

基于 YOLOv5 的 CCTSDB 小目标检测算法研究

刘顺财¹
LIU Shunca

摘要

为了提高全新中国交通标志检测数据集 2021 (CCTSDB 2021) 的小目标检测精度。在 YOLOv5 网络模型上, 融入归一化的注意力模块 (NAM) 和协调注意力模块 (CA), 同时新增加 160×160 的检测特征图, 增加小目标检测层, 用于检测 4×4 以上的目标。在 YOLOv5 中采用改进的 SIou 目标回归损失函数, 使得整个网络模型对图像特征的学习能力和目标检测精度得到一定的提升。实验表明, CCTSDB 2021 在改进的 YOLOv5 算法中, 小目标检测精度 $mAP@.5$ 和 $mAP@.5:.95$ 达到 85.87%、57.21%, 相比原 YOLOv5 网络 $mAP@.5$ 、 $mAP@.5:.95$ 分别提升了 5.72%、5.85%, 检测精度和精确率 - 召回率得到了明显提升, 减少了推理时间, 提高了整体网络的鲁棒性, 具有更好的检测性能和目标分辨能力。

关键词

交通标志; 小目标; 注意力; 检测算法; 精度

doi: 10.3969/j.issn.1672-9528.2024.03.010

0 引言

在自动驾驶过程中, 小目标检测效果不理想的原因之一是小目标样本的尺寸较小。YOLOv5 的采样倍数较大, 导致较深的特征图很难学习到小目标的特征信息。为了解决这个问题, 提出了一种增加小目标检测层的方法, 该方法将较浅特征图与深特征图拼接后再进行检测。通过这种方式, 可以更好地利用不同层次特征图的信息, 使模型能够更有效地学习和检测小目标。

1 网络模型

在 Facebook 开源神经网络框架 PyTorch 的支持下, ultralytics 团队发布了 YOLOv5^[1-2]。尽管 YOLOv5 在性能上稍逊于 YOLOv4, 但它在灵活性和速度上表现出色, 远胜于 YOLOv4。此外, 由于 YOLOv5 可以使用计算机的 GPU 来加载数据和进行训练。在自动驾驶的过程中, 检测某些复杂场景中的小目标, 存在较大的误差。原因是输入图片的尺寸较大, 小目标尺寸较小, 训练时图片中的小目标感知不到, 导致训练的网络模型检测精度不高。

1.1 改进的 YOLOv5 网络结构图

为了提高检测性能, 可以改进 YOLOv5 的网络结构, 系统采取以下措施。首先, 在网络中引入具有通道 + 空间的注

意力模块, 可以有效提高模型对通道间特征图信息的自适应学习能力, 对特征进行重新拟合和分配, 从而增加小目标特征图的占比权重。这样, 模型可以更好地训练小目标的特征, 提高其检测精度。

其次, 增加一个新的检测层并改进模型的融合方式。通过多尺度特征融合方法, 可以有效地检测不同尺度的目标。具体来说, 修改 Anchor 设置, 增加一组较小的 Anchor, 以更好地适应小目标的检测。同时, 对 head 部分进行修改, 在第 17 层后面进行上采样操作, 处理特征图并扩大其尺寸。

最后, 在第 20 层, 将骨干网络中第 5 层的特征图与获取到的 160×160 大小的特征图进行 concat 融合, 以获得更大的特征图。这样做的目的是提高模型对小目标的检测能力。改进的网络模型如图 1 所示。

1.2 改进 YOLOv5 下的注意力模块

为了提高目标检测的精度, 在计算机视觉领域中, 广泛使用注意力机制。在原有的网络模型中, 引入了归一化的注意力模块 (NAM)^[3] 和协调注意力模块 (CA)^[4]。NAM 模块通过降低不太显著特征的权重, 可以更加关注图像中的重要区域和特征。这种机制可以应用于稀疏的权重惩罚, 使计算更加高效, 同时保持相同的性能。Yichao Liu 作者将 NAM 和 SE^[5], BAM, CBAM^[6] 等注意力机制进行比较, 该方法可以达到更高的准确率。图 2 为 NAM 通道注意机制和空间注意力机制结构图。SENet 提出了 SE Attention, 但 SE 注意力^[7]中只关注通道之间的相互依赖关系, 忽略了空间特征。注意力模块结构图如图 2 所示。

