

# 改进YOLOv5的中国交通标志检测算法

罗开涛, 何清龙

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2022年10月23日; 录用日期: 2022年11月23日; 发布日期: 2022年11月30日

## 摘要

针对实际复杂道路环境下, 交通标志像素小而密集导致的检测定位精度较低, 较高漏检的问题, 本文提出了一种改进YOLOv5的中国交通标志检测算法。通过遗传算法和K-means算法优化先验锚点框来提高小目标的检测准确度; 引入Bi-FPN结构对语义信息和位置信息进行特征融合, 增强网络表征能力; 增加注意力机制GAM, 提升网络对于复杂环境下的抗干扰能力和特征提取能力; 改变边框回归损失函数SIoU loss, 实现模型的快速收敛和准确定位。实验结果表明, 与标准的YOLOv5算法相比, 本文改进的算法的平均精度均值mAP、精确度P、召回率、F1提升值分别为10.03个百分点、4.7个百分点、2.6个百分点和3.48个百分点, 验证了该算法对中国交通标志检测的有效性。FPS达到了208.33FPS, 满足实时性检测要求。

## 关键词

YOLOv5, 中国交通标志, 小目标检测, 注意力机制

# Chinese Traffic Sign Detection Based on YOLOv5 Algorithm

Kaitao Luo, Qinglong He

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Oct. 23<sup>rd</sup>, 2022; accepted: Nov. 23<sup>rd</sup>, 2022; published: Nov. 30<sup>th</sup>, 2022

## Abstract

In order to solve low accuracy and missing detection of traffic signs in sign recognition task, caused by complex road environment, small and dense signs filled of whole picture, we propose an improved Yolov5 algorithm. Firstly, we use genetic algorithm and K-means algorithm to get

perfect anchors fit for the dataset which is good for the small objects. Secondly, we introduce the Bi-FPN structure to the neck of mode to combine high level semantical features and low-level features, which improves the representational capacity of the network. Thirdly, embedding the Gam attention into the backbone and neck could enhance feature extraction ability and anti-interference ability to messy background. Furthermore, a modified SIoU loss function is applied to optimize the training process and improve accuracy. Experimental results show that the mean average precision (mAP), precision, recall and F1 score are improved by 10.03%, 4.7%, 2.6 % and 3.48 % compared with the benchmark YOLOv5 algorithm respectively, which shows that the proposed algorithm is effective for Chinese traffic sign detection. In addition, 208.33 FPS could meet the real-time requirement.

## Keywords

YOLOv5, Chinese Traffic Sign, Small Object Detection, Attention Mechanism

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

近年来,随着中国经济持续快速发展,以新能源汽车为主体的智能网联汽车行业呈现快速增长态势。一个完备且安全的智能网联的汽车控制系统是智能车辆研究和发展的一个重要方向,它能够辅助驾驶甚至实现无人驾驶,有效地保障驾驶安全,提升驾驶体验,减少交通事故,提高交通效率。而交通标志识别系统(Traffic Sign Recognition System, TSRS)作为车辆控制系统领域中的一个重要组成部分,近年来受到学界和工业界的广泛关注。

传统的交通标志的检测主要依靠手工设计的传统特征,如交通标志的形状和颜色等,根据分类目标提取图像合适的特征,然后用提取的特征训练机器学习分类器,分类器离线训练好之后,便可以用来对待识别图像进行分类[1]。手工设计的特征往往无法适应复杂的驾驶环境,多阶段的特征提取和分类导致计算量巨大,在识别的精度和速度上都无法满足实际的需求。得益于计算机硬件的提升和深度学习的发展,大量的优秀的目标检测算法不断被提出并成功的运用于交通标志的检测和识别[2] [3] [4]。

基于深度学习的目标检测算法主要分为两阶段(two stage)算法(如, RCNN [5]、fast RCNN [6]、faster-RCNN [7])和一阶算法(one-stage) (如, SSD [8]、YOLO [9] [10] [11] [12])两类。在交通标志的识别中,李旭东[13]等将 YOLOv3-tiny 检测算法作为基础网络,提出了一种三尺度嵌套残差结构的交通标志快速检测算法,并在长沙理工大学中国交通标志检测数据集上进行试验,指示、禁令、警告三大类交通标志 F1 分数分别为 92.41%、93.91%、92.03%,检测时间为 5.0ms。鲍敬源[14]等提出了一种 Strong Tiny-YOLOv3 目标检测模型,该模型通过引入 Fire Modul 层进行通道变换,在减小模型参数的同时能够加深模型深度。同时,模型在 Fire Module 层之间加入 shortcut 来增强网络的特征提取能力,在中国交通标志检测数据集上精度均值达到了 85.56%。刘万军[15]等提出了一种新型多尺度融合卷积神经网络(SF-RCN),在基础特征提取网络中加入多尺度空洞卷积池化金字塔模块(MASPP),在网络中增加两个快速拼接模块(F-concat),融合模型中高层与低层的信息,该模型在中国交通标志检测数据集上精度均值达到了 87.48%。

本文基于 YOLOv5 模型,针对交通标志在图像中密集且占比小的特点,通过增加注意力机制,引入

多尺度融合, 通过改变边框回归损失函数等方法, 提升交通标志的检测的精度, 加快模型训练的收敛。改进的 YOLOv5s 模型在长沙理工大学中国交通标志检测数据集(CCTSDB) [16]上进行训练和验证。实验结果表明, 改进后的 YOLOv5s 模型比原始 YOLOv5 网络的 mAP 指标提升了 10.03 个百分点, 准确度提升了 4.7 个百分点, 召回率提升了 2.6 个百分点, 并且在小目标的检测中表现良好。

本文整体安排如下: 第 2 部分从 Input、Backbone、Neck、Head 四个方面对原始 YOLOv5 模型算法进行简单介绍, 再对优化锚框结构、引入的 GAM 注意力机制、特征融合结构 Bi-FPN、加快模型收敛的 SIOU 分别进行详细的阐述。第 3 部分针对中国交通标志检测数据集(CCTSDB)的基本情况、实验环境的基本设定进行了简单交代, 最后对实验结果和实验效果进行了简单分析。

