DOI:10.19651/j.cnki.emt.2415632

基于改进 YOLOX-tiny 算法的交警手势识别

方吴逸　陈章进　唐英杰
(上海大学微电子研究与开发中心 上海 200444)

摘　要：为了在城市中实现无人驾驶,需要能够高效检测交警的现场指挥手势。针对现有手势识别算法识别精度低、检测速度慢、难以应对复杂道路环境等问题,提出一种改进的 YOLOX-tiny 交警手势识别算法。首先,使用改进后的 GhostNet 网络替换原主干网络,并且插入坐标注意力机制,全面提取输入图像特征,提高了网络的检测精度,同时提升了对中小型目标的检测效果;其次,改进解耦头部分,设计了 SCDE Head 结构,在减少计算量的同时过滤冗余信息,使得解耦头更有效率,并且解耦头融合了多尺度的特征,提升了目标检测准确率;最后,将 SIoU 应用到定位损失中,加快网络收敛的速度,提升回归精度。在自制的交警指挥手势数据集上进行测试,实验结果表明,与 YOLOX-tiny 模型对比,改进后算法参数量减少了 27.97%,模型计算量减少了 33.31%,且平均检测精度提高了 2.31%,检测速度提升了 45%,更适合汽车无人驾驶以及硬件部署方面的实际需求。

关键词：交警手势识别;YOLOX-tiny;网络轻量化;GhostNet;注意力机制
中图分类号：TP391.4;TN791　　**文献标识码**：A　　**国家标准学科分类代码**：520.60

Traffic police gesture recognition based on improved YOLOX-tiny algorithm

Fang Wuyi　Chen Zhangjin　Tang Yingjie
(Microelectronics Research and Development Center, Shanghai University, Shanghai 200444, China)

Abstract: In order to achieve autonomous driving in cities, it is necessary to be able to efficiently detect the on-site command gestures of traffic police. Aiming at the problems of low recognition accuracy, slow detection speed, and difficulty in dealing with complex road environments in existing gesture recognition algorithms, an improved YOLOX-tiny traffic police gesture recognition algorithm is proposed. Firstly, an improved GhostNet network was used to replace the original backbone network, and a Coordinate Attention mechanism was inserted to comprehensively extract input image features, improving the detection accuracy of the network and enhancing the detection performance for small and medium-sized targets; Secondly, the decoupling head was improved by designing the SCDE Head structure, which reduces computational complexity while filtering redundant information, making the decoupling head more efficient. The decoupling head also integrates multi-scale features, improving the accuracy of object detection; Finally, applying SIoU to localization loss accelerates network convergence and improves regression accuracy. Tested on a self-made traffic police command gesture dataset, the experimental results showed that compared with the YOLOX-tiny model, the improved algorithm reduced the number of parameters by 27.97%, the model's computational complexity by 33.31%, and the average detection accuracy increased by 2.31%, with a 45% increase in detection speed, which is more suitable for the practical needs of autonomous driving and hardware deployment.

Keywords: traffic police gesture recognition; YOLOX-tiny; network lightweight; GhostNet; attention mechanism

0　引　　言

随着经济发展,城市中机动车数量激增,这导致车辆碰撞事故频发[1]。虽然道路上已经设置了交通标志及交通信号灯,但是仍然需要交警在路口指挥交通。无人驾驶技术为了应对复杂的道路交通情况,除了要求能够快速检测信号灯、交通标志以外,对于现场交警指挥也必须能快速识别[2]。然而,道路目标检测有着环境复杂、车流量大导致遮挡、人物距离远目标小等特点,这对于无人驾驶算法的性能有巨大的挑战。此外,对于汽车这种计算资源有限的边缘设备来说,模型轻量化、硬件部署方面的研究也至关重要。

目前基于深度学习的交警指挥手势目标检测方法可以

收稿日期:2024-03-14

分为以下两种:一种是基于动态手势识别,另一种是基于静态手势识别。动态手势的识别通过划分每种手势的起始帧,输入到神经网络中学习训练[3]。张丞等[4]使用基于人体铰链特征建立交警手势的时空图模型,并提出一种 GCN 分区策略,解决了分区策略对图结构造成的限制,最后设计了一种 GCN 输出架构,以适配多对多标签序列预测模式。马天祥[5]利用 YOLOv5s 区分交警和行人,然后使用 Mediapipe 提取交警的身体关键点,以得到交警手势的时间序列模板,最后使用 FastDTW 进行特征的匹配,实现对动态的交警手势的识别。程贝芝等[6]使用时空图卷积网络识别交警指挥手势,使用 Transformer 对视频时间序列进行编码,获取全局信息从而提高交警手势识别的准确率。但是,基于动态手势识别方法的模型较大,准确率偏低,检测速度慢,不适用于无人驾驶的应用场景。

因此,一些研究人员开始采用静态手势识别的方法来进行目标检测。目前,广泛应用的静态目标检测算法可分为 One-Stage 检测算法和 Two-Stage 检测算法。Two-Stage 检测算法先生成候选框,然后对候选框区域进行分类,代表算法有 R-CNN[7]、Fast R-CNN[8] 和 Mask R-CNN[9]等。而 One-Stage 检测算法在一个网络中进行训练和检测,具有较快的检测速度,代表算法有 YOLO 系列[10-11]和 SSD[12]等。李超军[13]提出了一种结合 LK 光流法和迁移学习的交警手势识别方法。采用优化的 LK 光流法检测交警上肢的运动区域,再从运动块轨迹中提取关键手势信息并运用迁移学习对关键手势信息进行分类识别。李晓杰[14]提出了一种基于 CNN 的交警指挥手势分类识别方法,使用 AlexNet 和 GoogLeNet 为基础设计了 CNN-A 和 CNN-G 网络用于特征提取,并最终采用 Softmax 分类器对交警手势进行分类识别。康观龙[15]使用 Resnet18 网络对交警手势图片进行迁移学习,并用胶囊网络提升了在小样本情况下的准确率。王新等[16]通过在 YOLOv5 算法的基础上添加了自校准卷积、置换注意力模块、广义焦点损失函数,提高了检测精度。