1. 福州理工学院计算与信息科学学院 福建福州 350506

[基金项目] 校一流课程 (LGJG2022014)、2023 年度福建省中青年教师教育科研项目 (科技类) - 重点项目 (JZ230068)

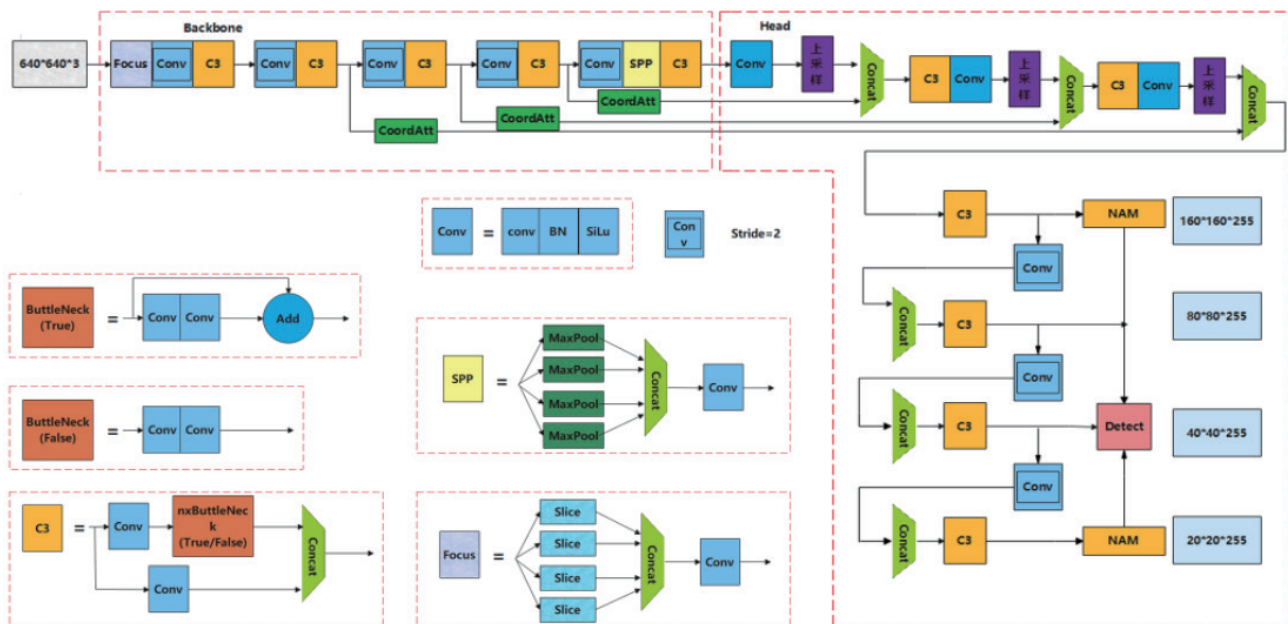
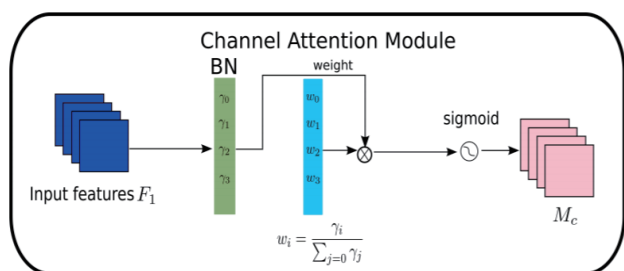
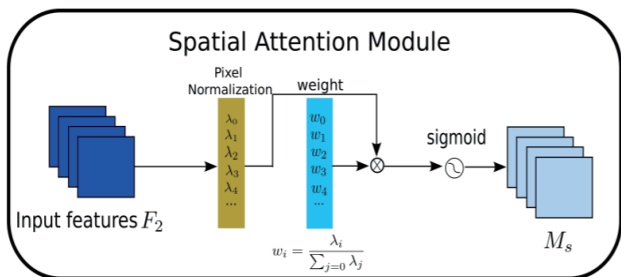


