DOI:10.19651/j.cnki.emt.2108071

基于改进 YOLOv5 算法的交警手势识别*

王 新 王 赛
(河南理工大学物理与电子信息学院 焦作 454000)

摘 要：为了解决交警手势在光照不均匀、背景复杂的环境下识别精准度低以及实时性差等问题，以 YOLOv5 网络模型为基础，针对标准卷积层感受野范围有限的问题，将部分卷积层替换为自校准卷积，增大感受野范围；引入置换注意力模块，提高算法的特征提取能力；针对交警所处环境复杂多变的问题，将焦点损失函数替换为广义焦点损失函数，提高算法在复杂环境下目标框的表示能力。实验结果表明，改进后的算法在满足实时性的基础上对于交警手势的检测平均精度高达 98.54%，相比于未改进的算法平均精度提高了 3.39%，且损失函数的损失值更小。

关键词：YOLOv5；自校准卷积；SA 模块；GFL 函数
中图分类号：TP391.4 **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 YOLOv5 algorithm

Wang Xin Wang Sai
(School of Physics & Electronic Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China)

Abstract: In order to solve the problems of low recognition accuracy and poor real-time performance of traffic police gestures in the environment of uneven illumination and complex background. Based on the YOLOv5 network model, part of the convolution layers are replaced by self-calibrated convolutions to increase the range of the receptive field. Shuffle attention module is introduced to improve the feature extraction ability of the algorithm. Aiming at the complex and changeable environment of traffic police, the focal loss function was replaced by generalized focal loss function to improve the expression ability of target frame in complex environment. Experimental results show that on the basis of real-time performance, the average accuracy of the improved algorithm for traffic police gesture detection is as high as 98.54%, which is 3.39% higher than that of the unimproved algorithm, and the loss value of the loss function is smaller.

Keywords: YOLOv5; self-calibrated convolutions; SA module; GFL function

0 引 言

伴随着计算机的发展，人工智能领域的发展越来越快，其应用包括方方面面，例如新兴的无人驾驶汽车。无人驾驶汽车作为一种新型的智能汽车，受到社会的很大关注，虽然无人驾驶汽车的研究已经相对完善，但其仍存在一些问题，比如对于交警手势的识别。

目前在手势识别方面已经有大量的学者进行过研究。传统的识别方法主要是根据颜色阈值的不同进行手势轮廓分割，通过提取出的手势轮廓特征进行区分[1-2]。这类方法的识别精准度低，速度慢，不能满足无人驾驶汽车的需求。现如今比较流行的方法是基于深度学习的识别方法，主要分为动态手势识别和静态手势识别。针对动态手势的识别，主要是通过划分每种手势的起始帧，然后输入到神经网络中进行学习训练[3-4]。这类方法对于每种手势的起始划分并没有明确的界限，且实时性较差，准确率偏低，模型较大，并不能满足无人驾驶汽车的需求。对于交警手势，通过观察可以得知，每种手势均可以通过一个关键手势表示，所以可以通过静态手势识别的方法实现对于交警手势的识别。伴随着目标检测算法的快速发展，越来越多的研究人员已经开始将目标检测算法应用到静态手势识别领域。王粉花等[5]通过对 YOLOv3-tiny 算法改进完成手势的识别；王晓华等[6]通过对 YOLOv3 算法进行改进完成对于缝纫手势的识别；睢丙东等[7]通过将 YOLOv3 进行改进完成手势识别。YOLO 算法对于手势的识别精准度高，速度快[8-12]，尤其是最新出来的 YOLOv5 算法相比于之前的 YOLO 系列算法具有更快的速度以及更高的精准度[13]。虽然 YOLOv5 网络模型已经在公开的数据集上验证具有不错的性能，但是根据所采用的不同数据集的特点，仍需进

