

基于 Cache-DCN YOLOX 算法的交通标志检测方法研究

高尉峰¹, 王如刚¹, 王媛媛¹, 周 锋¹, 郭乃宏²

(1. 盐城工学院 信息工程学院, 江苏 盐城 224051;

2. 盐城雄鹰精密机械有限公司, 江苏 盐城 224006)

摘要: 针对传统方式识别交通标志算法存在的检测精度较低的问题, 提出了一种基于 Cache-DCN YOLOX 算法的交通标志识别方法; 在该方法中, 使用 DCN 可变形卷积替换 backbone 中的普通卷积, 有效地增大了模型的感受野, 提高了特征提取能力; 使用 EIoU 损失函数代替 YOLOX 中的 GIoU 损失函数, 优化了训练模型, 提高了收敛的速度; 优化设计了 YOLOX 算法中的强弱两阶段的训练过程, 增强了模型的泛化性能, 同时加入 cache 方案, 进一步提高了检测精度; 在交通标志数据集 TT100K 上进行了实验, 提出方法的检测精度为 67.2%, 比原 YOLOX 算法的检测精度提升了 6.4%, 同时, 在被遮挡的小目标等多种受干扰的环境下, 提出的方法能够精确地检测出交通标志, 并有着较好的置信度, 满足实际需求。

关键词: 深度学习; YOLOX; 交通标志识别; 可变形卷积; 小目标检测

Research on Traffic Sign Detection Method Based on Cache-DCN YOLOX Algorithm

GAO Weifeng¹, WANG Ruguang¹, WANG Yuanyuan¹, ZHOU Feng¹, GUO Naihong²

(1. School of Information Technology, Yancheng Institute of Technology, Yancheng 224051, China;

2. Yancheng XiongYing Precision Machinery Company Limited, Yancheng 224006, China)

Abstract: Aiming at the problem of low detection accuracy of traditional traffic sign recognition algorithms, a traffic sign recognition method based on cache-convolutional network (Cache-DCN) YOLOX is proposed. In this method, the deformable convolutional network (DCN) is used to replace the ordinary convolution in the backbone, which effectively increases the receptive field of the model and improves the ability of feature extraction; The EIoU loss function is used instead of the GIoU loss function in YOLOX, which optimizes the training model and improves the convergence speed; The training processes of the strength and weak stages in the YOLOX algorithm are optimized and designed to enhance the generalization performance of the model. At the same time, the cache scheme is added to further improve the detection accuracy. Experiments are conducted on the traffic sign dataset TT100K, the detection accuracy of the proposed method reaches by 67.2%, it is 6.4% higher than that of the original YOLOX algorithm, and the proposed method can accurately detect traffic signs with good confidence and meet practical needs in various disturbed environments such as occluded small targets.

Keywords: deep learning; YOLOX; traffic signs recognition; deformable convolution; small target detection

0 引言

在无人驾驶系统中, 交通标志对于车辆的行驶具有非常重要的作用。作为自动驾驶中的关键技术, 交通标志检测识别精度的提升对于经济发展、社会和谐相处都具有十分重要的意义。传统的目标检测方法主要依靠人工自主选择特征从而进行训练, 如 HOG、SIFT、Haar 等^[1-3], 该类

方法不仅计算量大, 且检测速度和精度都不能满足实际的应用需求。随着深度神经网络的逐渐完善, 由于较强的特征提取能力, 深度学习技术被广泛应用于目标检测任务中。目前, 常用的目标检测算法有以 YOLO^[4-6] 系列算法、SSD^[7-8] 系列算法为代表的单阶段算法和以 R-CNN^[9-10] 系列算法为代表的两阶段算法。其中单阶段算法是将目标检测过程看作一个回归问题来处理, 而两阶段算法则是先

收稿日期: 2023-06-28; 修回日期: 2023-08-03。

基金项目: 江苏省研究生实践创新计划项目(SJCX22_1685, SJCX21_1517); 江苏省高等学校自然科学研究重大项目(19KJA110002); 国家自然科学基金项目(61673108); 江苏省高校自然科学研究面上项目(18KJD510010, 19KJB510061); 江苏省自然科学基金项目(BK20181050)。

作者简介: 高尉峰(1997-), 男, 硕士研究生。

通讯作者: 王如刚(1975-), 男, 博士, 教授。

引用格式: 高尉峰, 王如刚, 王媛媛, 等. 基于 Cache-DCN YOLOX 算法的交通标志检测方法研究[J]. 计算机测量与控制, 2024, 32(2): 71-77, 84.