图1 改进的YOLOv5网络结构图



(a) 通道注意力机制



(b) 空间注意力机制

图2 注意力模块结构图

文献[4]提出的协调注意力模块(CA)克服了这一缺陷。CA能够捕获跨通道的信息以及方向和敏感位置上的信息,考虑了通道信息和方向相关的位置信息。这种机制具有足够的灵活性和轻量级,可以插入到轻量级网络的模块中。CA的结构如图3所示。为了避免大量的计算开销并将空间信息全部压缩到通道中,CA没有使用全局平均池化。相反,它通过分解全局平均池化,对水平方向和垂直方向进行平均池化,从而捕获具有精准位置信息的远程空间交互。这产生了两个一维向量,然后对这些向量进行空间信息编码。最后,将空间信息通过在通道上加权的方式进行融合。

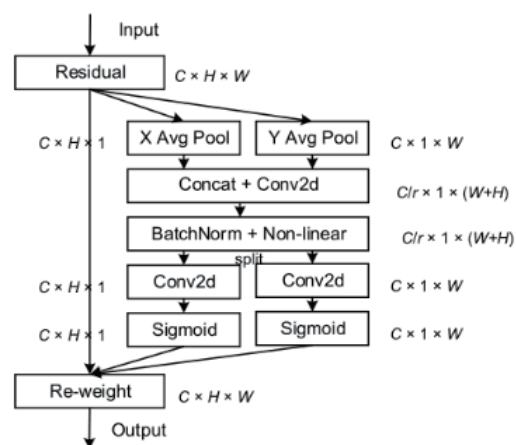


图3 CA协调注意力模块结构图

在改进的YOLOv5s网络模型中,采取了两个主要的改进措施。首先,在C3模块后面引入了CA模块,通过它来建立特征映射关系,对特征图进行注意力重构。这种改进有助于模型在检测交通标志时,能够更加关注小目标特征,从而在训练过程中赋予它们更高的权重,进而提高对小目标的检测精度。其次,将NAM模块加入到网络模型中,协调注意力模块作为输出,进一步优化网络模型的性能。通过这种方式,进一步提高特征值的精度,从而对整个网络模型进行优化。改进的模型在处理交通标志等复杂场景中的小目标检测问题时,能够更好地关注重要区域和特征,提高模型的检测精度和鲁棒性。

2 算法优化

2.1 传统目标检测算法

传统的目标检测损失函数依赖于边界框回归指标的聚合,只考虑距离、重叠区域和纵横比。以上的算法没有考虑

到所需真实框与预测框之间不匹配的方向问题。这种不足导致收敛速度较慢且效率较低，预测框可能在训练过程中“四处游荡”并产生更差的模型。

2.2 改进算法 SIou

Zhora Gevorgyan 提出了一种新的损失函数 SIou^[8]，该函数考虑了所需回归之间的向量角度，并重新定义了惩罚指标。通过在传统的神经网络和数据集上应用 SIou。结果表明，它提高了训练速度和推理准确性。SIou 由四部分组成：角度损失、距离损失、形状损失和 Iou 损失。在图 4 中，第一个 Angle cost 表示目标框 B 与回归框 B^{GT} 之间的夹角。当 B 到 B^{GT} 的夹角小于 α 时，该函数会向最小的 α 收敛；反之，它会向 β 收敛。这种新的损失函数 SIou 能够更好地衡量预测框与目标框之间的相似性，从而提高了模型的性能。