## 2. YOLOv5 模型架构

### 2.1. 原始 YOLOV5 模型

2016 年, Joseph Redmon 等人提出了一种单阶段(one-stage)的目标检测网络 YOLO, 并经过不断技术迭代至 YOLOv5, 逐渐成为 one-stage 目标检测算法的基准算法。YOLO 的核心思想是将目标检测视为一个空间上分离的预测框和有联系的类概率的回归问题。YOLOv5 模型结构可以分为 Input、Backbone、Neck、Head 四个主要部分, 并根据模型的深度和宽度不同, 衍生出 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 四个模型, 根据不同设备, 适配于不同的数据集, 从而达到精度和速度的平衡。

Input 输入端主要包括 Mosaic 数据增强、自适应 anchor 计算, 自适应图片缩放三个方法。Mosaic 数据增强通过随机缩放、随机裁剪、随机排布的方式对 4 张不同图像进行拼接, 不仅能丰富检测目标的背景, 提高小目标的检测效果, 且能减少显存的使用, 加强 BN 效果。自适应 anchor 计算是指利用 K-Means 聚类和遗传算法, 自适应的计算不同训练集中的最佳锚框值, 根据最佳锚框生成的预测框能更好的匹配 Ground Truth, 进而加快模型收敛和提高定位精度。自适应图片缩放是对不同长宽比的图片自适应地添加最少的黑边, 减少冗余信息提高推理速度。

Backbone 主要包括 C3 [17]和 SPPF [18]模块。C3 模块基于 BottleneckCSP 模块, 将梯度的变化从头到尾地集成到特征图中, 支持特征传播, 鼓励网络重用特征, 增强网络提取特征的能力的同时减少网络参数数量。SPPF 通过不同池化核的最大池化进行特征提取, 提高网络的感受野。

Neck 主要包括特征金字塔网络(feature pyramid networks, FPN [19])和与路径聚合网络(path aggregation network, PAN [20])相结合的结构。FPN 是自顶向下的, 将高层的特征信息通过上采样的方式进行传递融合。PAN 是自底向上的, 将低层的位置信息传递融合。FPN 层自顶向下传达强语义特征, 而特征金字塔则自底向上传达强定位特征, 实现高层特征与低层特征融合互补, 使模型获得丰富的定位和分类特征信息。

Head 设置了三个检测头, 在三个检测层中图像分别下采样 32 倍、16 倍和 8 倍。不同尺寸的特征图用于检测不同尺寸的目标对象, 利用计算出来的先验框进行回归预测, 有利于大中小不同尺寸的目标, 尤其是小目标的检测。如图 1 所示为 YOLOv5s 的网络架构。

### 2.2. 改进 YOLOv5

#### 2.2.1. 先验锚框优化

大量实验证明, 合适的先验锚框能够很大程度的增加网络训练的稳定性 and 收敛速度, 提高网络检测准确率。由于不同的数据集的检测目标大小各不相同, YOLOv5 基于 COCO 数据生成的锚框不再适合交通标志的数据集。本文利用 K-Means 聚类和遗传算法, 计算出适合不同特征层的先验锚框, 如表 1 所示。

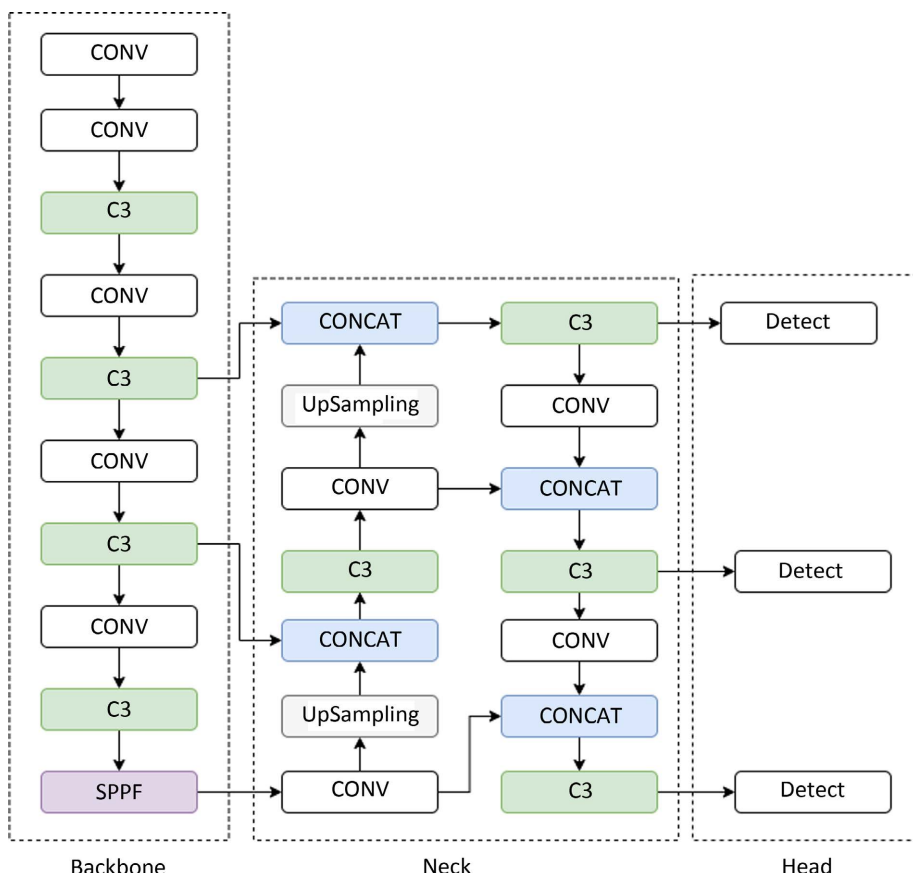


Figure 1. YOLOv5s network architecture  
 图 1. YOLOv5s 网络结构

Table 1. Different anchor box size  
 表 1. 不同尺度的先验锚框

尺寸		先验锚框	
80*80	[7,7]	[11,11]	[14,14]
40*40	[18,18]	[24,23]	[32,30]
20*20	[26,25]	[54,54]	[96,113]