收稿日期：2021-10-12
* 基金项目：国家重点研发计划(2016YFC0600906)、国家自然科学基金(61403129)项目资助

行一定的改进，如 Jia 等[14]通过在 YOLOv5 中融合三重注意力机制来提高算法的性能。

所以，本文以 YOLOv5 网络模型为基础，针对所采用的数据集光线明暗不一，环境背景复杂多变的特点，进行了进一步的改进。为了增加所用交警手势数据集的输出特征，增大卷积神经网络的感受野，将 YOLOv5 算法中第 18 与 21 层的 3×3 卷积层替换为自校准卷积[15]；通过在网络模型中添加置换注意力（shuffle attention，SA）模块[16]，加快算法对于手势特征的学习；针对交警指挥交通的环境复杂问题，将 YOLOv5 中原有的焦点损失（focal loss，FL）函数替换为广义焦点损失（generalized focal loss，GFL）函数[17]，提升算法对复杂环境下正负样本不平衡的解决能力。实验结果表明，改进后的算法在满足实时性的基础上平均精度更高且损失函数的损失值更低。

1　数据集处理

本文将文献[3]公开的数据集按照 5 帧/s 的速度进行分割，然后通过人工选取每种手势的图片。为了增加数据的可靠性，将所得到的手势关键帧进行数据增强操作，如图 1 所示。将所有的图片数据集统一处理成 512×512 大小，最后得到的 8 种手势总计 16 000 张，每种手势对应 2 000 张。

(a) 亮度调整　(b) 添加噪声　(c) 平移

图 1　数据增强

2　YOLOv5 算法及改进

2.1　YOLOv5 算法

YOLOv5 是于 2020 年 5 月份提出，经过更新迭代主要分为 4 个版本，分别为 YOLOv5m，YOLOv5l，YOLOv5x，YOLOv5s，其中 YOLOv5s 网络模型相对于其他 3 种版本模型最小，且检测速度最快，所以本文选用 YOLOv5s 模型进行交警手势的检测。其网络结构如图 2 所示，从图中可以看出 YOLOv5 主要由 3 部分组成：用于特征提取的主干网络、用于特征融合的颈部网络以及用于目标检测的输出部分。主干网络是一种卷积神经网络，它通过多次卷积和合并从输入图像中提取不同大小的特征映射。从图 2 中可以看出，主干网络中生成了 4 层特征图，颈部网络通过融合这些不同大小的特征图获得更多的上下文信息，减少信息丢失。输出部分根据颈部网络所生成的新的特征映射实现对目标的检测与分类。在 YOLOv5 网络体系结构中 Focus 模块将图片切片并进行连接，目的是在下采样期间更好的提取特征。CBL 模块由卷积、归一化和 Leaky ReLU 激活函数组成。在 YOLOv5 中存在两种跨阶段部分（cross-stage partial，CSP）网络，一种在主干，一种在颈部。CSP 网络旨在减小模型尺寸来提高推理速度，同时保持精度。此外 SPP 模块是指空间金字塔模块，通过不同大小的内核来进行最大池化，并通过将特征进行连接来完成融合。池化层通过执行降维操作来模拟人类视觉系统，以在更高的抽象级别上来表示特征，它主要是对输入的特征图进行压缩。一方面，它使特征映射更小，简化了网络的计算复杂度；另一方面，它进行特征压缩并提取主要特征。Concat 模块表示向量的串联操作，Upsample 表示上采样。