产生一个建议区域,再对该区域进行分类及边界框回归,提取特征。由于具有良好的实时性,因此,单阶段检测被更多地应用在交通标志检测领域,许多研究人员对单阶段算法进行了优化以便更好地检测交通标志。如,为了解决天气的干扰对交通标志检测的影响问题,代小宇等研究人员在 YOLOv5 中加入了 SE 注意力机制并融合双向特征金字塔 BiFPN,相比于原算法,检测精度和召回率分别提高了 7.94% 和 4.98%^[11]。针对交通标志在图像中占据位置小、检测中精度较低的问题,尹宋麟等研究人员在 YOLOv4 的基础上,使用 152×152 的尺度检测层替代原有的 19×19 的大感受野检测层,并加入了 ECA 模块去提高算法的检测能力,在 TT100K 数据集上训练的准确率相较于原算法提升了 4.58%^[12]。为解决交通标志检测任务中,在复杂路况出现错检漏检的情况,赵宏等研究人员提出了一种改进的 CGS-Ghost YOLO 算法,将 Focus 模块用 StemBlock 代替进行采样,降低了模型的参数数量的同时提高了模型的检测精度^[13]。为了提高各个类别的检测精度,Dubey 等研究人员在 RetinaNet 的框架上以更多层数的 ResNet 为基础网络,同时提出一种两步 TSR 方法,在交通标志数据集 CCTSDB 上进行分类检测,得到了较好的检测精度^[14-16]。为解决目标检测过程中计算量大的问题,Rachmadi 等研究人员提出了一种新型的轻量级空间金字塔卷积神经网络分类结构,该结构将每一层金字塔所学习到的特征与第一层叠加,使得每一层网络所能得到的学习区域更加平滑,最后的检测训练实现了在较低的计算量基础上有着良好的检测精度^[17]。为解决交通标志检测速度低的问题,党宏社等研究人员在 YOLOv5 算法上引入了 MobileNetv3 主干网络,结合了 RFB 模块与 ECA-Net 模块,并采用了 Matrix NMS 筛选候选框,最终对交通标志实现了更快的检测速度^[18]。为解决传统检测方法速度慢、精度低的问题,吕禾丰等研究人员使用加权 Cluster 非极大值抑制 (NMS) 代替 YOLOv5 的加权 NMS 算法,提高了生成检测框的准确率,改进后的算法比原算法提高了 6.23% 的准确率^[19]。为了解决交通标志检测实时性不足的问题,王林等研究人员在 YOLOv3 算法中融入了金字塔池化模块 SPP,并充分利用上下文信息增强对小目标的检测,实现了更快的检测速度,并提高了对小目标的检测精度^[20]。从现有的研究结果来看,众多学者基于单阶段目标检测算法做出改进,能够良好地提升检测精度,但是对于受干扰较大的交通标志例如遮挡、阴影或小目标交通标志等,无法实现更好的特征提取,使得检测精度难以进一步提高。

为了加强模型的特征提取能力,本

文提出了一种基于缓存方案和可变形卷积的 YOLOX 优化算法。首先将骨干网络中普通的卷积模块替换成 DCN 可变形卷积,使得模型在应对不规则形状的目标物体时,能够更加精确地感知和定位;然后使用 EIOU 损失函数替换原有的 GIOU 损失函数,提高收敛速度和回归精度;最后使用多种数据增强的模式来提升模型的性能,并加入 cache 方案提升其检测速度及精度。

1 YOLOX 算法

YOLOX-Darknet53 算法的骨干网络采用了 5 个 ResX 残差组件和 32 倍下采样的特征提取方式,并保留了先验的 YOLOv3 算法的 Darknet-53 骨干网络和 FPN 颈部网络。这个模型设计了 3 个解耦头,采用分类和定位分离的方法来缓解深骨干网络所带来的性能瓶颈。在输入为 $640 \times 640 \times 3$ 的图像下,3 个解耦头输出尺度大小分别为 $20 \times 20 \times 85$ 、 $40 \times 40 \times 85$ 、 $80 \times 80 \times 85$,最终拼接转置输出尺度大小为 85×8400 。此外,该算法还使用了 SPP 子模块和 FPN 子模块实现多尺度空间特征的融合。跟之前的 YOLO 系列算法相比,YOLOX 算法具有更好的性能,能够实现更高的准确率及更快的速度。YOLOX-Darknet53 的网络结构如图 1 所示。从图 1 中可以看出,YOLOX-Darknet53 中引入了大量残差结构,主干网络部分,用卷积层代替 Darknet19 中的最大池化下采样层,输入为 $640 \times 640 \times 3$ 图像通过堆叠的下采样卷积和 n 个残差块,得到 3 张分别为原图像 1/8、1/16、1/32 大小的特征图 (feature map),随后对 3 张特征图进行上采样和下采样,通过解耦检测头后,进行多尺度特征融合,最终输出 85×8400 的图像。