### 2.2.2. 引入特征融合结构 Bi-FPN

卷积神经网络通过卷积不断的提取图像的特征，随着网络深度的增加，高级的语义特征被提取，能够大幅度提高网络对图片的识别，但与此同时，目标的位置信息也在特征提取的过程中不断被弱化。Yolov5 采用 FPN + PAN 相结合的结构实现特征的融合，自下而上的 FPN 结构能够提供更强的语义信息(有利于物体分类)，自上而下的 PAN 结构能够提供更强的位置信息(有利于物体定位)，这样使得预测特征层同时具备较高的语义信息和位置信息，提高了目标检测任务精度(见图 2)。

传统的特征融合只是简单的将不同特征层进行叠加和融合，而不对特征层进行区分。不同特征层对于模型的性能贡献是不同的，针对这一问题，本文采用加权双向(自顶向下 + 自低向上)特征金字塔网络 Bi-FPN [21]。它引入了可学习的权重来衡量不同输入特征的重要性，能够更好的平衡不同尺度的特征信息，同时重复应用自顶向下和自底向上的多尺度特征融合。Bi-FPN 的操作可如下公式(1)表示：

$$P = \sum_i \frac{w_i}{\varepsilon + \sum_j \omega_j} \cdot P_i \tag{1}$$

其中,  $P$  表示融合后的特征层,  $P_i$  表示不同的特征层,  $w_i$  是一个可学习的权重参数, 参数  $\varepsilon = 0.0001$  可以避免除零导致的数值计算问题。通过网络的自我学习能够在不同特征层提取关键信息, 促进特征融合。

为了更好的检测出小目标交通标志, 在网络结构中, 多个 Bi-FPN 结构能够融合来自多个低层的特征, 如第 13 层, 第 17 层、第 21 层, 第 25 层。并通过网络训练权重, 平衡语义和位置信息, 从而提高目标的分类和定位准确率。