总体来说,目前对于静态手势识别方法在交警指挥手势识别中的研究还较少,并且对于计算资源有限的边缘设备来说,未进行算法轻量化及模型部署的研究。

针对此现状,本文提出了一种基于 YOLOX-tiny 网络的轻量化交警指挥手势快速检测方法,取名为 GSCA-YOLOX 模型。本文主要工作如下:

1)在 GhostNet 中引入坐标注意力机制(coordinate attention, CA)设计了 GhostNet-CA 网络,并使用 GhostNet-CA 网络替换了 CSPDarknet 网络结构,使得网络轻量化,加强了主干网络对输入图像特征的全面提取能力,提升对中小型目标的检测效果;

2)引入空间通道重构卷积和全局特征调节模块,设计了 SCDE Head(spatial and channel reconstruction decoupling head, SCDE Head),使得解耦头更有效率且减少了计算量;

3)将损失函数(intersection over union, IoU)替换为具有角度代价的 SIoU,促进真实框和检测框的拟合,提高收敛速度和精度。

1 YOLOX 算法简介

YOLOX 算法[17]是旷视科技于 2021 年提出的单阶段目标检测算法。在 YOLOv3 的基础上,融合了输入端的数据增强、解耦头等目标检测领域的前沿成果,实现了检测精度和推理速度的巨大提升。YOLOX 网络由 Input、Backbone、Neck、Prediction 4 部分组成,网络结构如图 1 所示。

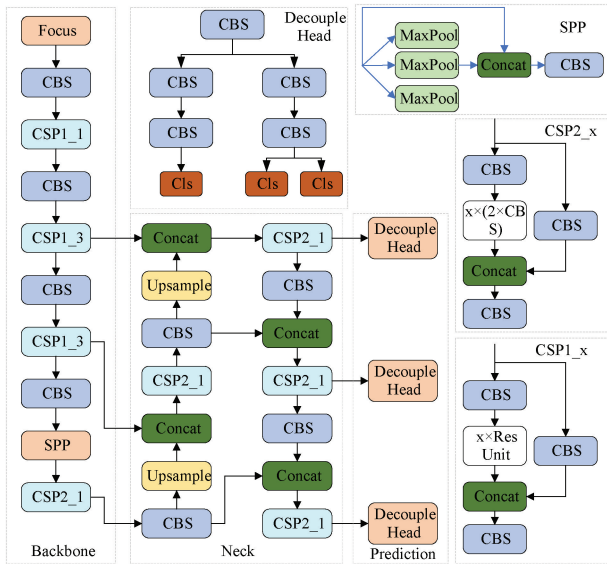


图 1 YOLOX 网路结构图

Input:在网络的输入端,YOLOX 主要采用了 Mosaic、Mixup 两种数据增强方法。Mosaic 主要思想是将 4 张图片进行拼接到一张图上作为训练样本。Mixup 主要思想是将两张图片和其标签,按权重进行叠加,生成新的数据集和其所对应的标签。这两种数据增强提升了模型对小目标的检测能力。

Backbone:YOLOX 的主干网络采用了 CSPDarknet 网络结构。图片输入后首先进入 Focus 结构,通过将图片每隔一个像素点取出一个值,得到 4 个特征层,再将这 4 个特征层堆叠,此时宽高信息就集中到了通道信息,通道数扩充 4 倍。CSPDarknet 网络结构大量使用了残差网络,特点是容易优化,可以通过模型深度的增加来提高检测精度,并且其内部的残差块使用了跳跃连接,缓解了由于增加深度带来的梯度消失问题。

Neck:YOLOX 的颈部网络由特征金字塔网络(feature pyramid networks, FPN)与路径聚合网络(path aggregation networks, PAN)组成。FPN 和 PAN 形成的 PAFPN,强调不同层次特征的双向交流和融合,提升了多

尺度目标的检测性能。此设计提升了对小目标的捕捉能力，减少了小目标的误检及漏检。

Prediction：在预测方面，以往的 YOLO 系列模型将分类和回归任务都在卷积中实现，使得预测主体聚集在一起。然而，在 YOLOX 中，为了提高结果表现力，分类和回归任务被分开处理，并通过解耦头(decoupled detection head)进行并行处理。同时采用了 Anchor Free 方法进行目标框标注，既减少了模型参数量，又提高了模型的识别准确率。此外，引入了 SimOTA 方法来动态匹配正样本和负样本，以保证模型的检测精度，并加快训练时间。

2　GSCA-YOLOX 网络结构设计

由于道路目标检测有着环境复杂、车流量大导致遮挡，人物距离远目标小等特点，为了使得 YOLOX 算法更加契合交警指挥手势的目标检测任务，本文对 YOLOX-tiny 做出改进，分别是使用坐标注意力机制改进 GhostNet 并用其替换主干网络、设计 SCDE Head 解耦头来提升识别的精确度、使用 SIoU 改进损失函数来优化预测结果。改进后的 GSCA-YOLOX 结构如图 2 所示。