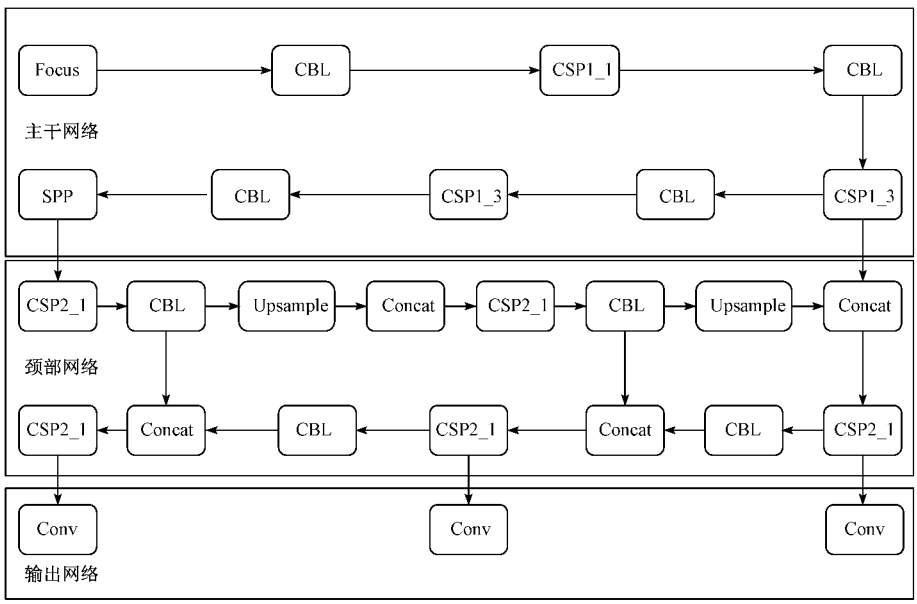


图 2　YOLOv5 网络结构

2.2　YOLOv5 算法的改进

1)基于自校准卷积网络的改进

对于标准卷积来说,一般是通过统一的小卷积核来完成特征的提取,这种操作一般只能学习到通道间的相似性,不能学习到彼此间的差异性。单纯的采用标准卷积限制了神经网络的感受野,得到的特征信息有限,为了增大卷积神经网络的感受野,提高在复杂环境下对于交警手势的特征提取能力。本文采用自校准卷积网络代替普通的卷积层从而提高每种手势的输出特征,其结构如图 3 所示。虽然相比于普通卷积层自校准卷积网络并没有引入多余的参数与复杂度,但在实际使用中仍会延长算法的推理时间,为了保证算法的实时性,本文只将 YOLOv5 中的第 18 与 21 层的 3×3 卷积替换为自校准卷积。该卷积自适应的构建了每个空间位置周围长期空间与信道间的依赖关系,从而可以通过合并得到更多信息。具体来说,自校准卷积首先是通过下采样将输入转换为低维嵌入,然后采用一个滤波器部分变换后的低维嵌入来校正另一部分滤波器的卷积变换。这种不均匀卷积和滤波之间的通信,使得每个空间位置的接受域可以有效地扩大。

自校准卷积操作流程:首先将输入 X 分成两部分 $\{X_1,X_2\}$,然后将 X_1、X_2 分别传入到不同的路径中去收集不同类型的上下文信息,其中 X_1 所在的路径完成自校准操作,输出 Y_1。X_2 所在的路径通过一个简单的卷积操作得到输出 Y_2,其中 $Y_2=\mathcal{F}_1(X_2)=X_2*K_1$,该路径使其保持原来的空间信息。将两条路径的输出 Y_1、Y_2 进行合并得到输出 Y。通过上述操作,自校准网络可以有效增大神经网络的感受野,提高算法的特征提取能力,提高检测精度。