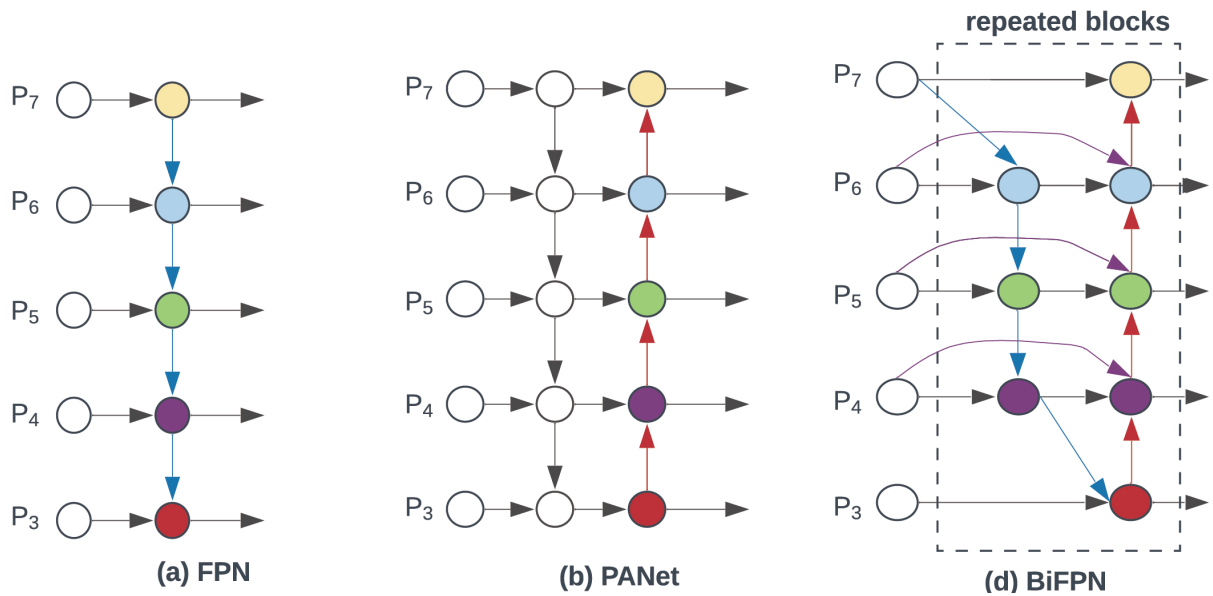


Figure 2. FPN and Bi-FPN architecture comparison  
图 2. FPN 和 Bi-FPN 对比图

### 2.2.3. 添加注意力机制(GAM)

注意力机制的本质就是通过观察图像的全局信息, 提取感兴趣的关键信息, 抑制甚至过滤掉无用的背景和冗余信息, 提高信息处理效率和提取关键信息的准确率。

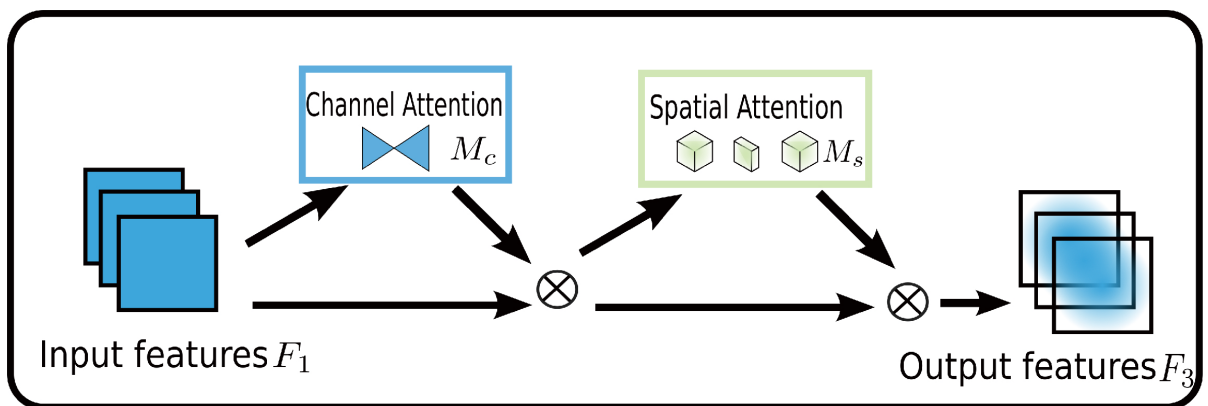


Figure 3. The overview of GAM attention  
图 3. GAM attention 过程示意图

GAM 注意力机制(Global Attention Mechanism) [22]不同于 CBAM 模块(Convolutional Block Attention Module) [23], 它不仅仅考虑了将注意力映射沿特征图的通道与空间两个独立的维度进行注入, 更考虑了通道与空间的相互作用, 引入并放大了跨维度的交互作用, 在所有三个维度上捕捉重要特征。GAM attention 整个过程如图 3 所示。

GAM attention 包括通道注意力子模块和空间注意力子模块。通道注意力子模块使用三维排列在三个维度上保留信息, 用一个两层的 MLP (多层感知器)放大跨维通道 - 空间依赖性。对于输入的特征图, 首先进行维度转换, 经过维度转换的特征图输入到 MLP, 再转换为原来的维度, 并进行 Sigmoid 处理输出(见图 4)。空间注意力子模块使用两个卷积层进行空间信息融合, 删除了池化层对空间位置信息的损失影响, 利用通道缩减和 7\*7 的卷积来减少通道数量, 在减少计算量的同时专注于对于空间维度信息的提取(见图 5)。

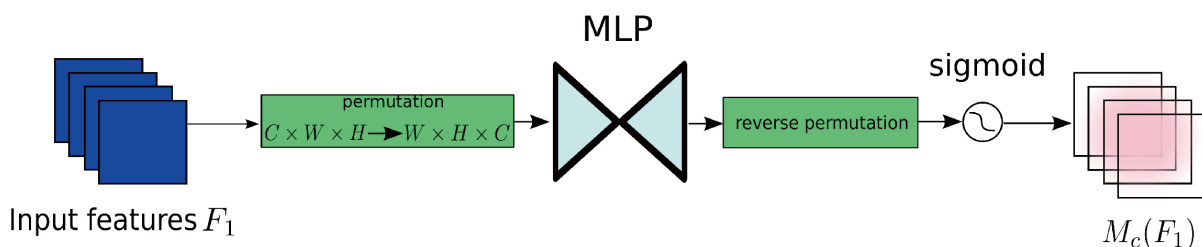


Figure 4. Channel attention submodule  
图 4. 通道注意力子模块

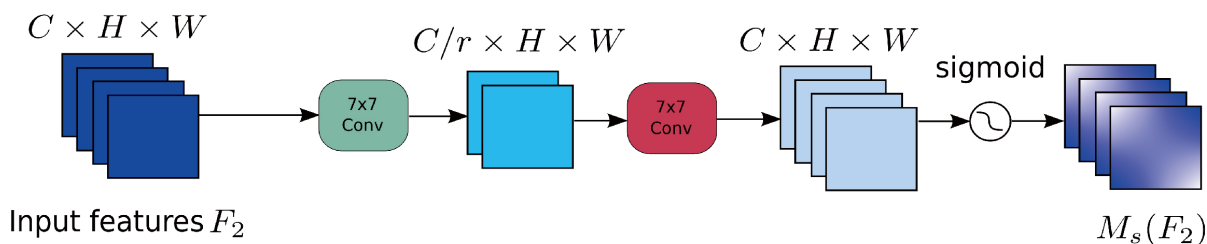


Figure 5. Spatial attention submodule  
图 5. 空间注意力子模块

### 2.2.4. 边界回归损失函数的改进

YOLOv5 的损失函数由分类损失(classification loss)、目标置信度损失(object loss)和定位框 bbox 回归损失(boundingbox regression loss)三部分组成。其中定位框 bbox 回归损失采用完全交并比损失函数(Complete-IoU loss, CIoU Loss [24])实现目标框预测。具体计算如公式(2)~(5)所示。

$$\mathcal{L}_{CIoU} = 1 - CIoU \tag{2}$$

$$CIoU = IoU(B, B^{gt}) + \frac{\rho^2(B, B^{gt})}{C^2} + \alpha v \tag{3}$$

$$v = \frac{4}{\pi} \left( \arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right) \tag{4}$$

$$\alpha = \frac{v}{1 - IoU(B, B^{gt}) + v} \tag{5}$$

其中,  $B, B^{gt}$  代表预测框和真实框,  $\rho$  两个框的中心点坐标的欧式距离, 而  $c$  则是能够同时包含预测框和

真实框的最小闭包区域的对角线距离,  $w^{gt}$ 、 $h^{gt}$  代表真实框框的宽度和高度,  $w$ 、 $h$  代表预测框的宽度和高度。 $v$  为预测框和真实框长宽比例差值的归一化,  $\alpha$  为平衡因子, 用来平衡长宽比例造成的损失和 IoU 部分造成的损失。

传统的 IoU 损失仅考虑了预测框和真实框之间的交并集造成的结果损失, 而忽略预测框在回归过程中 IoU 损失对于模型收敛的影响, 比如预测框和真实框完全重叠和完全不重叠的情况在计算 IOU 损失中没有区别, 无法帮助模型收敛。GIoU [25] 引入能够同时框住真实框与预测框的最小的框, 解决了真实框与预测框完全分离时 IoU 无法衡量两者距离的问题。DIoU 引入了衡量中心点坐标的距离差异的惩罚参数  $p$  和  $c$ , 解决了预测框和真实框完全重叠时, 使得预测框中心点快速收敛于真实框的中心点。YOLOv5 采用 CIoU 损失函数是在 DIoU 的基础上考虑到了预测框与真实框的长宽比例问题, 更好的解决区域重叠问题带来的收敛低效问题。然而目前为止提出的方法中, 都没有考虑到真实框与预测框框之间的方向而导致收敛速度慢的问题。这种不足导致收敛速度较慢且效率较低, 因为预测框可能在训练过程中“四处游荡”并最终产生更差的模型。Gevorgyan Z. 等人提出了全新的 SIoU [26] 损失函数。