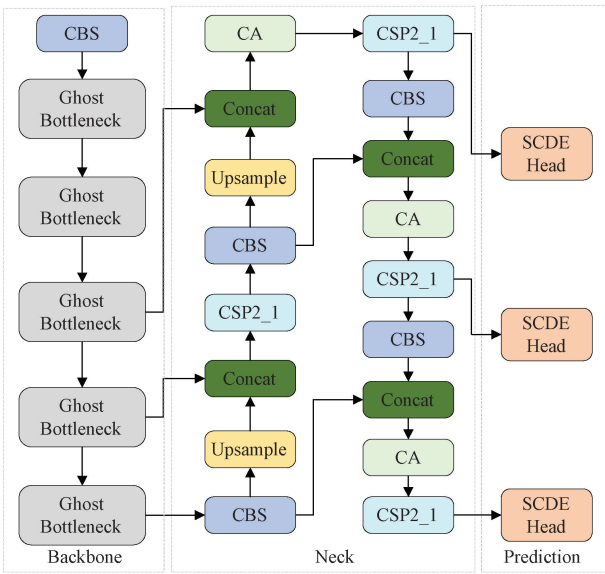


图 2　改进后的 GSCA-YOLOX 网络结构

2.1　改进的 GhostNet 主干网络

YOLOX 主干网络 CSPDarknet53 采用了大量的 3×3 卷积核来进行特征提取，过程中产生的相似特征图通常被当作冗余信息。冗余特征图虽然能使网络更具泛化性，但同时也带来了更大的数据量并降低算法效率。为了加快主干网络特征提取性能以及减少模型的参数和计算量，本文使用改进的 GhostNet[18] 网络替换了原始的 CSPDarknet 网络。

GhostNet 中提出使用 Ghost Module 替换传统卷积。Ghost Module 将标准的卷积运算分为两部分，首先使用 1×1 卷积生成通道数为输入特征一半的内在特征图，然后通过廉价操作生成 Ghost 特征图，Ghost Module 输出的结果是内在特征图和 Ghost 特征图的融合。Ghost Module 的操作流程如图 3 所示，其中，Φ_n 代表廉价操作。

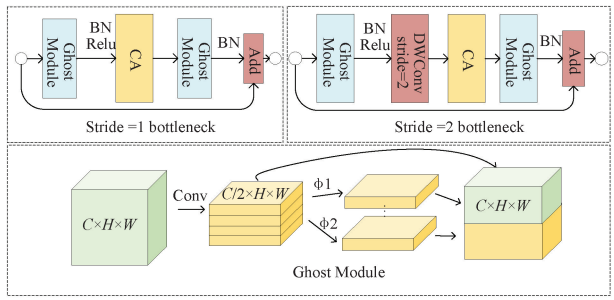


图 3　Ghost 模块结构图

假设输入特征图大小为 $H\times W$，输入通道数为 C，输出特征图大小为 $H'\times W'$，输出通道数为 $m\times s$，卷积核大小为 $k\times k$。那么，传统的卷积计算如下：

$$(m\times s)\times H'\times W'\times C\times k\times k \tag{1}$$

为了保持 Ghost Module 输出的特征图数量与传统卷积一致，Ghost Module 中卷积核大小、步幅和填充需要与传统卷积相同。假设 Ghost Module 的第 1 部分的卷积生成 m 个特征图，每个特征图生成 $s-1$ 个新的特征图，第 2 部分得到 $m\times(s-1)$ 个特征图。因此，Ghost Module 的总计算量为：

$$m\times H'\times W'\times C\times k\times k+m\times(s-1)\times H'\times W'\times k\times k \tag{2}$$

式中：$H'\times W'$ 为输出特征图的大小；C 为输入通道数；$k\times k$ 为卷积核大小。

Ghost Module 与传统卷积的理论参数压缩比为：

$$\frac{(m\times s)\times H'\times W'\times C\times k\times k}{m\times H'\times W'\times C\times k\times k+m\times(s-1)\times H'\times W'\times K\times K}=\frac{C\times s}{C\times s-1} \tag{3}$$

通常，C 远大于 s，因此：

$$\frac{C\times s}{C\times s-1}\approx s \tag{4}$$

从式(4)可以看出，使用 Ghost Module 进行卷积操作可以显著减少卷积过程中的参数量和计算量。

Ghost Bottleneck 旨在调用 Ghost Module，如图 3 所示。Ghost Bottleneck 主要由两个堆叠的 Ghost Module 组成。其中第 1 个 Ghost Module 用作增加特征维度的扩展层；第 2 个 Ghost Module 用于减少特征维度，再添加由深度可分离卷积层及普通卷积层构成的下采样 shortcut 旁路。

整个 GhostNet 结构由一个深度可分离卷积块和多个 Ghost Bottlenecks 组成，根据 Ghost Bottleneck 的数量可分为 5 个阶段。每个阶段用“GBMx”表示。“x”表示 Ghost Bottleneck 的数量，每个阶段的特征图的输出大小如图 4 所示。最终输出的是不同大小的特征层 P3(80×80×40)、

P4(40×40×112)和 P5(20×20×160),这 3 个特征图被输入到后续的特征融合阶段。

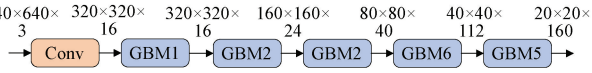


图 4 GhostNet 结构图

但是 GhostNet 在对特征信息建模方面有局限性。为了解决这个问题,原始 GhostNet 结构中引入了通道注意力模块(squeeze-and-excitaion, SE)[19]进行通道关系建模,使得模型可以收集全局信息。然而,SE 模块只考虑了通道间信息的编码,而忽略了位置信息的重要性。此外,SE 模块中的全连接层具有大量参数,不符合轻量化的目标。而 CBAM 模块[20]通过减少通道数量,然后使用大尺寸卷积来获取位置信息,从而获取局部相关信息,但缺乏提取远距离关系的能力。

针对此问题,本文引入了坐标注意力机制替换了原本的 SE 模块。CA 模块[21]将位置信息嵌入到通道注意力中,不仅可以捕获跨通道信息,还可以获得关于方向和位置的感知信息,这有助于模型更准确地识别和定位目标。CA 模块结构如图 5 所示。