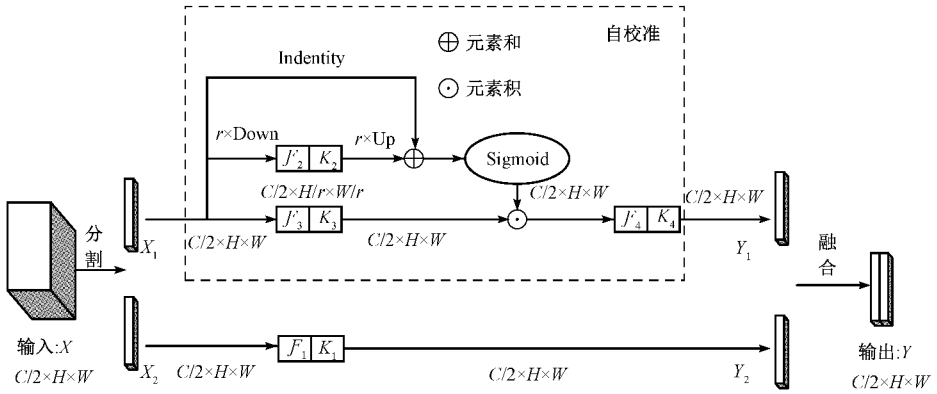


图 3　自校准卷积结构

2)SA 模块的添加

为了得到更多的有效信息,抑制其他通道的无效信息,本文引入了注意力机制。注意力机制可以使得计算机在有限的计算能力下,快速准确地完成对于目标的检测。注意力机制主要分为空间注意力机制、通道注意力机制以及两者相结合的注意力机制。空间注意力机制可以理解为让计算机重点关注某个区域,在该区域中找到自己想要的内容,而通道注意力机制则是根据事物不同通道所表示的重点信息不同,通过计算通道间的权重值,将重点放在权重值较大的通道。一般来说空间注意力机制与通道注意力机制单独使用时的效果相对两者结合起来时效果稍差,但将两者相结合的注意力机制模块虽然会提高一定的准确度,却也不可避免地增加了一定的计算量。针对这个问题,提出了 SA 模块,其流程如图 4 所示,它在空域与通道结合的基础上又引入了特征分组与通道置换,得到了一种超轻量型的注意力机制,实现了低复杂度高精准度的极佳效果。图中 $\sigma(\cdot)=sigmoid(\cdot)$,$\mathcal{F}_{fc}(x)=Wx+b$,$W$、$b$ 分别表示权重和偏置。

构建流程如下:

(1)特征分组:首先将给定的特征映射 $X\in R^{C\times H\times W}$ 根

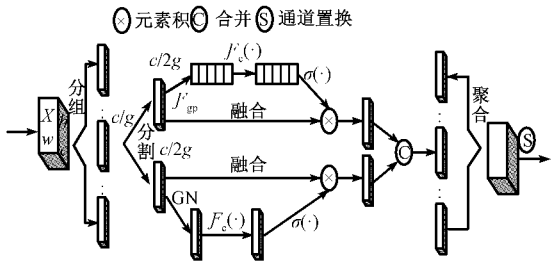


图 4　SA 流程

据通道维度特征划分为 G 组。则有 $X=[X_1,X_2,\cdots,X_G]$,其中 $X_K\in R^{C/G\times H\times W}$,$C$、$H$、$W$ 分别表示通道数、空间高度与宽度。对于划分后的子模块在训练过程中学习特定的语义反应,然后每个子模块生成相关的重要系数。具体来说就是在每个注意力单元开始时,输入 X_K 便根据通道维度划分成了两个分支 X_{K1},X_{K2},其中 X_{K1},$X_{K2}\in R^{C/2G\times H\times W}$,一个分支通过利用通道间的相互关系生成通道注意力图,另一个分支利用特征之间的空间关系生成空域注意力图,从而保证模块在空域与通道中均是有意义的。

(2)通道注意:首先使用全局平均池化嵌入到全局信息中,生成基于通道的统计信息 s,其中 $s=\mathcal{F}_{gp}(X_{K1})=$

$\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}X_{K1}(i,j)$，然后通过一个门机制实现精准与自选择指导。最后通道注意的输出结果如式(1)所示。

$$X'_{K1}=\sigma(\mathcal{F}_{\psi}(s))\cdot X_{K1}=\sigma(W_1 s+b_1)\cdot X_{K1} \tag{1}$$

其中，$W_1\in R^{C/2G\times1\times1}$，$b_1\in R^{C/2G\times1\times1}$ 用来对 s 进行缩放与变换。

(3)空间注意：首先通过组归一化(group norm，GN)对 X_{K2} 进行处理获得基于空间的统计信息，然后通过 $\mathcal{F}_{\psi}(\cdot)$ 进行增强，最终的空间注意结果输出如式(2)所示，其中 W_2 和 b_2 是生成 $R^{C/2G\times1\times1}$ 的参数，然后将两个分支融合到一起使得通道数与输入时的通道数相等。