2 YOLOX 算法改进

2.1 backbone 部分改进优化

各算法基本结构如图 2 所示,从图 2 (a) 可以看出,

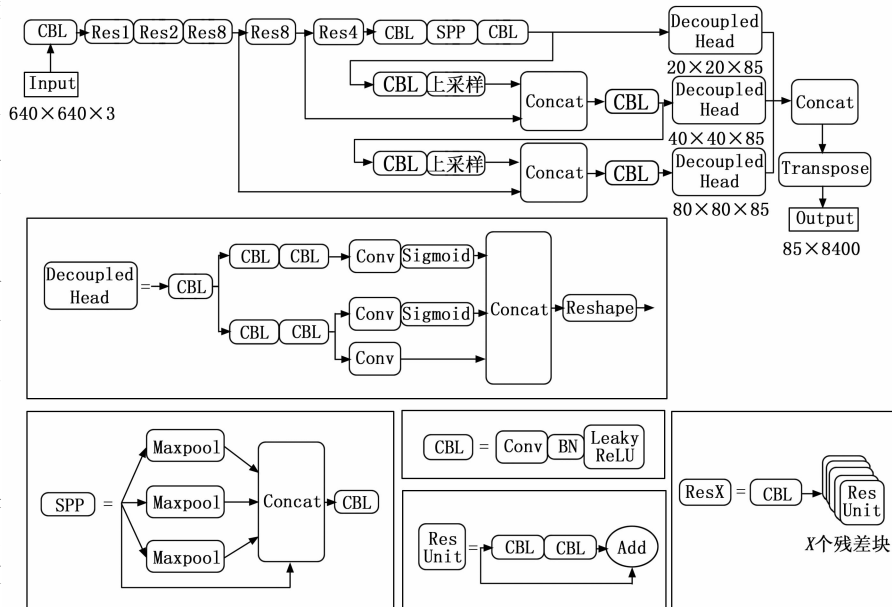


图 1 YOLOX-Darknet53 网络结构图

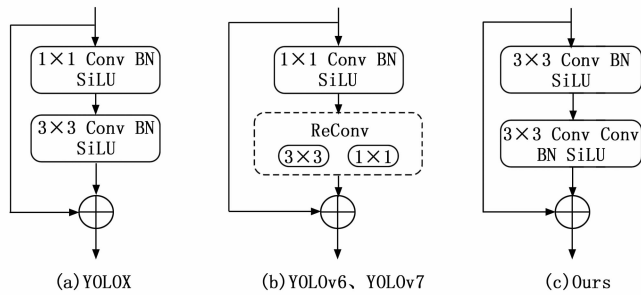


图 2 算法的 Basic Block 对比示意图

YOLOX 中使用的是 1×1 与 3×3 卷积的 Basic Block。从图 2 (b) 可以看出, 在 YOLOv6、YOLOv7 中, 使用的是重参数化 Block, 但重参数化的训练代价高, 且不易量化, 需要其他方式来弥补量化误差。如图 2 (c) 所示, 本文则将 Darknet 中的 1×1 卷积替换成 3×3 卷积, 并将原有模块中的 3×3 卷积替换成了 3×3 可变形卷积, 利用可变形卷积的形变建模能力, 提高网络模型的特征提取能力, 从而更精确地检测目标。改进后的模型整体框架如图 3 所示, 从图 3 中可以看出, 本文的算法改进了 CspLayer 中的 CspBlock 部分, 在 YOLOX 的主干网络中, 残差卷积的主干部分为 1×1 卷积和 3×3 卷积, 改进模型则为 CspBlock 中加入了 3×3 的可变形卷积来增大模型的感受野, 加强模型主干特征提取网络的性能。输入的特征图首先经过 DCNConv (Deformable Convolutional Networks Convolution), 在可变形

卷积中, 每个位置的卷积核都会根据该位置的偏移量进行调整, 以更好地适应目标形状和姿态变化, 然后再通过一些附加的卷积层进行特征提取和压缩, 最终得到预测框的类别和位置信息。在检测交通标志时, 可变形卷积所带来了更大的感受野, 能够使得模型在检测时能够获取到更丰富的关键特征, 对于小目标、遮挡等交通标志的检测能够实现更高的检测精度。

可变形卷积 (Deformable Convolution) 是一种基于空间变形的卷积运算方法, 在卷积神经网络中常被用来处理具有空间变形特征的图像。相较于传统的卷积操作, 可变形卷积具有更强的自适应能力和更好的感受野控制能力, 能够对目标物体的不规则形状进行更加精确的感知和定位。

标准的卷积操作通常可以分为两部分: 1) 在输入的特征图上使用标准固定网格进行采样; 2) 对各个采样点的数值进行加权运算。特征图的标准卷积可以用式 (1) 表示:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n) \quad (1)$$

其中: p_0 为特征图的原始位置; p_n 为包含采样点中所列位置; R 为每个分块的索引编号; $w(p_n)$ 为权重; $x(p_0 + p_n)$ 为原始图。由式 (1) 可知, 在标准卷积中, 每个卷积核所对应的权重是相同的。而可变形卷积则引入了额外的偏移量, 使得卷积核可以基于输入特征图的空间变化进行自适应调整, 因此特征图的映射关系可以用式 (2) 表示:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (2)$$