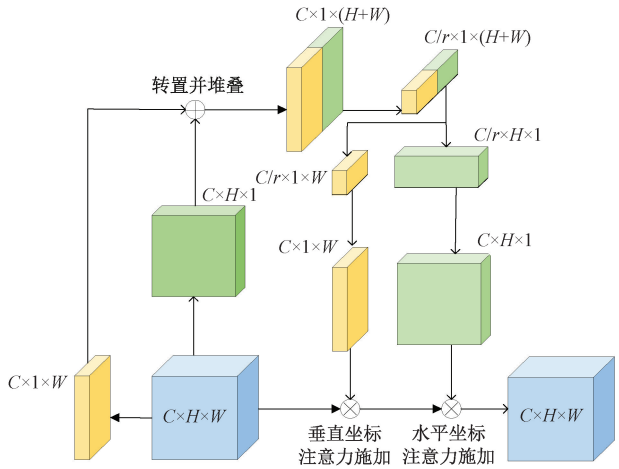


图 5 坐标注意力机制结构图

CA 模块首先对输入特征图的每个通道在水平和垂直方向上分别使用全局平均池化进行编码。对于第 c 个通道,维度 h 和 w 的垂直和水平方向的输出表示为:

$$Zc^{h}(h)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}Xc(h,i) \tag{5}$$

$$Zc^{h}(h)=\frac{1}{H}\sum_{0\leqslant i\leqslant H}Xc(j,w) \tag{6}$$

两个变换沿着两个空间方向进行特征聚合,不仅能让注意力模块捕捉到沿着一个空间方向的长距离依赖,并保存着另一个空间方向的精确位置信息,有助于网络更准确地定位感兴趣的目标。然后级联这两个方向的特征图,然后使用一个共享的 1×1 卷积进行卷积变换具体如下:

$$f=\delta(F_1([Z^h,Z^w])) \tag{7}$$

式中:[.,.]表示连接操作;δ 表示批量归一化层和非线性激活函数;F_1 表示用于通道缩减的 1×1 卷积。f 是一个维度为 $R^{\frac{c}{r}\times(H+W)}$ 的中间特征图,其中 r 表示通道缩减率。

随后,f 在空间上被分成两个特征图,$f\in R^{\frac{c}{r}\times H}$ 和 $f\in R^{\frac{c}{r}\times W}$。这些特征图用于通过 1×1 卷积恢复通道数,然后使用 Sigmoid 函数进行归一化以创建两个空间方向的注意力图,如下所示:

$$g^{h}=\sigma(F^{h}(f^{h})) \tag{8}$$

$$g^{w}=\sigma(F^{w}(f^{w})) \tag{9}$$

式中:σ 表示 Sigmoid 函数。

最后,将输入特征图与获得的注意力相乘,得到 CA 模块的最终输出如下:

$$y_c(i,j)=x_c(i,j)\times g_c{}^{h}(i)\times g_c{}^{w}(j) \tag{10}$$

改进后的 GhostNet 主干网络结构相较于原模型在参数量上有较大提升,同时引入 CA 模块保证了对输入图像特征的全面提取,提高了网络的检测精度,同时提升了特征提取网络对中小型目标的检测效果。

2.2 改进解耦头

尽管 YOLOX 中引入的解耦头可以提高模型精度和收敛速度,但在交警手势识别的过程中,可能会出现目标互相遮挡的情况,因此,需要准确的定位信息来确定各个目标的位置。为此设计了一种改进的空间通道重构解耦检测头 SCDE Head 来进行目标检测。SCDE Head 可以生成包含更多语义信息的特征编码,从而为分类任务提供帮助。同时,它还可以生成包含了更多边缘信息的高分辨率特征图,从而提高定位任务的准确率。改进的 SCDE Head 结构如图 6 所示。

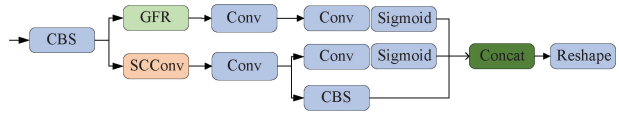


图 6 SCDE Head 模块原理图

解耦头的输出包含分类分支和定位分支两部分。在分类分支中,设计了全局特征调节(global feature regulation, GFR)模块。该模块是为了使用深度特征调制浅层特征而设计的,能够捕获关键的局部图像区域。如图 7 所示,分类任务的起始阶段中,GFR 利用来自两个不同尺度 P_l 和 P_{l+1} 的特征图,随后对深度特征进行上采样,然后进行 1×1 卷积以匹配 P_{l+1} 的通道计数。然后将这些与 P_{l+1} 连接以产生最终 $G_l{}^{cls}$。

$$G_l{}^{cls}=\mathrm{Concat}(\mathrm{Conv}(\mathrm{Upsample}(p_l),p_{l+1})) \tag{11}$$

在定位分支中,使用 SCConv[22](spatial and channel reconstruction convolution, SCConv)替换了原本的 3×3 卷积。SCConv 可以利用特征图之间的空间和通道冗余来进行 CNN 压缩,以减少广泛存在于标准卷积中的空间和

通道冗余并促进代表性特征的学习以实现定位信息的精炼。SCConv 将输入的特征先经过空间重构单元(space reconstruction unit，SRU)，得到空间细化的特征。再经过通道重构单元(channel reconstruction unit，CRU)，得到通道提炼的输出特征。SCConv 的结构如图 8 所示。

输入特征图 X 首先进入 SRU 进行空间重构。对于输入特征图 X，首先进行组归一化，通过归一化层中的尺度缩放因子 γ 评估不同特征图的空间信息量，进而得到特征图的重要性权重 W_γ，过程如式(12)所示。