$$X'_{K2}=\sigma(W_2\cdot GN(X_{K2})+b_2)\cdot X_{K2} \tag{2}$$

(4)聚合：将所有子模块进行聚合，使得 SA 模块最终的输出与输入 X 的维度大小相等。

为了保证算法在复杂环境下可以学习到更多的交警手势特征，本文在原网络模型中的第 5、8、12 三层均嵌入 SA 模块，从而提高算法的泛化能力。

3)FL 的改进

单阶段目标检测算法为了提高算法的检测速度，去掉了目标候选框的生成，这样就可能造成对于一张图片，若所要检测的目标较少，大部分的检测框就会位于背景区域，便会造成正负样本的极不平衡。针对这种问题，YOLOv5 在置信度损失和分类损失方面采用了 FL 函数，作为常用的解决正负样本比例不平衡的方法，FL 函数的使用在单阶段目标检测方面取得了较广的应用。但 FL 仍然存在一定的缺陷，如训练与测试中定位质量评估与分类评分的使用是不一致的，在训练时是将两者分开使用，而测试时又将两者合并在一起使用，且在复杂环境下，边界框的表示不够灵活。鉴于交警指挥交通时目标单一且环境复杂，本文采用 GFL 函数代替 FL 函数。GFL 函数主要由质量焦点损失(quality focal loss，QFL)以及分布焦点损失(distribution focal loss，DFL)两部分组成，QFL 的计算如式(3)所示，其中 $y\in[0,1]$，$y=0$ 表示负样本，$0<y\leqslant1$ 表示正样本。其中 sigmoid 算子 $\sigma(\cdot)$ 的输出记为 σ。

$$QFL(\sigma)=-|y-\sigma|^{\beta}((1-y)\log(1-\sigma)+y\log(\sigma)) \tag{3}$$

式中：$|y-\sigma|^{\beta}$ 表示调制因子，当一个样本的质量评估不够准确偏离标签 y 时，这个调制因子就会变得相对较大，QFL 便会花费较多的注意去学习困难样本，其中 β 控制着样本权值下降比例。相比于传统的 FL 离散标签，QFL 实现了标签的连续性，且对于 FL 中的标准交叉熵部分以及动态缩放因子均进行了扩充，实现了定位质量评估与分类评分的统一化。对于边界框，FL 通常采用狄拉克函数分布作为边界框的表示，如式(4)所示。这种边界框表示的方法在复杂场景下不够灵活，所以将式(4)进行进一步的广义推广，给定标签 y 的最大值 y_n 与最小值 y_o，可以从模型中获得估计值 $Y(y_o\leqslant Y\leqslant y_n,n\in N^+)$，推广后的公式如式(5)所示。为了与卷积神经网路保持一致，将式(5)进行离散化，离散化后的表达式如式(6)所示。将 $P(y_i)$ 简化为 S_i，则有 DFL 表达式如式(7)所示。DFL 通过快速增大目标 y 周围 y_i，y_{i+1} 的概率值使得网络对于边界框的表示更加灵活。将 QFL 与 DFL 整合在一起即为 GFL，假设一个模型对于两个变量 y_l，y_r($y_l<y_r$) 的估计概率为 P_{yl}，P_{yr}($P_{yl}\geqslant0$，$P_{yr}\geqslant0$，$P_{yl}+P_{yr}=1$)，则最后的预测值即为它们的线性组合，如式(8)所示。根据这个原则，将 QFL 与 DFL 进行合并，合并后的 GFL 公式如式(9)所示。GFL 函数相比于 FL 函数可以更好地适用于环境复杂、检测目标较少的情况。对于交警手势的检测有了进一步的改善，增加了算法的泛化能力与鲁棒性。

$$y=\int_{\infty}^{\infty}\sigma(x-y)x\,\mathrm{d}x \tag{4}$$

$$Y=\int_{\infty}^{\infty}P(x)x\,\mathrm{d}x=\int_{y_o}^{y_n}P(x)x\,\mathrm{d}x \tag{5}$$

$$Y=\sum_{i=0}^{n}P(y_i)y_i \tag{6}$$

$$DFL(S_i,S_{i-1})=-((y_{i+1}-y)\log(S_i)+(y-y_i)\log(S_{i+1})) \tag{7}$$

$$y_1=y_lP_{yl}+y_rP_{yr}(y_l\leqslant y_1\leqslant y_r) \tag{8}$$

$$GFL(P_{yl},P_{yr})=-|y-(y_lP_{yl}+y_rP_{yr})|^{\beta}((y_r-y)\log(P_{yl})+(y-y_l)\log(P_{yr})) \tag{9}$$