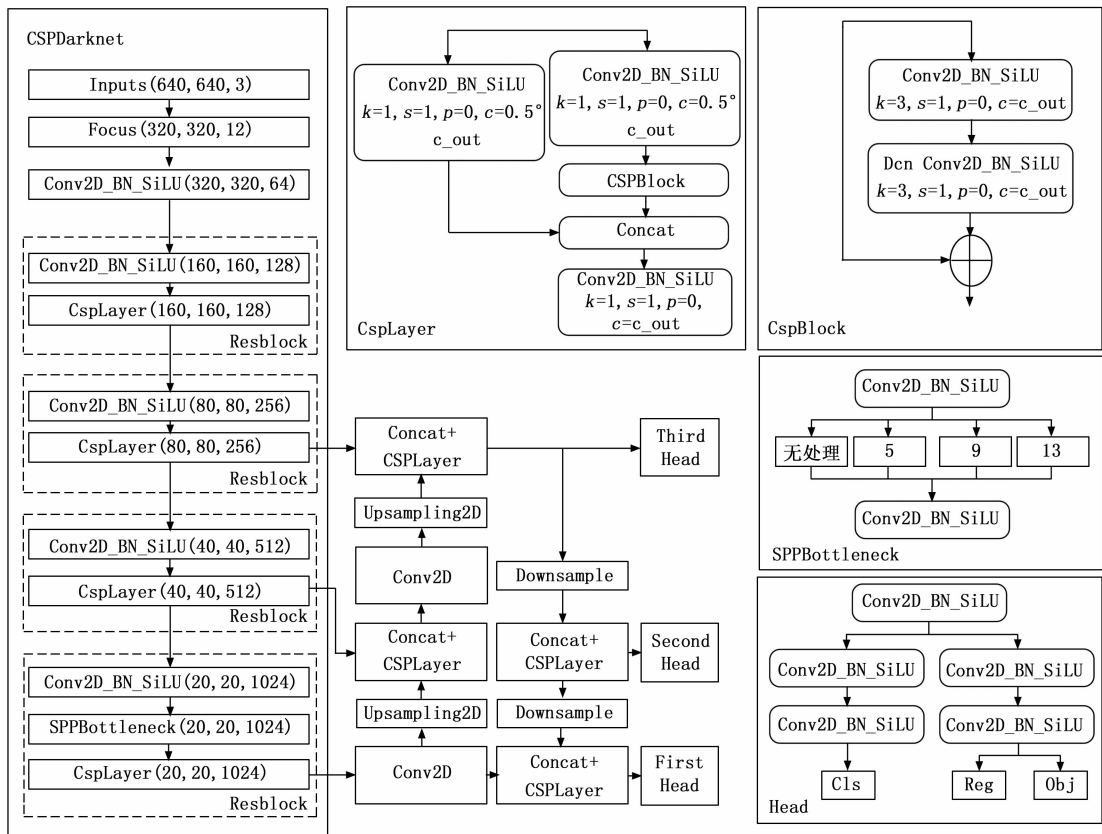


图 3 改进模型整体框架图

其中: Δp_n 为偏移量; $x(p_0 + p_n + \Delta p_n)$ 和 $y(p_0)$ 为原始图以及卷积所得特征图的映射关系, 可变形卷积的操作过程如图 4 所示, 首先使用传统的卷积核提取特征图, 得到特征响应映射 (Feature Response Map)。将上一步得到的特征响应映射作为可变形卷积操作的输入, 对其施加一个新的卷积层, 生成 $2N$ 个通道的偏移量特征图, 其中 N 表示卷积核大小。每两个通道代表着一个坐标偏移量, 即用来描述卷积核在该位置上的位移距离, 这些偏移量可以随着训练进行不断地调整, 以适应不同的输入图像。对偏移量特征图进行双线性插值, 将其上采样至与输入特征图相同大小, 以便后面能够使用偏移量特征图来调整卷积核的位置和大小。最后, 结合上采样后的偏移量特征图和原始输入特征, 在反向传播算法的优化下同时学习它们, 以获得动态调整可变形卷积核的大小和位置的能力。

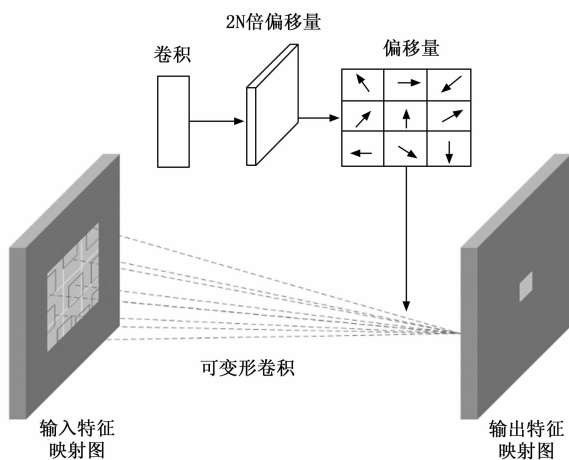


图 4 可变形卷积结构示意图

本文基于可变形卷积构建 basic block, 在基础的 3×3 卷积后, 再使用 3×3 的可变形卷积进行特征提取加强, 增大了其感受野, 从而能够更好地提取特征, 提高检测的准确率。

2.2 损失函数改进

损失函数是一个衡量模型预测值和真实值之间差距的函数, 在目标检测中扮演着非常重要的角色, 选择一个合适的良好的损失函数能够加强模型的性能, 使模型更快地收敛。而 IoU 损失函数则是一种在目标检测任务中常用的评估方法, 它能够很好地反映模型预测框与真实标注框之间的重叠程度。该损失函数可以帮助模型优化预测结果, 使其更加准确地匹配实际标注, IoU 及 IoU 的损失函数可分别用式 (3) 和式 (4) 表示:

$$IoU = \frac{|B \cap B_i|}{|B \cup B_i|} \quad (3)$$

$$L_{IoU} = 1 - \frac{|B \cap B_i|}{|B \cup B_i|} \quad (4)$$

其中: B 表示预测框的面积; B_i 表示真实框的面积。 IoU 损失函数可以将预测框与真实框之间的重叠程度表示为 $0 \sim 1$ 之间的值以评估检测效果。但是在现实情况下, 存