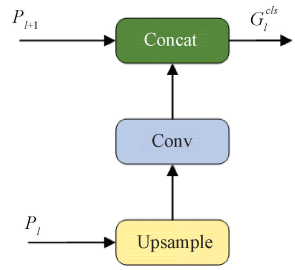


图 7　GFR 模块结构图

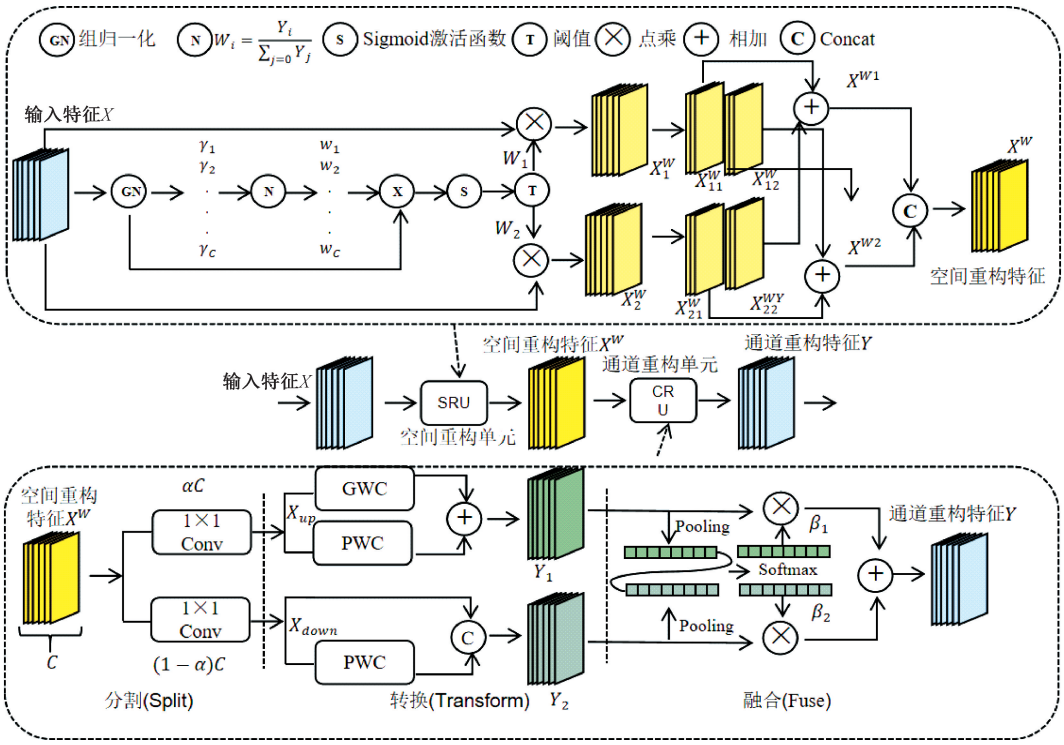


图 8　SCConv 结构图

$$X_{out}=GN(X)=\gamma\frac{X-\mu}{\sqrt{\sigma^2+\varepsilon}}+\beta \tag{12}$$

式中：GN 代表组归一化层；μ 和 σ 分别是输入特征图 X 中的均值和标准差；γ 和 β 是可训练的仿射变换参数；ε 是一个为了稳定除法而添加的小的正数。

接着通过 Sigmoid 函数将权重 W_γ 映射到范围(0，1)，取 0.5 为阈值，将权重 W_γ 分解成重要信息权重 W_1 和冗余信息权重 W_2，其中 W_1 有丰富的空间信息，W_2 只有很少的信息。

随后，将输入特征 X 分别乘以 W_1 和 W_2，得到信息丰富的特征 $X_1{}^W$ 和信息较少的特征 $X_2{}^W$。为了进一步减少空间冗余，采用交叉相加的方式充分结合两个不同的信息的特征，以生成更丰富的信息特征并节省空间。最终，将交叉重建后的特征 $X_1{}^W$ 和 $X_2{}^W$ 连接起来以获得空间重构特征图 X^W。

$$\begin{cases}X_1^W=[X_{11}^W,X_{12}^W]\\X_2^W=[X_{21}^W,X_{22}^W]\\X_{11}^W\oplus X_{22}^W=X^{W1}\\X_{21}^W\oplus X_{12}^W=X^{W2}\\X^{W1}\cup X^{W2}=X^W\end{cases} \tag{13}$$

式中：[·，·]表示沿通道拆分操作；⊕表示逐元素相加操作；∪表示通道拼接操作。

空间重构完后，继续对中间特征图 X^W 进行通道维度的重构。首先按照分裂率 α 将中间特征图分成两部分，分别是 αC 和 $(1-\alpha)C$ 个通道，再通过 1×1 卷积压缩特征图的通道数，提高计算的效率。此操作后，中间特征图 X^W 被分成了上半部分特征图 X_{up} 和下半部分特征图 X_{down}。

X_{up} 作为丰富的特征提取器，采用高效的逐分组卷积(GWC)和逐点卷积(PWC)来代替昂贵的标准 k×k 卷积来提取高级具有代表性的特征信息并降低计算成本。随后将输出相加合并成一个具有代表性的丰富特征图 Y_2。

X_{down} 作为特征信息补充,采用轻量的 1×1PWC 操作来生成具有细节信息的特征图,将生成和复用的特征沿通道拼接起来,形成下层转换阶段的输出 Y_1。

$$Y_1 = M^G X_{up} + M^{P1} X_{UP} \tag{14}$$

$$Y_2 = M^{P2} X_{down} \cup X_{down} \tag{15}$$

式中:M^G、M^{P_1} 分别是 GWC 和 PWC 的可学习权重;Y_1 是上半部分输出的特征图;M^{P_2} 是 PWC 的可学习权重;∪表示 Concat;Y_2 是下半部分输出的特征图。

随后,利用简化的 SKNet 方法自适应地合并上转换阶段和下转换阶段的输出特征 Y_1 和 Y_2[23]。运用全局平均池化来获得具有全局空间信息的通道描述符 S_1、S_2。随后,将上下分支的通道描述符 S_1、S_2 堆叠在一起,并使用通道软注意力操作生成重要性特征向量 β_1、β_2。

$$S_m = Pooling(Y_m) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} Y_c(i,j), m = 1,2 \tag{16}$$

$$\boldsymbol{\beta}_1 = \frac{e^{s_1}}{e^{s_1} + e^{s_2}}, \boldsymbol{\beta}_2 = \frac{e^{s_2}}{e^{s_1} + e^{s_2}}, \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 = 1 \tag{17}$$