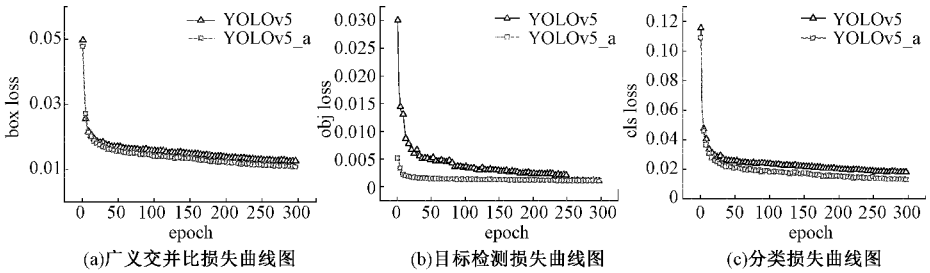
3　实验结果分析

3.1　实验环境

本文的实验平台是 64 位 Windows10 系统，处理器为 Intel(R) Xeon(R) W-2265 CPU，显卡为 NVIDIA Quadro RTX 4000，采用 pytorch 深度学习框架，开发环境为 pytorch1.8.1，python3.8。

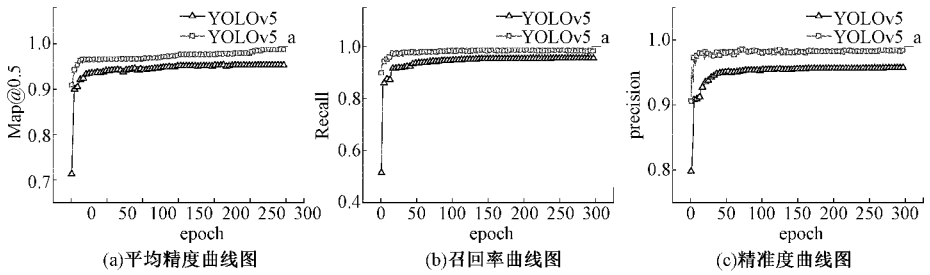
3.2　实验结果分析

本文将数据集按照训练集与测试集 7∶3 的比例进行划分，将改进后的算法命名为 YOLOv5_a，将改进前后的算法迭代次数统一设定为 300 次，初始学习率为 0.01，动量因子为 0.937。通过比较改进前后算法的损失函数、平均精度(Map)、召回率(Recall)、精准度(percision)的曲线图变化，分析算法改进前后的效果。其中损失函数值越低说明算法的效果越好，平均精度、召回率以及精准度的值越高说明算法的性能越好。图 5 为 3 种损失函数曲线图，其中 box loss 表示广义交并比损失，obj loss 表示目标检测损失，cls loss 表示分类损失。由图 5 可知改进后的算法 3 种损失函数值更小，如图 6 所示为改进前后算法召回率、精准度、以及平均精度的曲线变化图，由图 6 可知改进后的算法 3 种性能均比改进前的算法更好。为了进一步说明算法的适用性，将数据集又按照 8∶2 与 6∶4 的比例进行划分，结合之前的 7∶3，通过比较 3 次实验的平均精度的平均值与检测速度的平均值，进一步说明改进算法的有效性，实验数据如表 1 所示。



(a)广义交并比损失曲线图　(b)目标检测损失曲线图　(c)分类损失曲线图

图 5　3 种损失函数曲线图



(a)平均精度曲线图　(b)召回率曲线图　(c)精准度曲线图

图 6　3 种性能曲线图

表 1　实验数据对比

算法名称	划分比例	平均精度	平均值	检测速度/(帧·s^{-1})	平均值/(帧·s^{-1})
YOLOv5	6∶4	0.947 6	0.951 5	161	158
	7∶3	0.952 6		159	
	8∶2	0.954 4		154	
YOLOv5_a	6∶4	0.988 0	0.985 4	119	117
	7∶3	0.983 6		115	
	8∶2	0.984 7		116	

