模型检测出的小人脸个数增加了 5.8%, 但是计算速度仍然限制在 0.01s 级以内。

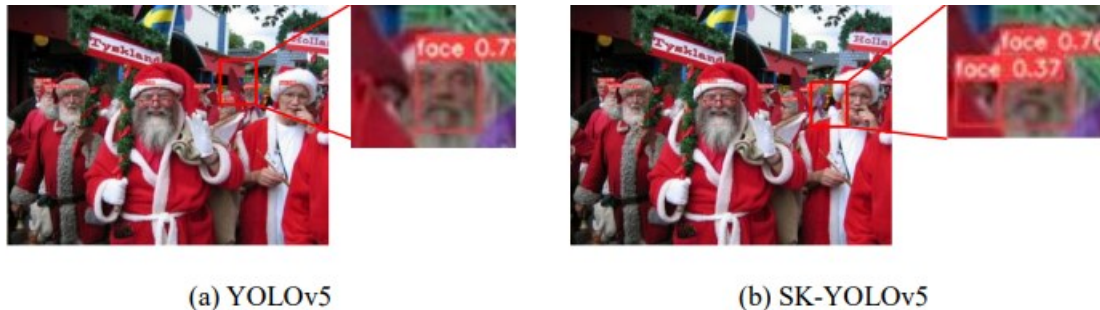


图 6 YOLOv5 人脸检测结果

表 7 不同人脸检测模型的效率对比实验结果

网络模型	检测出的人脸个数	用时(s)
Baseline	361	0.051s
SK-YOLOv5s	382	0.075s

5 结束语

本文提出了一种基于卷积核注意力机制而改进的模型 SK-YOLOv5s 用于小人脸检测，针对小人脸检测的技术难点设计了一种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 作为目标损失函数以及采用 FReLU 作为激活函数，增强特征图的非线性表达能力，实验结果证明，该改进后的模型相较于原模型在真实场景下进行小人脸检测更加可行和有效。



图 7 SK-YOLOv5s 检测结果

参考文献:

- [1] 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):11. DOI: 10.3778/j.issn.1002-8331.2202-0093.
- [2] 郭凡,卢铎宇,李嘉怡,王红梅.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). DOI: 10.12132/ISSN.1673-5048.2022.0226.
- [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]. IEEE International Conference on Computer Vision, Kyoto, Japan, 2009: 365-372.
- [5] SCHAPIRE R E. The strength of weak learnability[J]. Machine Learning, 1990, 5(2): 197-227.
- [6] SCHAPIRE R, FREUND Y. A decision-theoretic generalization of on-line learning and an application to boosting[C]. Second European Conference on Computational Learning Theory, Barcelona, Spain, 1995: 23-37.
- [7] LIENHART R, MAYDT J. An extended set of haar-like features for rapid object detection[C]. IEEE Conference on Image Processing, Rochester, USA, 2002, 1: I-I.
- [8] MITA T, KANEKO T, HORI O. Joint haar-like features for face detection[C]. IEEE Conference on Computer Vision, Beijing, China, 2005, 2: 1619-1626.
- [9] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, USA, 2001, 1: I-I.
- [10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 7263-7271.
- [11] Babiloni F, Marras I, Kokkinos F, et al. Poly-NL: Linear Complexity Non-local Layers with Polynomials[J]. 2021. DOI: 10.1109/ICCV48922.2021.01035.
- [12] Wang, Chien-Yao, I-Hau Yeh, and Hong-Yuan Mark Liao. "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information." arxiv preprint arxiv:2402.13616 (2024).
- [13] Wang, Chien Yao, A. Bochkovskiy, and H. Y. M. Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." arXiv arXiv, 2022.
- [14] ZHANG S, ZHU X, LEI Z, et al. S3fd: single shot scale-invariant face detector[C]. IEEE Conference on Computer Vision, Venice, Italy, 2017: 192-201.
- [15] HU P, RAMANAN D. Finding tiny faces[C]. IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 951-959.
- [16] Deng, Jiankang, et al. "Retinaface: Single-stage dense face localisation in the wild." arxiv preprint arxiv:1905.00641 (2019).
- [17] ZHANG B, LI J, WANG Y, et al. Asfd: automatic and scalable face detector[EB/OL]. (2020-3-30) [2023-3-15]. <https://arxiv.org/abs/2003.11228>.
- [18] YANG S, LUO P, LOY C C, et al. Wider face: a face detection benchmark[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 5525-5533.
- [19] Zhu, Yanjia, et al. "Tinaface: Strong but simple baseline for face detection." arxiv preprint arxiv:2011.13183 (2020).
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778.
- [21] Wang J, Yuan Y, Yu G. Face Attention Network: An Effective Face Detector for the Occluded Faces[J]. 2017. DOI: 10.48550/arXiv.1711.07246.
- [22] SUN X, WU P, HOI S C H. Face detection using deep learning: an improved faster rcnn approach[J]. Neurocomputing, 2018, 299: 42-50.
- [23] 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.
- [24] 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):9. DOI: 10.3778/j.issn.1002-8331.2212-0336.
- [25] ZHENG Z, WANG P, LIU W, et al. Distance-iou loss: faster and better learning for bounding box regression[C]. AAAI Conference on Artificial Intelligence, New York, USA, 2020, 34(07): 12993-13000.
- [26] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient iou loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [27] Zhang K, Zhang Z, Li Z, et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing Letters, 2016, 23(10): 1499-1503. DOI: 10.1109/LSP.2016.2603342.
- [28] LI X, WANG W, HU X, et al. Selective kernel networks[C]. IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 510-519.
- [29] Albiero V, Chen X, Yin X, et al. img2pose: Face alignment and detection via 6dof, face pose estimation[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 7617-7627.
- [30] Qi D, Tan W, Yao Q, et al. YOLO5Face: Why Reinventing a Face Detector[J]. 2021. DOI: 10.48550/arXiv.2105.12931.
- [31] Guo J, Deng J, Lattas A, et al. Sample and Computation Redistribution for Efficient Face Detection[J]. 2021. DOI: 10.48550/arXiv.2105.04714.
- [32] Liu Y, Wang F, Deng J, et al. Mogface: Towards a deeper appreciation on face detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 4093-4102.
- [33] Xu Y, Yan W, Yang G, et al. CenterFace: Joint Face Detection and Alignment Using Face as Point[J]. Scientific Programming, 2020, 2020: 1-8. DOI: 10.1155/2020/7845384.
- [34] Yoo Y J, Han D, Yun S. EXT-D: Extremely Tiny Face Detector via Iterative Filter Reuse[J]. 2019. DOI: 10.48550/arXiv.1906.06579.
- [35] 高强,唐福兴,李栋,et al. 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.