在重要性特征向量 β_1、β_2 的指导下,合并上层特征 Y_1 和下层特征 Y_2,可以得到通道精细化的特征 Y。

$$Y = \boldsymbol{\beta}_1 Y_1 + \boldsymbol{\beta}_2 Y_2 \tag{18}$$

本文使用文献[22]实验中的超参数值对 SCConv 模块进行参数量和标准卷积进行比较。比较结果是 SCConv 模块可以将网络模型参数量压缩 5 倍,并且具有比标准卷积更优秀的性能。

2.3 改进损失函数

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,在目标检测中起着非常重要的作用,选择一个优秀的损失函数能够加强模型的性能,使模型更快更好地收敛。IoU 损失函数[24]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 和 IoU 的损失函数 L_{loss} 可以分别用式(19)和(20)表示。

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{19}$$

$$L_{loss} = 1 - IoU \tag{20}$$

式中:A 代表预测框的面积;B 代表真实框的面积。

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,本文使用 SIoU 损失函数[25]代替 IoU 损失函数,SIoU 损失函数考虑了角度、距离和形状 3 部分。

首先,引入角度变量 α 以矫正偏差的预测框角度,α 为真实框与预测框中心点连线与 x 轴的夹角。预测框沿着 x 轴方向逼近真实框,如果 $\alpha \geqslant \frac{\pi}{4}$,则沿另一方向继续逼近。

其次,考虑真实框与预测框的距离,引入距离相关变量 Δ 以解决预测框与真实框的位置偏移问题,Δ 定义为:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma \rho t}) \tag{21}$$

其中,

$$\gamma = 1 + 2\sin^2(\alpha - \frac{\pi}{4}) \tag{22}$$

$$\rho_x = (\frac{b_{c_x}^{gt} - b_{c_x}}{C_w})^2 \tag{23}$$

$$\rho_y = (\frac{b_{c_y}^{gt} - b_{c_y}}{C_h})^2 \tag{24}$$

式中:C_w 和 C_h 分别为预测框和真实框最小外接矩形的宽和高;$b_{C_x}^{gt}$ 和 $b_{C_y}^{gt}$ 分别为真实框中心的横坐标和纵坐标;b_{C_x} 和 b_{C_y} 分别为预测框中心的横坐标和纵坐标;α 为真实框与预测框中心点连线与 x 轴的夹角。当 α 趋近 0 时,Δ 越小,α 趋近 $\frac{\pi}{4}$,Δ 越大。

最后,考虑预测框的形状。形状损失的定义如下:

$$\Omega = \sum_{t=w,h} (1 - e^{-\omega t})^4 \tag{25}$$

$$\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})} \tag{26}$$

$$\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})} \tag{27}$$

式中:w 和 h 分别为预测框的宽和高;w^{gt} 和 h^{gt} 分别为真实框的宽和高。

根据以上 3 部分,SIoU 损失函数定义为:

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \tag{28}$$

SIoU 从预测框与真实框之间的距离、长宽比、重叠率等因素出发,引入回归向量角度,重新设定损失函数,降低回归自由度。经过对比实验证明,SIoU Loss 能带来更快的收敛和更好的性能,在模型保证精度的同时加快训练速度。

3 实验分析

3.1 数据集采集与制作

本文自制的数据集包含交警交通指挥手势和交通标志两类。交通标志的数据集来源于 CCTSDB2021[26],按交通标志的含义分为强制(mandatory)、禁止(prohibitory)和警告(warning)。交警交通指挥手势来自于文献[3],按照现实中交警指挥的 8 个手势,分别为直行(go)、停止(stop)、右转弯(turn_right)、左转弯待转手势(wait)、左转弯手势(trun_left)、变道(lane_change)、减速慢行(slow)、靠边停车(pull_over)。将文献[3]中公开的视频数据集,按

照 10 帧/s 进行划分，然后逐帧筛选，得到 8 种交警指挥手势关键手势的图片。再通过数据增强方式，如图 9 所示，扩展了交警交通指挥手势数据集的数量，提高了该自制数据集的可靠性。本文将以上两部分混合的数据集称为交通信息数据集，汇总得到 11 个种类共计 16 000 张，每个种类 1 400 张。在实验中，数据集划分为训练集、验证集和测试集，比例为 6∶2∶2。

(a) 原始图片　(b) 添加噪声和旋转　(c) 添加噪声和平移

图 9　数据集增强

3.2　实验环境与参数设置

本文自本文自制本文中模型训练和性能测评基于 AutoDL 云平台，硬件配置为：CPU 12 vCPU Intel(R) Xeon(R) Gold 5320 CPU@2.20 GHz；GPU 为 NIVDIA RTX A4000，显存为 16 GB；软件环境为：Python 3.8，Pytorch 1.9.0，CUDA 版本为 11.1。

实验所传入的图片大小设置为 640×640，总训练次数(epoch)为 100，批处理量(batchsize)为 16，初始学习率(learning_rate)为 0.005，权重衰减系数(weight_decay)为 0.000 5。启用 Mosaic 和 Mixup 数据增强对图片进行拼接处理，增强背景复杂度和多尺度的检测目标。使用随机梯度下降优化器在网络训练时对学习率进行微调。模型训练损失函数曲线如图 10 所示，本文模型在第 10 个 Epoch 之后 Loss 曲线开始稳定，拥有更快的收敛速度。