SIoU 考虑了真实框和预测框之间的向量角度, 重新定义了惩罚指标。具体包含四个部分: 角度损失 (Angle cost)、距离损失 (Distance cost)、形状损失 (Shape cost)、IoU 损失 (IoU cost)。

a) 角度损失

$$\Lambda = 1 - 2 * \sin^2 \left( \arcsin x - \frac{\pi}{4} \right) \tag{6}$$

$$x = \frac{c_h}{\sigma} = \sin(\alpha) \tag{7}$$

$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2} \tag{8}$$

$$c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y}) \tag{9}$$

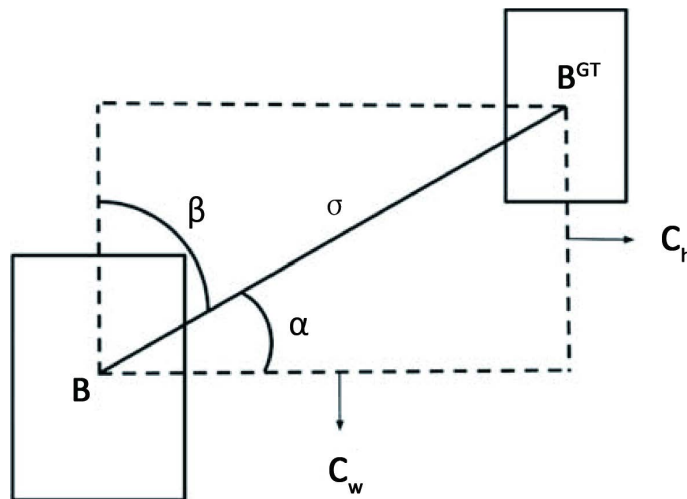


Figure 6. Angle cost  
图 6. 角度损失示意图

如图 6 所示, 角度损失为  $\Lambda$ , 计算公式为公式(6)~(9), 其中  $(b_{c_x}, b_{c_y})$  为预测框中心点坐标,  $(b_{c_x}^{gt}, b_{c_y}^{gt})$  为真实框中心点坐标,  $\sigma$  表示真实框和预测框中心点的距离,  $c_h$  为真实框和预测框中心点的高度差。

b) 距离损失

$$\Delta = \sum_{t=x,y} (1 - e^{-r\rho t}) \tag{10}$$

$$\rho = \frac{(b_{c_x}^{gt} - b_{c_x})^2}{c_w}, \rho = \frac{(b_{c_y}^{gt} - b_{c_y})^2}{c_h}, r = 2 - \Lambda \tag{11}$$

距离损失为  $\Delta$ , 计算公式为公式(10)~(11), 其中  $c_w$  和  $c_h$  为真实框和预测框最小外接矩形的宽和高。

c) 形状损失

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega t})^\theta \tag{12}$$

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})} \tag{13}$$

$$\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \tag{14}$$

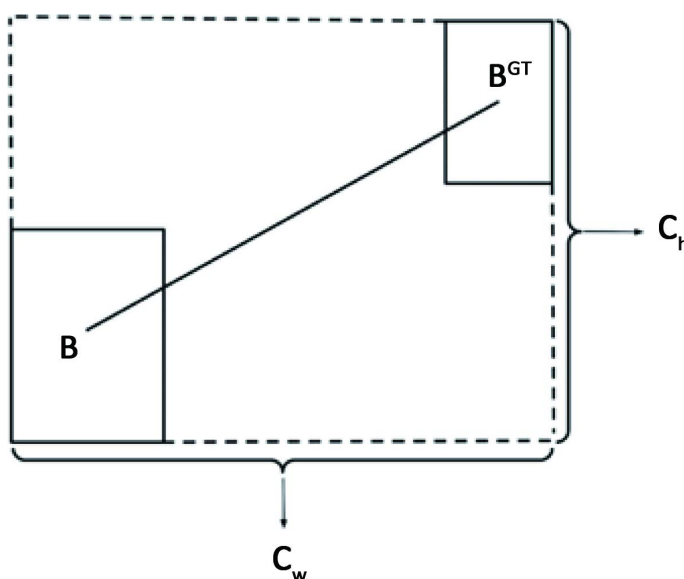


Figure 7. Shape cost  
图 7. 形状损失示意图

如图 7 所示, 形状损失为  $\Omega$ , 计算公式为公式(12)~(14), 其中  $w, h, w^{gt}, h^{gt}$  分别为预测框和真实框的宽和高。 $\theta$  为超参数, 控制对形状损失的关注程度, 为了避免过于关注形状损失而降低对预测框的移动。最终 SIOU 损失表达式如公式(15)所示。

$$\mathcal{L}_{SIOU} = 1 - \text{IoU}(B, B^{gt}) + \frac{\Omega + \Delta}{2} \tag{15}$$

在边框回归损失中, SIOU Loss 中各项损失考虑全面, 使得收敛速度更快, 精度更高, 优于原网络中的 CIoU Loss, 因此本文采用性能更优的 SIOU Loss 边框回归损失函数。如图 8 所示为本文改进后的 YOLOv5 网络模型结构。