在预测框与真实框不相交的情况, 这样就无法进行有效地学习。此外, IoU 并没有考虑预测框与真实框之间的距离, 无法精确地反映两者之间的重合度情况。因此, 在这些情况下, IoU 损失函数的效果可能会受到影响。

因此 YOLOX 算法使用了 GIoU 损失函数。GIoU 损失函数的公式可用式 (5) 表示:

$$L_{GIoU} = 1 - \frac{|B \cap B_i|}{|B \cup B_i|} + \frac{|C - B \cup B_i|}{|C|} \quad (5)$$

其中: C 为最小外接矩形的面积。GIoU 在计算时, 是以公式中的 $C - B \cup B_i$ 除去两个框的其余面积, 然而在预测框和真实框在相同距离的情况下, 此部分面积最小, 对 Loss 的贡献也就越小, 从而导致在垂直水平方向上回归效果较差, 模型的收敛速度慢。

因此, 本文采用了 EIoU 损失函数来替换 YOLOX 所使用的损失函数。相比于 GIoU 损失函数, EIoU 用分别计算宽高的差异值取代了纵横比, 同时引入 Focal Loss 解决难易样本不平衡的问题。EIoU 损失函数将最小外接矩形替换为最小化两个框中心点的标准化距离, 并分开计算目标框的长和宽。EIoU 损失函数可用式 (6) 表示:

$$L_{EIoU} = 1 - \frac{|B \cap B_i|}{|B \cup B_i|} + \frac{\rho^2(b, b^{gt})}{d^2} + \frac{\rho^2(\omega, \omega^{gt})}{c_\omega^2} + \frac{\rho^2(h, h^{gt})}{c_h^2} \quad (6)$$

其中: b 和 b^{gt} 分别为预测框和真实框的中心点, ρ 为两个中心点的欧氏距离, d 为预测框和真实框最小外接矩形的对角线距离, $\omega, \omega^{gt}, h, h^{gt}$ 分别为预测框和真实框的宽度与长度, c_ω, c_h 为覆盖两个框的最小外接矩形的宽度与长度。EIoU 损失函数考虑了预测框和真实框的重叠部分对结果的影响, 在计算 IoU 时, 使用一个权重因子来平衡两个框的面积, 并将重叠部分的贡献归属到各自的框中。这样可以更准确地评估预测框和真实框的位置关系, 避免了 GIoU 中出现两个框并集等于其中一个框面积的情况, 提升了收敛的速度和回归精度。

为了解决 GIoU 损失函数在目标框水平和垂直方向上误差较大的问题, 并提高收敛速度和回归精度, 本文算法采用 EIoU 损失函数替换原来的损失函数。EIoU 损失函数将目标框的长度和宽度分开计算, 从而避免了 GIoU 损失函数误差较大的问题, 同时提高了收敛速度和回归精度, 更准确地确定两个框的位置关系。在模型训练时, 能够带来更好的收敛效果, 从而提高模型对于交通标志的检测精度。

2.3 数据增强改进

在模型的训练过程中, 为了保证训练的样本充足, 需要对样本做数据增强, 来提高样本的质量和数量, 从而提升模型的泛化能力。在很多目标检测器中, Mosaic+Mixup 的混合数据增强方式被广泛应用, 但是 Mosaic+Mixup 的数据增强方式失真度比较高, 持续使用这种太强的数据增强对模型产生负面影响。于是 YOLOX 中采用的是强弱两阶段的数据训练方式, 但是 YOLOX 中的训练引入了旋转、切片以及翻转, 导致 box 标注产生误差。

于是对于此模块，本文采用多种数据增强的方式来增加模型的性能，主要包括单图数据增强以及混合类数据增强，其中包括交通标志图像色相调整、饱和度调整、明度调整、平移和 Mosaic 以及 MixUp。但是其中取消了旋转、切片以及翻转的数据增强方式，因为部分交通标志经过旋转及翻转等方式后，将会导致其含义改变，如图 5 所示的部分交通标志展示图，列举的两种不同的交通标志在旋转或翻转后，将会导致检测的误差。



图 5 旋转后会改变含义的部分交通标志图

模型的数据增强模块采用强弱两阶段训练，在前 280 个 epoch 中，模型采用了单图数据增强和不带旋转、翻转及剪切的 Mosaic+MixUp 混合数据增强，同时，将每一个训练样本中的混合图像数量由原有的 4 加至 8，以提升强度以及样本的数量。在后 20 个 epoch 中，模型使用了比较小的学习率，关闭了 Mosaic+Mixup 高强度数据增强，同时采用了指数移动平均（EMA），使模型在比较弱的增强下进行微调，通过 EMA 将参数缓慢更新至模型。如果关闭太早则不能发挥 Mosaic 等高强度数据增强模式的效果，关闭太晚则由于之前已经过拟合，此时再关闭没有任何增益，选择在剩余 20 个 epoch 关闭可以避免持续使用原有数据增强所带来的负面影响，同时减少误差，提升准确率。