3.3　评价指标

为了客观评价改进后的 YOLOX-tiny 算法的性能，采

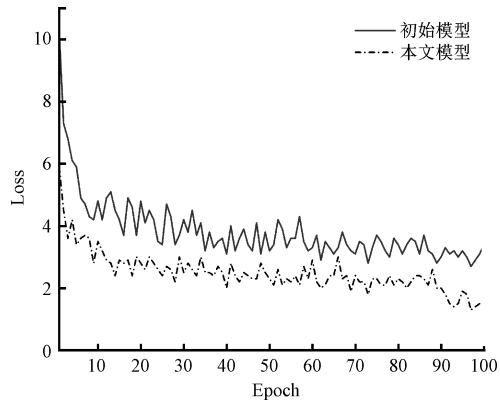


图 10　损失函数曲线

用查准率(precision，P)、召回率(recall，R)、平均精度(mean average precision，mAP)、每秒传输帧数(frames per second，FPS)作为模型性能的评价指标。其式为：

$$P = \frac{TP}{TP + FP} \tag{29}$$

$$R = \frac{TP}{TP + FN} \tag{30}$$

$$AP = \sum_i^N \int_0^1 P(R)\mathrm{d}R \tag{31}$$

式中：TP 表示被分配为正样本，而且分配正确的样本；FP 表示被分配为正样本，但是分配错误的样本；FN 表示被分配为负样本，而且分配错误的样本；FPS 用以衡量模型的检测速度。

3.4　消融实验

为了验证本文网络模型的改进对交警指挥手势的检测效果，按照 GhostNet-CA 主干网络、SCDE Head、SIoU 的顺序依次添加相应改进。消融实验结果如表 1 所示，“√”表示使用了该方法。

表 1　消融实验结果

实验模型	GhostNet	GhostNet-CA	SCDE Head	SIoU	P/%	R/%	mAP@0.5%	参数量/10^6	FPS
√	—	—	—	—	94.39	89.14	88.29	5.04	55.21
√	√	—	—	—	93.87	90.90	87.62	4.39	76.92
√	—	√	—	—	93.96	91.12	88.82	4.03	78.23
√	—	√	√	—	95.43	91.23	90.01	3.63	80.11
√	—	√	√	√	**97.88**	**92.58**	**90.60**	**3.63**	**80.11**

表 1 中可以看到，使用 GhostNet 替换了原始的 CSPDarknet53 后，mAP@0.5 略微降低，但是参数量减少了 13%，模型的计算量减少了 20%，FPS 提高了 21.71 帧/秒。而将 GhostNet 改进为 GhostNet-CA 后，提升了对全局特征的提取能力，mAP@0.5 提升了 1.2%，这说明网络在改进后提升了对中小模型的检测识别能力。对比原始的 CSPDarknet，改进后的 GhostNet-CA 实现了更小的模型体积、更快的检测速度和更高的准确率。

在 GhostNet-CA 的基础之上，继续使用 SCDE Head 替换原始的解耦头，mAP@0.5 提升了 1.19%，且 FPS 有所提升，这说明设计的 SCDE Head 不仅精简了解耦头结构，过滤了冗余信息，并且由于 GFR 模块学习了多尺度特征，进一步提升了对中小目标的检测识别能力。

继续改进损失函数，使用 SIoU Loss 代替 IoU Loss，解决了网络收敛速度慢的问题，并且有效减少了过拟合现象，并且在平均检测精度、精确率、召回率都有提示。

综合对比本文模型和原始模型，模型改进后检测速度提升了 45%，且缩减了 33.31%的计算量，使其更加契合无

人驾驶的应用场景,mAP@0.5 提升了 2.31%,召回率提升了 3.44%,准确率提升了 3.49%。

3.5 模型对比

为验证本文方法在交警指挥手势应用中的检测性能,将本文算法与 Faster-RCNN、YOLOv3-tiny、YOLOv7-tiny、YOLOv4-tiny、YOLOX-S 等方法进行准确率、权重大小(weights)和检测速度的比较,几种算法的训练过程及方法均相同。比较结果如表 2 所示。

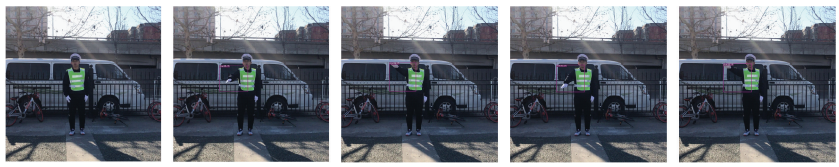
表 2 不同算法的结果对比

实验模型	参数量/10^6	GFlops	mAP@0.5%	权重大小/MB	FPS
Faster-RCNN	136.72	370.4	87.21	159.0	22.62
YOLOv3-tiny	8.70	13.0	87.82	16.65	30.21
YOLOv4-tiny	5.95	16.18	86.53	22.50	36.72
YOLOv7-tiny	6.0	13.0	88.26	12.3	62.36
YOLOX-S	9.01	26.77	89.82	34.3	40.38
YOLOX-tiny	5.04	15.25	88.29	19.4	55.21
GCAS-YOLOX	**3.63**	**10.17**	**90.60**	**16.3**	**80.11**

表 2 实验结果可以看出,与轻量级 YOLOv3-tiny、YOLOv4-tiny 算法相比,本文算法的参数量分别下降了 58.27% 和 38.99%,计算量分别下降了 21.76% 和 37.14%,精度方面分别提升了 2.78% 和 4.07%,且 FPS 有较高程度的优越性。同时,与 YOLOX 系列的网络 YOLOX-tiny、YOLOX-S 算法相比,在保持原始 YOLOX-tiny 模型的高 FPS 前提下,本文算法的参数量分别下降了 27.97% 和 59.71%,计算量分别下降了 33.31% 和 62.00%,精度方面分别提升了 2.31% 和 0.78%,模型的检测速度提升了分别提升了 98.39% 和 45.10%。由此可以认为,本文改进算法不仅有较快的检测速度,而且对网络进行了轻量化处理,有着更小的参数量和模型计算量,可以达到汽车无人驾驶过程中准确且快速的检测效果。

3.6 可视化分析

为了直观展现本文算法的改进效果,此处使用 YOLOX-tiny 算法和 GSCA-YOLO 算法进行可视化分析,分别用这两种算法对“减速慢行”和“左转弯待转弯”两种交警指挥手势进行检测,为了展示 YOLO 算法对手势动作的检测效果,分别截取了两种手势动作各 5 帧,运行结果如图 11 所示。