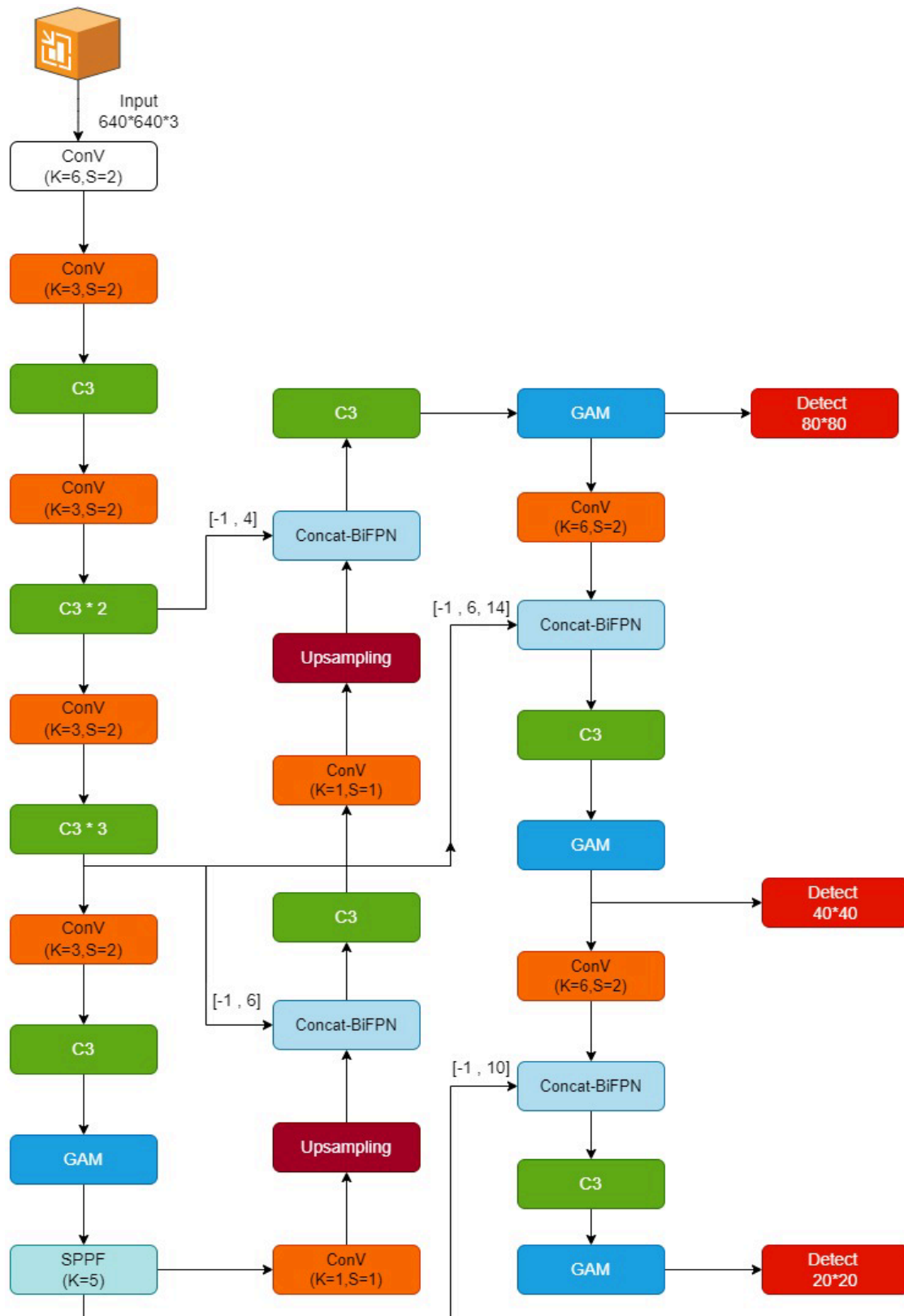


Figure 8. Improved YOLOv5 network architecture  
 图 8. 改进的 YOLOv5 网络结构图

### 3. 实验结果与分析

#### 3.1. 数据集

中国交通数据集 CCTSDB2021 (CSUST Chinese Traffic Sign Detection Benchmark 2021)由长沙理工大学综合交通运输大数据智能处理湖南省重点实验室张建明老师团队制作完成[16]。数据集有 17,856 张图片, 其中训练集 16,356 张, 测试集 1500 张(见表 2)。目前标注数据有三大类: 指示标志(Mandatory)、禁止标志(Prohibitory)、警告标志(Warning) (如图 9 所示)。图片分辨率有 860\*480, 1280\*720, 1920\*1080 不等。交通标志像素大小不同并按照 5 个标准(XS, S, M, L, XL)进行划分(见表 3)。



**Figure 9.** Chinese traffic signs with meaningful classification  
**图 9.** 交通标识示意图

**Table 2.** Traffic sign dataset classification and amount  
**表 2.** 数据集交通标志类别及数量

分类	training 集数量	test 集数量
Prohibitory (禁止)	13,867	2177
Waring (指示)	4598	333
Mandatory (警告)	8363	718

**Table 3.** Cropped image size and amounts  
**表 3.** 目标大小分类及数量

分类	尺寸(像素)	数量
XS	$\leq 210$	813
S	[210,400]	807
M	[400,1000]	828
L	[1000,2000]	408
XL	$\geq 2000$	372

### 3.2. 实验环境和设定细节

本文实验在 Linux 操作系统 Ubuntu18.04 下进行, 显卡为 Geforce GTX 1080Ti, 11GB 运行内存。整个实验基于深度学习框架 pytorch 1.7, 实验环境是 python3.8, GPU 加速软件为 CUDA11.2。实验的超参数设定如下: 训练 100 epochs, 批样本数量(batch size)为 32, 随机梯度下降法(SGD), 动量(momentum) 0.957, 权重衰减(weight decay)为 0.00048, 初始学习率为 0.012。

### 3.3. 评价标准

本文主要使用目标检测模型常用的评价指标。包括精确度(precision, P)、召回率(recall, R)、平均精度均值(mean average precision, mAP)、F1 分数(F1 score)和每秒处理帧率(FPS)。各项评价指标计算公式如下:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F_1 = \frac{2PR}{P + R}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N \int_0^1 P(R) dR$$

其中,  $TP$  为将正类预测为正类的数量,  $FN$  为将正类预测为负类的数量,  $FP$  为将负类预测成正类的数量,  $TN$  为将负类预测成负类的数量,  $N$  表示类别数。