同时加入 cache 方案，由于 Mosaic&MixUp 涉及多张图片的混合，在每次迭代中需要加载多张图片来进行训练，会引入更多的数据加载成本并减缓训练的速度。如在 YOLOv5 中，每次做 Mosaic 时，4 张图片的信息都需要从硬盘中重新加载。缓存方案则是重新载入当前的一张图片，其余参与混合增强的图片则从缓存队列中获取，通过牺牲一定内存空间的方式提升效率。整体流程如图 6 所示，从图 6 中可以看出，cache 队列中预先储存了 N 张已加载的图像与标签数据，每一个训练步骤只需加载一张新的图片及其标签数据并更新到 cache 队列中，当需要进行混合数据增强时，只需要从 cache 中随机选择需要的图像进行拼接等处理，通过利用缓存，可以将训练流水线中混合图像的时间成本显著降低到处理单个图像的水平，提升了模型的性能。

由于在数据增强阶段中取消了旋转、切片和翻转的数据增强方式，模型在检测交通标志时避免了 box 标注产生的误差，并使得部分交通标志的检测不会出现错误，同时 cache 方案的引入减少了对数据加载的需求，能够提升交通标志检测的速度。

3 实验与分析

3.1 数据集选取

本文的实验采用了我国 Tsinghua-Tencent100K 交通标

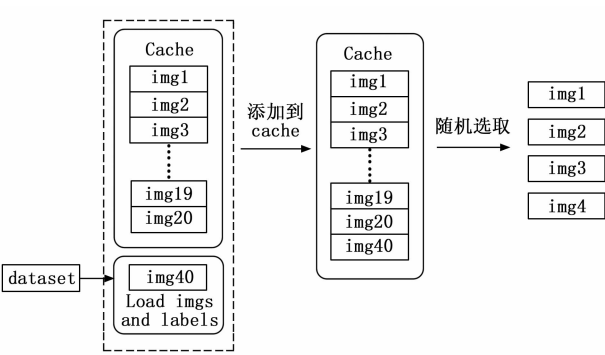


图 6 cache 方案示意图

志数据集，该数据集是腾讯和清华一起合作制作的交通标志数据集，是目前国内唯一基于真实场景的公开交通标志数据集。数据集包含 10 万幅街景图像，经过计算，训练数据集总共包含 16 527 个实例，测试数据集包含 8 190 个实例。这些图片包含了各种不同情形下的交通标志图，例如遮挡、能见度低、小目标等情形。该数据集下的交通标志图所处的环境复杂，面临的挑战较多，能够很好地检验模型的性能，并真实地反映现实中不同情况下的场景。本次实验对于训练集、验证集和测试集按照 8 : 1 : 1 的比例来划分。

3.2 实验平台及结果分析

实验使用一张显存大小为 24 GB 的 NVIDIA GeForce3090 显卡训练和测试所有模型。使用 pytorch1.10.0 框架，Cuda 版本为 11.3，使用 python3.8 版本语言进行编程。训练过程中对网络设定的部分超参数如表 1 所示。其中，batch 为一次训练所抓取的数据样本数量，它的大小影响训练速度和模型优化，设置得合理可以让梯度下降的方向更为准确，这里 batch 的大小设置为 16；weight_decay 为权重衰退，它主要是为了防止过拟合，避免梯度爆炸；max_epochs 为对完整数据进行迭代的次数，一般在 200 以上网络才会收敛；base_lr 为网络的基础学习率，学习率过高可能会导致无法梯度下降，过低则可能会使网络收敛速度缓慢；momentum 为动量，它可以帮助损失函数在训练的过程中更好地达到全局最优的状态。从表中可以看出，基础学习率为 0.001，最大训练代数固定为 300。

表 1 部分超参数设置

参数	值
batch	16
weight_decay	0.01
max_epochs	300
base_lr	0.001
momentum	0.9

图 7 为两个算法模型在训练过程中的损失函数折线图，可以看出本模型的损失函数收敛效果明显优于原算法。模型经过 280 个 epoch 后，进入后 20 个 epoch 的弱增强阶段，进行进一步训练，由图可见由于 YOLOX 算法采用的数据

增强方式，导致前 280 个 epoch 和后 20 个 epoch 的训练差距较大，图中 A 点为 280 epoch 点，B 点为 290 epoch 点，可以看出两点之间的曲率有一定的差距，说明模型的收敛效果较差，验证了本文之前数据增强模块的理论分析内容，而本文算法的收敛效果整体平稳。训练结束，最终 YOLOX 算法模型总损失在 2 左右振荡稳定，而本模型最终则稳定在 0.3 左右。