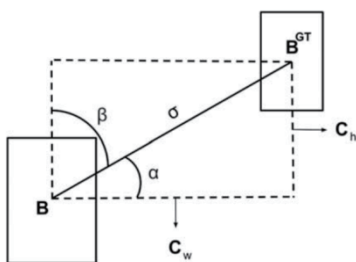


图 4 角度成本对损失共享图

(1) 角损失函数

角度损失函数公式如 (1) 所示。

$$\delta = 1 - 2 \times \sin^2(\arcsin(x) - \frac{\pi}{4}) \quad (1)$$

$$x = \frac{c_h}{\sigma} = \sin(\alpha) \quad (2)$$

$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \quad (3)$$

$$c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y}) \quad (4)$$

对公式 (1) 进行化简之后为公式 (5) ~ (7)。对于公式 (7)，当 $x = \frac{\pi}{4}$ 时， δ 取最大值。

$$\Lambda = 1 - 2 \times \sin^2(x - \frac{\pi}{4}) \quad (5)$$

$$\Lambda = \cos^2(x - \frac{\pi}{4}) - \sin^2(x - \frac{\pi}{4}) \quad (6)$$

$$\Lambda = \sin(2x) \quad (7)$$

(2) 距离损失函数

在计算损失函数时，直接使用角度损失是不合理的。当角度为平角但距离很长，和一个角度相反但距离较近的情况相比，肯定优先选择距离近的。为了平衡考虑距离和角度，可以将角度损失和距离损失结合起来考虑。通过综合考虑距离和角度等不同因素，设计一个更加平衡和有效的损失函数，以优化模型的训练和推理性能。

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho_t}) \quad (8)$$

式中： $\rho_x = (\frac{b_{c_x}^{gt} - b_{c_x}}{cw})^2$ 、 $\rho_y = (\frac{b_{c_y}^{gt} - b_{c_y}}{cw})^2$ 、 $\gamma = 2 - n$ 。

式 (8) 中独立的距离为 $1 - \frac{1}{e^{(2-\gamma)\rho_t}}$ 。其中 ρ_t 是距离损失使

用 2 次幂来赋权重。说明距离的影响要大于角度的影响，根据式 (7)，当角度 $x = \frac{\pi}{4}$ 时， r 为 1，因为负指数幂，角度损失增大。角度为 $x=0$ 时候， $r=2$ ，退化为距离损失。

(3) 形状损失函数

$$\Omega = \sum_{t=w,h} (1 - e^{-wt})^\theta \quad (9)$$

式中： $\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$ ， $\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$ 。

把角损失函数、距离损失函数、形状损失函数累加得到：

$$L_{box} = 1 - IOU + \frac{\Delta + \Omega}{2} \quad (10)$$

3 实验分析

3.1 实验数据

CCTSDb 2021^[9-11] 是一个全新的中国交通标志检测数据集，用于评估交通标志检测算法的性能。该数据集在 CCTSDb 2017 的基础上进行了扩展，添加了 4000 多幅真实的交通场景图像，并附有详细的注释。这些图像和注释准确地反映了现实世界中交通标志各种可能的表示形式和复杂的背景环境。

CCTSDb 2021 包括训练集和阳性样本测试集，共有 17 856 幅图像。这些图像中的交通标志根据其含义被分为三类：强制、禁止和警告。这三种类型的标志在交通规则和安全提示方面具有不同的意义，对于自动驾驶和智能交通系统的正常运行至关重要。该数据集的设计旨在提供一个更具挑战性的检测环境，通过使用更难以检测的样本替换了原始的易于检测的图像。这有助于研究人员开发更有效的交通标志检测算法，以适应复杂多变的实际交通环境。

3.2 实验结果分析

在系统中加入 SIou 算法、注意力机制下、新增 160×160 检测特征图小目标三种网络模型和原始 YOLOv5s.yaml 网络模型进行验证比较。在三种不同环境下对 CCTSDb 2021 数据集训练 100 轮，数据集输入图片大小设置为 640×640 实验结果如表 1 所示。

表 1 改进算法与原 YOLOv5s 模型实验结果表

算法	mAP_0.5/%	mAP_0.5:0.95/%
CIou	80.15	51.36
SIou	82.11	53.17
CoordAtt+NAM	81.74	52.66
新增 160×160 检测特征图	82.47	54.01
Ours	85.87 (+5.72)	57.21 (+5.85)