### 3.4. 实验结果及分析

为了验证改进的 YOLOv5 网络模型的目标识别性能, 本文复现了原始的 YOLOv5 (v6.1)模型在交通标志数据集集中的表现(实验 Repo), 以此为参考基线, 对所有改进点进行了消融实验, 并将融合所有改进点的改进 YOLOv5 网络模型(实验 Imp-YOLO)与参考基线进行了对比。实验结果如表 4 所示。

**Table 4.** Experiments results

**表 4.** 实验结果

Model	P	R	mAP (0.5)	mAP (0.5:0.95)	F1	FPS
original	90.80	69.2	76.3	52.1	78.542	123.46
Repo	88.10	73	79.33	50.87	79.84	246.7
GAM	91.77 (+3.67)	76.23 (+3.23)	82.3 (+2.97)	54.07 (+3.2)	83.28 (+3.44)	202.99
Bi-FPN	92.1 (+4.0)	75.4 (+2.4)	82.13 (+2.8)	53.73 (+2.87)	82.91 (+3.07)	232.89
SIoU	91.5 (+3.4)	74.97 (+1.97)	82.03 (+2.7)	54.13 (+3.27)	82.41 (+2.57)	250.42
Imp-YOLO	92.8 (+4.7)	75.6 (+2.6)	85.7 (+6.37)	60.9 (+10.03)	83.32 (+3.48)	208.33

由实验数据可知, 改进的 YOLOv5 模型在各项指标中都获得了显著提升。在原始模型的基础上引入 GAM 注意力模型, 能够更好利用空间位置信息和特征通道信息以及二者之间的交互作用, 在三个维度捕

捉特征信息,对于依赖位置信息的密集检测任务具有更好的检测效果,召回率  $R$  提升了 3.23%,定位  $mAP$  提升了 3.2%。 $Bi-FPN$  结构使用加权的方式进行特征融合,可以将低层特征与高层特征进行有效的平衡和融合。针对目标数据集小目标较多的特点,高效的特征融合有效的提升了模型对于特征的提取能力,使模型准确度  $P$  提升了 4.0%。最后引入  $SIoU$  作为预测框回归的损失函数,解决了原始  $CIoU$  没有考虑真实框与预测框框之间的方向不同而导致收敛速度慢的问题,提高模型边界框的定位精度 3.27%,检测速度也提升了 3.72FPS。

总体而言,相比较原模型  $YOLOv5$ ,准确度  $P$  达到了 92.8%,提升了 4.4%,召回率 75.6%,提升了 2.6%, $F1$ -score 为 83.32,提升了 3.48%。 $mAP@0.5:0.95$  提升 10.03%,模型高精度边界拟合能力显著增强。在保证较高的检测精度的同时,检测速率达 208.33FPS 能够满足实时性需求。实验结果表明注意力机制、特征融合和边框回归损失函数的共同优化可以有效地提升交通标志的检测的效果。图 10 为改进前后检测结果对比。



Figure 10. Improved YOLOv5 detection result comparison

图 10. 改进 YOLOv5 算法对比检测结果

由图 10 可以看出,原始的  $YOLOv5$  算法训练的模型在检测中存在漏检的情况,尤其是在连续密集小目标和夜晚环境复杂的情况下,检测效果欠佳。第一张图漏检一个指示标识,第二张图在夜晚漏检两个禁止标识,第三张图漏检中间的禁止标识。而本文改进的  $YOLOv5$  算法成功的在图像上检测出了数据集中标注的所有交通标志,并且定位准确,无漏检现象。

在不同类别的检测上,改进的  $YOLOv5$  模型相较于原始的  $YOLOv5$  定位准确度提升更为明显(见表 5)。在三大类别交通标志的定位  $mAP@0.5:0.95$  上分别达到了 62.5%、61.2%和 58.9%。在禁止标志定位精度提升了 9.7%,在警告标志定位精度提高了 8.27%,在指示标志上定位精度提高了 12.1%。在不同尺寸的目标检测当中,对于密集小目标的提升相较于大目标更为明显(见表 6)。对于尺寸小于 210 像素的  $XS$  目标提升达到了 13.83%,引进的注意力机制和特征融合对于大量密集的小目标更为关注,使之相较于大目标提升明显。

**Table 5.** Experiments result for classification  
**表 5.** 不同类别的实验检测结果

Model	Prohibitory		Waring		Mandatory	
	mAP (0.5)	mAP (0.5:0.95)	mAP (0.5)	mAP (0.5:0.95)	mAP (0.5)	mAP (0.5:0.95)
Repo	80.90	52.8	86.17	52.93	71	46.8
GAM	84.53 (+3.63)	56.77 (+3.97)	88.2 (+2.03)	55.9 (+2.97)	74.17 (+3.17)	49.5 (+2.7)
Bi-FPN	84.17 (+3.27)	56.43 (+3.63)	87.8 (+1.63)	55.07 (+2.14)	74.53 (+3.53)	49.67 (+2.87)
StoU	84.83 (+3.93)	56.7 (+3.9)	87.63 (+1.46)	56.43 (+3.5)	73.57 (+2.57)	49.3 (+2.5)
Imp-YOLO	85.7 (+4.8)	62.5 (+9.7)	89.8 (+3.63)	61.2 (+8.27)	81.5 (+10.5)	58.9 (+12.1)