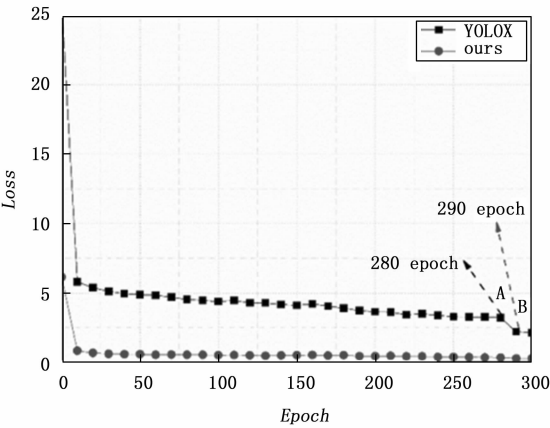


图 7 Loss 曲线图

在模型训练完成后，对比两个模型的 mAP 值，如图 8 为两个模型的 mAP 值随 epoch 变化的折线图， mAP 表示所有检测类别平均准确率的均值，可以用公式 (7) 表示：

$$mAP = \frac{\sum AP}{N} \tag{7}$$

其中： $\sum AP$ 为各类别的平均准确率的总和， N 为类别个数。作为模型训练性能的指标， mAP 值越高，代表模型的性能越好，检测精度越高。由图 8 中可以看出，YOLOX 算法在训练至 280 epoch 时其 mAP 值甚至有下降趋势，在后 20 个 epoch 的弱数据增强的作用下才进一步提升，图中 C 点为 280 epoch 点，图中 D 点为 290 epoch 点，可以看出两点之间准确率有一定的差距，验证了本文之前数据增强模块的理论分析内容，而本模型在强弱两阶段训练过程中， mAP 值稳定提升。最终经过 300 个 epoch，本文模型训练所得的 mAP 值达到了 67.2%，而 YOLOX 算法在同样的条件下 mAP 值则为 60.8%。相较于原算法，本算法模型的 mAP 值提升了 6.4%，由此可以看出该算法对于网络模型检测精度的提升是有效果的。

为了进一步验证改进算法的有效性，本文在相同的配置与参数下对加入的模块进行了消融实验，消融实验的结果对比如表 2 所示。从表 2 中可以看出，加入的 3 个模块相比较于原算法而言， mAP 值都有一定程度的提升，其中加入可变形卷积模块时，提升了 2.8 个点，这是因为加入的 DCNConv 有着很强的感受野控制能力，加强了模型的特征提取，原有的不规则交通标志例如残缺交通标志等，在 DCNConv 的作用下能够更准确地识别出来。另外，在加入 cache 方案下的数据增强模块时， mAP 又提升了 3.1 个点，

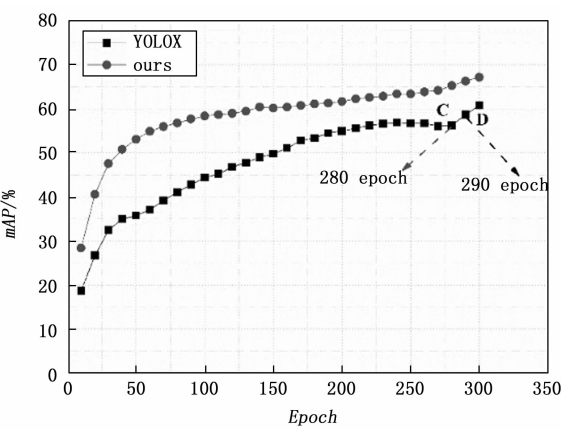


图 8 mAP 折线图

主要是 cache 方案减少了对数据加载的需求，使得模型收敛得更好，训练效率大大提升。同时改进的算法在数据增强模块中取消了裁剪及旋转等增强方式，避免了框注释和输入之间的错位，从而避免部分交通标志的检测出现错误，提升了检测的精度。

表 2 消融实验结果对比

变量	$mAP/\%$
Baseline	60.8
+DCN	63.6(+2.8)
+DCN+EIou	64.1(+3.3)
+DCN+EIou+Cached Mosaic& Mixup	67.2(+6.4)

同时将本文算法与目前主流的 YOLO 系列算法进行对比实验，实验对比结果如表 3 所示。由表 3 可以看出，在相同的配置环境下，改进的算法的 mAP 值要明显优于其他 YOLO 系列的算法。与同系列的 YOLOv3 与 YOLOv5 相比， mAP_{50} ：95 值分别提升了 10.9、6.9 个点，与原算法相比，改进的算法提升了 6.4 个点。本文算法的推理速度为 33.4，能够良好地满足实际检测需求。综上所述，与其他主流的 YOLO 算法相比，改进的 YOLOX 算法在能够满足良好的实时性的同时，有着更高的检测精度。

表 3 不同算法实验结果对比

算法名称	$mAP_{50}/\%$	$mAP_{50:95}/\%$	FPS
YOLOv3	77.2	56.3	29.2
YOLOv5	81.1	60.3	40.3
YOLOX	81.8	60.8	37.7
本文算法	86.1	67.2	33.4

最后为了直观地看出所改进后模型的网络性能，以及在不同情形下的检测效果，分别将原模型与改进后的模型在 4 种不同的情形下进行检测，分别是无遮挡大目标、遮挡大目标、密集小目标以及遮挡小目标环境下进行检测。检测框中的数字为置信度，即该样本为正样本的概率，它与准确率成正比，置信度越高越好。其环境中的交通标志都是数据集中实际拍摄所得，能够真实反映现实情况。

图 9 为检测无遮挡大目标时的效果图, 由图可以看出, 在检测无遮挡大目标交通标志 pm20 时, YOLOX 算法和本文算法都实现了成功检测, 但是 YOLOX 算法的检测置信度为 68.9%, 而本文算法的检测置信度为 95.8%, 远高于原算法。