(a) YOLOX-tiny对“减速慢行”手势的检测结果

(b) GCAS-YOLOX对“减速慢行”手势的检测结果

(c) YOLOX-tiny对“左转弯待转弯”手势的检测结果

(d) GCAS-YOLOX对“左转弯待转弯”手势的检测结果

图 11 改进效果对比图

从检测结果对比图可以看出，在较复杂的环境背景之下，对于偏小尺寸的人物目标，同样的左转弯待转弯手势，改进后算法平均准确率提升了 3%，且整个手势动作都可以保持较高的识别率。而同样的左转弯待转弯手势，绝大多数时候可以准确识别出此手势，且算法在改进后平均识别率提升了 2%。

综上所述，本文算法的改进相比于 YOLOX-tiny 算法来说改进是有效的，不仅实现了模型的轻量化，而且在性能上也实现了提升，在复杂背景和人物尺寸偏小条件下识别情况良好。

4　结　　论

为了解决汽车驾驶过程中对于交警交通指挥手势的检测精度低、实时性差的问题，本文提出了一种基于 YOLOX-tiny 的轻量级交警指挥手势检测方法。在 YOLOX-tiny 基础上，使用 CA 模块改进 GhostNet 并用其替换主干网络、设计 SCDE Head 解耦头、损失函数改进为 SIoU。经对比试验分析，改进后的模型检测效果良好，相较于 YOLOX-tiny，不仅提升了 40% 的检测速度，检测精度也提升了 2.31%，召回率提升了 3.44%，准确率提升了 3.49%。后续将继续研究对网络结构的精简，以及在边缘设备上的移植与部署。

参考文献

[1] 常津津，罗兵，杨锐，等. 基于深度学习的交警指挥手势识别[J]. 五邑大学学报(自然科学版)，2018，32(2)：38-44，66.

[2] 郑颖，李培峰，罗恒杰，等. 交警手势识别的研究进展[J]. 计算机科学，2018，45(10)：87-91.

[3] HE J，ZHANG C，HE X L，et al. Visual recognition of traffic police gestures with convolutional pose machine and handcrafted features[J]. Neurocomputing，2019，390(5)：248-259.

[4] 张丞，侯义斌，何坚. 高度分层分区的图卷积交警手势识别技术[J]. 计算机辅助设计与图形学学报，2022，34(7)：1037-1046.

[5] 马天祥. 基于目标检测和模板匹配的交警手势识别研究[J]. 现代信息科技，2022，6(20)：60-64，70.

[6] 程贝芝，伍鹏，寇静雯，等. 结合全局上下文信息的交警手势识别方法[J]. 中南民族大学学报(自然科学版)，2023，42(3)：349-356.

[7] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2014：580-587.

[8] REN S Q，HE K M，GIRSHICK R，et al. Faster R-CNN：Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2015，39(6)：1137-1149.

[9] HE K，GKIOXARI G，DOLLAR P，et al. Mask R-CNN，international conference on computer vision[C]. Proceedings of the 2017 IEEE International Conference on Computer Vision(ICCV)，Venice，Italy，25 December 2017：2980-2988.

[10] REDMON J，FARHADI A. YOLOv3：An incremental improvement[J]. ArXiv Preprint，2017，ArXiv：1710：09412.

[11] BOCHKOVSKIY A，WANG C Y，LIAO H Y M. YOLOv4：Optimal speed and accuracy of object detection[J]. ArXiv Preprint，2020，ArXiv：2004：10934.

[12] LIU W，ANGUELOV D，ERHAN D，et al. SSD：Single shot MultiBox detector[C]. Computer Vision-ECCV 2016：14th European Conference，Amsterdam，The Netherlands，October 11-14，2016，Proceedings，Part Ⅰ 14. Springer International Publishing，2016：21-37.

[13] 李超军. 面向自动驾驶的交警手势识别算法研究[D]. 北京：北京工业大学，2019：26-35.

[14] 李晓杰. 基于计算机视觉的交警手势识别技术研究[D]. 哈尔滨：哈尔滨工程大学，2019：36-50.

[15] 康观龙. 基于胶囊网络的手势识别应用研究[D]. 江西：景德镇陶瓷大学，2022：36-50.

[16] 王新，王赛. 基于改进 YOLOv5 算法的交警手势识别[J]. 电子测量技术，2022，45(2)：129-134.

[17] GE Z，LIU S，WANG F，et al. YOLOX：exceeding YOLO series in 2021[J]. ArXiv Preprint，2021，ArXiv：2107：08430.

[18] HAN K，WANG Y，TIAN Q，et al. Ghostnet：More features from cheap operations[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020：1580-1589.

[19] HU J，SHEN L，SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2018：7132-7141.

[20] WOO S，PARK J，LEE J Y，et al. CBAM：

Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 3-19.

[21] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.

[22] LI J F, WEN Y, HE L H. SCConv: Spatial and channel reconstruction convolution for feature redundancy [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 6153-6162.

[23] LI X, WANG W H, HU X L, et al. Selective kernel networks[C]. Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, Piscataway, NJ: IEEE Press, 2019: 510-519.

[24] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 784-799.

[25] GEVORGYAN Z. SIoU loss: More powerful learning for bounding box regression[J]. ArXiv Preprint, 2022, ArXiv: 2205:12740.

[26] 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, DOI: 10.22967/HCIS.2022.12.023.

作者简介

方吴逸(通信作者),硕士研究生,主要研究方向为深度学习、神经网络硬件加速、集成电路设计。

E-mail:965932506@qq.com

陈章进,博士,教授,主要研究方向为深度学习、集成电路设计与平板显示。

E-mail:zjchen@shu.edu.cn

唐英杰,硕士研究生,主要研究方向为深度学习、集成电路设计。

E-mail:chnjstyj@outlook.com