**Table 6.** Experiments results for cropped image size  
**表 6.** 不同目标大小的实验结果

Model	XS		S		M		L		XL	
	mAP (0.5)	mAP (0.5:0.95)	mAP (0.5)	mAP (0.5:0.95)	mAP (0.5)	mAP (0.5:0.95)	mAP (0.5)	mAP (0.5:0.95)	mAP (0.5)	mAP (0.5:0.95)
Repo	73.47	35.87	86.93	53.97	92.93	63.07	96.9	69.07	96.4	74.23
GAM	80.9 (+7.43)	41.73 (+5.87)	87.37 (+0.43)	56.93 (+2.97)	94.13 (+1.2)	65.77 (+2.7)	97.4 (+0.5)	71.63 (+2.57)	95.9 (-0.5)	73.53 (-0.7)
Bi-FPN	79.8 (+6.33)	41.57 (+5.7)	87.23 (+0.3)	55.90 (+1.93)	93.6 (+0.67)	65 (+1.93)	97.07 (+0.17)	70.97 (+1.9)	95.63 (-0.77)	73.13 (-1.1)
StoU	80.43 (+6.97)	42.87 (+7)	85.80 (-1.13)	56 (+2.03)	93.1 (+0.17)	64.95 (+1.88)	98 (+1.1)	72 (+2.93)	96.95 (+0.55)	74.85 (+0.62)
Imp-YOLO	82.2 (+8.73)	49.7 (+13.83)	87.40 (+0.47)	60.40 (+6.43)	92.9 (-0.03)	68.20 (+5.13)	97.4 (+0.5)	75.3 (+6.23)	94.8 (-1.60)	75.5 (+1.27)

#### 4. 结论

本文针对交通标志数据集中目标小且密集, 背景复杂且多变的特点, 提出了一种改进的 YOLOv5 的目标检测算法。通过引入 GAM 注意力机制, 提高网络的特征提取能力, 忽略无关的特征, 引入 Bi-FPN 结构加强高级语义和底层特征之间的平衡和融合, 通过替换边界框回归损失函数, 加快模型的收敛速度。实验表明, 本文提出的改进算法与原始算法相比, 能够显著提升目标定位精度, 减少漏检。

#### 基金项目

本文获得感谢国家自然科学基金(资助号: 12261021, 11801111)、贵州省科技计划项目([2019]1122)和贵州省科技平台人才资助项目([2018]5781)联合资助。

#### 参考文献

- [1] 孙颖, 葛平淑, 刘德全. 交通标志的检测与识别研究综述[J]. 大连民族大学学报, 2019, 21(5): 412-417.
- [2] 陈飞, 刘云鹏, 李思远. 复杂环境下的交通标志检测与识别方法综述[J]. 计算机工程与应用, 2021, 57(16): 65-73.

- [3] 伍晓晖, 田启川. 交通标志识别方法综述[J]. 计算机工程与应用, 2020, 56(10): 20-26.
- [4] 马健, 张敏, 张丽岩, 段晓科. 交通标志识别系统研究综述[J]. 物流科技, 2021, 44(10): 69-74.  
<https://doi.org/10.13714/j.cnki.1002-3100.2021.10.022>
- [5] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [6] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. <https://doi.org/10.1109/ICCV.2015.169>
- [7] Ren, S., He, K., Girshick, R. and Sun, J. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149.  
<https://doi.org/10.1109/TPAMI.2016.2577031>
- [8] Liu, W., Anguelov, D., Erhan, D., et al. (2016) SSD: Single Shot Multibox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016. Lecture Notes in Computer Science*, Springer, Cham, 21-37.  
[https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2)
- [9] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788.  
<https://doi.org/10.1109/CVPR.2016.91>
- [10] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
- [11] Redmon, J. and Farhadi, A. (2018) Yolov3: An Incremental Improvement. ArXiv: 1804.02767.
- [12] Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. (2020) Yolov4: Optimal Speed and Accuracy of Object Detection. ArXiv: 2004.10934.
- [13] 李旭东, 张建明, 谢志鹏, 王进. 基于三尺度嵌套残差结构的交通标志快速检测算法[J]. 计算机研究与发展, 2020, 57(5): 1022-1036.
- [14] 鲍敬源, 薛榕刚. 基于 YOLOv3 模型压缩的交通标志实时检测算法[J]. 计算机工程与应用, 2020, 56(23): 202-210.
- [15] 刘万军, 李嘉欣, 曲海成. 基于多尺度卷积神经网络的交通标示识别研究[J]. 计算机应用研究, 2022, 39(5): 1557-1562. <https://doi.org/10.19734/j.issn.1001-3695.2021.10.0457>
- [16] Zhang, J., Zou, X., Kuang, L.-D., Wang, J., Sherratt, R.S. and Yu, X. (2022) CCTSDB 2021: A More Comprehensive Traffic Sign Detection Benchmark. *Human-Centric Computing and Information Sciences*, **12**, Article No. 23.  
<https://doi.org/10.22967/HCCIS.2022.12.023>
- [17] Wang, C.Y., Liao, H.Y.M., Wu, Y.H., et al. (2020) CSPNet: A New Backbone That Can Enhance Learning Capability of CNN. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, 14-19 June 2020, 390-391. <https://doi.org/10.1109/CVPRW50498.2020.00203>
- [18] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1904-1916.  
<https://doi.org/10.1109/TPAMI.2015.2389824>
- [19] Lin, T.Y., Dollár, P., Girshick, R., et al. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 936-944.  
<https://doi.org/10.1109/CVPR.2017.106>
- [20] Liu, S., Qi, L., Qin, H., et al. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 8759-8768.  
<https://doi.org/10.1109/CVPR.2018.00913>
- [21] Tan, M., Pang, R. and Le, Q.V. (2020) Efficientdet: Scalable and Efficient Object Detection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 10778-10787.  
<https://doi.org/10.1109/CVPR42600.2020.01079>
- [22] Liu, Y., Shao, Z. and Hoffmann, N. (2021) Global Attention Mechanism: Retain Information to Enhance Channel-Spatial Interactions. ArXiv: 2112.05561.
- [23] Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. *Proceedings of the European Conference on Computer Vision (ECCV)*, Cham, 3-19.  
[https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1)
- [24] Zheng, Z., Wang, P., Ren, D., et al. (2021) Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Transactions on Cybernetics*, **52**, 8574-8586.

<https://doi.org/10.1109/TCYB.2021.3095305>

- [25] Rezatofighi, H., Tsoi, N., Gwak, J.Y., *et al.* (2019) Generalized Intersection over Union: A Metric and a Loss for Bounding Box Regression. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 658-666. <https://doi.org/10.1109/TCYB.2021.3095305>
- [26] Gevorgyan, Z. (2022) SIoU Loss: More Powerful Learning for Bounding Box Regression. ArXiv: 2205.12740.