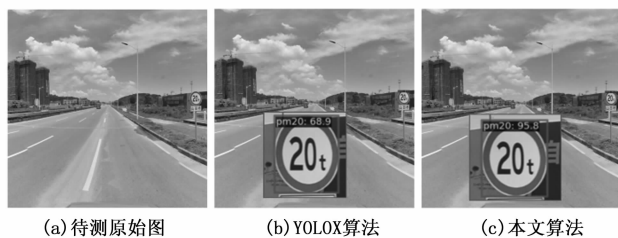


图 9 检测无遮挡大目标

图 10 为检测遮挡大目标时的效果图, 由图可以看出, 在检测出现部分遮挡的大目标交通标志时, YOLOX 算法由于其数据增强引入了旋转和切片等方式, 导致预测框出现了重叠、错乱的情况, 未能成功标注出两个交通标志, 其中交通标志 w55 的置信度为 34.9%, pl30 的置信度为 42.4%。而在本文算法模型中, 预测框与图像中交通标志的重合度较高, 两个交通标志都被正常地识别出来, 其中, 交通标志 w55 的置信度为 93%, pl30 的置信度为 45.9%。从检测结果中可以看出, 无论是检测精度还是预测框的重合度, 本文算法都要优于 YOLOX 算法。

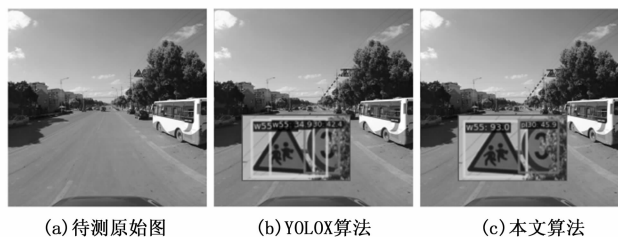


图 10 检测遮挡大目标

图 11 为检测密集小目标的效果图, 由图可以看出, 在检测密集的小目标时, 两个算法的置信度基本相同, 但是 YOLOX 算法出现了错检, 将交通标志 p12 错检成了 p26, 其置信度为 32.8%。而本文算法则成功地检测出了交通标志 p12, 其置信度为 33.6%。可以看出本文算法在检测密集的小目标时, 也有较好的效果。

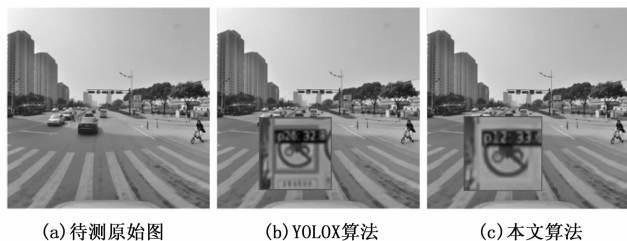


图 11 检测密集小目标

YOLOX 算法在检测遮挡的小目标时, 出现了漏检的情况, 未能检测出交通标志 p23, 本文算法则是成功检测出了遮挡的小目标交通标志 p23, 其置信度为 21.1%。从检测对比图中可以看出, 本文的算法模型相对于 YOLOX 算法而言, 具有更好的环境适配性, 能够较好地应对不同的环境。

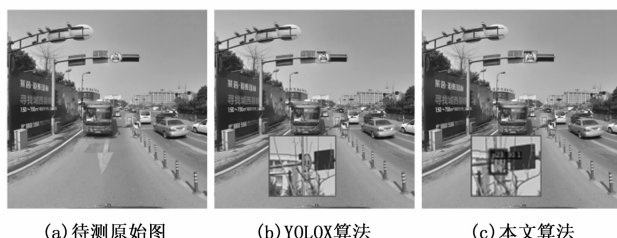


图 12 检测遮挡小目标

4 结束语

为了提高模型对交通标志的检测精度, 本文提出了一种基于 Cache 方案和 DCN 可变形卷积的 YOLOX 优化算法。在 YOLOX 算法的基础上, 进行了三点改进: 首先将 backbone 的基本框架中的普通卷积替换成 3×3 普通卷积 + 3×3 DCN 可变形卷积, 提高感受野, 增强模型的特征提取能力; 然后将原有的 GIOU 损失函数替换成 EIOU 损失函数, 解决了 GIOU 损失函数误差较大的问题, 并且提高了收敛的速度和回归的精度。最后采用多种数据增强的方式来增加模型的性能, 且加入 Cache 方案, 通过缓存队列提升检测效率, 并进一步提升其检测精度。实验证明本文算法能够在多种应用场景下更好地检测识别交通标志。

参考文献:

- [1] NAVNEET D, BILL T. Histograms of oriented gradients for human detection [C] // Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C. USA: IEEE Press, 2005: 4-8.
- [2] DAVID G. Object recognition from local scale-invariant features [C] // Proceedings of International Conference on Computer Vision. New York, USA: ACM Press, 1999: 1150-1157.
- [3] PAPAGEORGIOU C P, OREN M, POGGIO T. A general framework for object detection [C] // Proceedings of the 6th Computer Vision, Washington D. C., USA: IEEE Press, 1998: 555-562.
- [4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 779-788.
- [5] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 6517-6525.

图 12 为检测遮挡小目标的效果图, 由图可以看出,

(下转第 84 页)