由表 1 可知,YOLOv5_a 算法的平均精度的平均值相比于 YOLOv5 算法提高了 3.39%,检测速度只是降低了 2.2 ms。在基本不影响算法检测速度的前提下,算法的性能有很大地提升。

4　结　　论

本文针对交警手势在光照不均、环境复杂情况下识别率低,实时性差等问题,以 YOLOv5 网络模型为基础进行了进一步的改进,通过修改卷积层、添加 SA 模块提高算法复杂环境下的特征提取与学习能力;通过替换 FL 函数,提高算法在检测目标较少、环境复杂情况下目标框的表示能力。实验结果表明,改进后的算法更加适用于交警手势背景复杂、目标单一的特点,且在基本不影响速度的前提下对于交警手势的检测平均精度提高了 3.39%,损失函数值更小,具有很好的推广性。在后续研究中将针对数据集中未考虑多个交警指挥交通的问题,对数据集进行扩充,并将模型部署到无人驾驶汽车上,通过云平台观察模型效果。

参考文献

[1]　王景中,李萌. 基于轮廓 PCA 的字母手势识别算法研究[J]. 电子技术应用,2014,40(11):126-128,135.

[2]　CAI Z, FAN G. Max-covering scheme for gesture recognition of Chinese traffic police [J]. Pattern Analysis & Applications,2015,18(2):403-418.

[3]　HE J,ZHANG C,HE X, 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,何坚,王伟东. 空间上下文与时序特征融合的交警指挥手势识别技术[J]. 电子学报,2020,48(5):966-974.

[5]　王粉花,黄超,赵波,等. 基于 YOLO 算法的手势识别[J]. 北京理工大学学报,2020,40(8):873-879.

[6]　王晓华,姚炜铭,王文杰,等. 基于改进 YOLO 深度卷积神经网络的缝纫手势检测[J]. 纺织学报,2020,41(4):142-148.

[7]　睢丙东,张湃,王晓君. 一种改进 YOLOv3 的手势识别算法[J]. 河北科技大学学报,2021,42(1):22-29.

[8]　化嫣然,张卓,龙赛,等. 基于改进 YOLO 算法的遥感图像目标检测[J]. 电子测量技术,2020,43(24):87-92.

[9]　姜文志,李炳臻,顾佼佼,等. 基于改进 YOLOv3 的舰船目标检测算法[J]. 电光与控制,2021,28(6):52-56,67.

[10]　徐晓光,李海. 多尺度特征在 YOLO 算法中的应用研究[J]. 电子测量与仪器学报,2021,35(6):96-101.

[11]　彭继慎,孙礼鑫,王凯,等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报,2021,42(10):161-170.

[12]　刘素行,吴媛,张军军. 基于 YOLOv3 的交通场景目标

检测方法[J]. 国外电子测量技术，2021，40(2)：116-120.

[13] 张宏群，班勇苗，郭玲玲，等. 基于 YOLOv5 的遥感图像舰船的检测方法[J]. 电子测量技术，2021，44(8)：87-92.

[14] JIA W，XU S，LIANG Z，et al. Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector [J]. IET Image Processing，2021，15(14)：3623-3637.

[15] LIU J J，HOU Q B，CHENG M M，et al. Improving convolutional networks with self-calibrated convolutions [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition，2020：10093-10102.

[16] ZHANG Q L，YANG Y B. SA-Net：Shuffle attention for deep convolutional neural networks[C]. ICASSP，2021：2235-2239.

[17] LI X，WANG W H，HU X L，et al. Generalized focal loss V2：Learning reliable localization quality estimation for dense object detection[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition，2021：11627-11636.

作者简介

王新，博士生导师，主要研究方向为信号处理、故障诊断。
E-mail：wangxin@hpu.edu.cn

王赛(通信作者)，在读研究生，主要研究方向为信号处理、手势识别。
E-mail：1397972649@qq.com