从表 2 的实验结果来看，改良的 SIou 算法、加入注意力机制以及新增 160×160 检测特征图小目标模型在 CCTSDb 2021 数据集上均表现出色。相比原始的 YOLOv5s 框架，这些改进模型在 mAP_0.5 指标上分别提高了 1.96%、1.59% 和

2.32%，达到了约 82.11%、81.74% 和 82.47%。这意味着这些模型在检测较小目标方面具有更好的性能。

当考虑到 mAP_{0.5:0.9} 指标时，这些改进模型同样表现出优于原始 YOLOv5s 框架的性能。其中，改良的 SIou 算法提升了 1.81%，加入注意力机制的网络模型提升了 1.3%，而新增 160×160 检测特征图小目标模型则提升了 2.65%。这表明这些改进模型在检测较大目标方面也具有较好的性能。

综合比较 mAP_{0.5} 和 mAP_{0.5:0.9} 两个指标，可以发现改良的 SIou 算法、加入注意力机制以及新增 160×160 检测特征图小目标模型在总体检测性能上都有所提升。其中，改良的 SIou 算法和新增 160×160 检测特征图小目标模型在两个指标上都表现得相对更好。

另外值得注意的是，当同时使用三种改进算法的网络模型时，其在 mAP_{0.5} 和 mAP_{0.5:0.9} 指标上的性能进一步提升。这表明这些改进算法之间可能存在协同作用，可以相互促进以提升目标检测性能。图 5 是最终检测对比结果。



(a) 基于 YOLOv5s 模型检测结果 (b) 基于本文模型检测结果

图 5 最终检测对比结果

4 结语

实验在 YOLOv5 目标检测模型中增加了一个 160×160 的小目标检测算法，用于提升了自动驾驶对 4×4 以上小目标的检测性能。实验中采用了 SIou 目标检测损失函数，提高目标框回归损失，提高目标检测的精度，减少了达到最高检测精度的轮次。在网络模型中加入了两种注意力机制，得到特

征值进一步提高优化网络的模型。相较于原 YOLOv5s 网络模型，其检测精度有着显著提升，提高了整体网络的鲁棒性，具有更好的检测性能和目标分辨能力。

参考文献：

- [1] ULTRALYTICS. YOLOv5 官网 [EB/OL]. [2020-08-09]. <https://github.com/ultralytics/yolov5>.
- [2] LI P, ZHAO W. Image fire detection algorithms based on convolutional neural networks[J]. Case Studies in Thermal Engineering, 2020, 19: 100625.
- [3] LIU Y, SHAO Z, TENG Y, et al. NAM: normalization-based attention module [EB/OL]. [2023-10-06]. <https://arxiv.org/abs/2111.12419.pdf>.
- [4] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design [EB/OL]. [2023-10-26]. <https://arxiv.org/pdf/2103.02907.pdf>.
- [5] Mobile network design [EB/OL]. (2021-01-27) [2023-10-12]. <https://arxiv.org/pdf/2103.02907.pdf>.
- [6] ZHOU X Y, GAI S Y, DA F P. An occluded pedestrian detection algorithm based on attention mechanism[J]. Acta optica sinica, 2021, 41(15): 1515001.
- [7] WOO S, PARK J, LEE J-Y, et al. CBAM: convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 3-19.
- [8] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 42(8): 2011-2023.
- [9] ZHORA G. SIou loss: more powerful learning for bounding box regression [EB/OL]. (2022-05-25) [2023-10-08]. <https://arxiv.org/abs/2205.12740.pdf>.
- [10] ZHANG J M, ZOU X, KUANG L D, et al. CCTSDB 2021: A more comprehensive traffic sign detection benchmark[J]. Human-centric computing and information sciences, 2022, 12: 23.
- [11] ZHANG J M, WANG W, LU C Q, et al. Lightweight deep network for traffic sign classification[J]. Annals of telecommunications, 2020, 75: 369-379.

【作者简介】

刘顺财（1986—），男，福建漳州人，硕士，副教授，研究方向：嵌入式 AI。

（收稿日期：2024-01-02）