doi:10.6041/j.issn.1000-1298.2024.04.023

基于改进 YOLO v5n 的舍养绵羊行为识别方法

翟亚红[1]　王　杰[1]　徐龙艳[1]　祝　岚[1]　原红光[2]　赵逸凡[1]
(1.湖北汽车工业学院电气与信息工程学院,十堰 442002;2.沁阳市北盛牧业有限公司,沁阳 454550)

摘要:日常行为是家畜健康状况的重要体现,在传统的行为识别方法中,通常需要人工或者依赖工具对家畜进行观察。为解决以上问题,基于 YOLO v5n 模型,提出了一种高效的绵羊行为识别方法,利用目标识别算法从羊圈斜上方的视频序列中识别舍养绵羊的进食、躺卧以及站立行为。首先用摄像头采集养殖场中羊群的日常行为图像,构建绵羊行为数据集;其次在 YOLO v5n 的主干特征提取网络中引入 SE 注意力机制,增强全局信息交互能力和表达能力,提高检测性能;采用 GIoU 损失函数,减少训练模型时的计算开销并提升模型收敛速度;最后,在 Backbone 主干网络中引入 GhostConv 卷积,有效地减少了模型计算量和参数量。实验结果表明,本研究提出的 GS-YOLO v5n 目标检测方法参数量仅为 1.52×10^6,相较于原始模型 YOLO v5n 减少 15%;浮点运算量为 3.3×10^9,相较于原始模型减少 30%;且平均精度均值达到 95.8%,相比于原始模型提高 4.6 个百分点。改进后模型与当前主流的 YOLO 系列目标检测模型相比,在大幅减少模型计算量和参数量的同时,检测精度均有较高提升。在边缘设备上进行部署,达到了实时检测要求,可准确快速地对绵羊进行定位并检测。

关键词:舍养绵羊;智慧养殖;行为识别;注意力机制;YOLO v5n;绵羊数据集
中图分类号:TP391;S828　　**文献标识码**:A　　**文章编号**:1000-1298(2024)04-0231-10　　**OSID**:

Behavior Recognition of Domesticated Sheep Based on Improved YOLO v5n

ZHAI Yahong[1]　WANG Jie[1]　XU Longyan[1]　ZHU Lan[1]　YUAN Hongguang[2]　ZHAO Yifan[1]
(1. *School of Electrical and Information Engineering, Hubei University of Automotive Technology, Shiyan* 442002, *China*
2. *Qinyang Beisheng Pastoral Industry Co., Ltd., Qinyang* 454550, *China*)

Abstract: Daily behavior is an important manifestation of the health status of livestock. In traditional behavior recognition methods, livestock usually need to be observed manually or rely on additional tools. In order to solve the above problems, an efficient sheep behavior recognition method was proposed based on the YOLO v5n model, which used the target recognition algorithm to recognize the feeding, lying and standing behaviors of domesticated sheep from the video sequence above the sheepflod. Firstly, the daily behavior images of sheep in the farm were collected by cameras, and the data set of sheep behavior was constructed. Secondly, SE attention mechanism was introduced into the Backbone feature extraction network of YOLO v5n to enhance the global information interaction and expression capability and improve the detection performance. The GIoU loss function was utilized to reduce the computational cost and improve the convergence speed of the model. Finally, GhostConv convolution was integrated into Backbone network, which effectively reduced the calculation and parameter number of the model. The experimental results showed that the parameter number of GS-YOLO v5n object detection method proposed was only 1.52×10^6, which was reduced by 15% compared with the original model YOLO v5n. The FLOPs was 3.3×10^9, which was 30% less than the original model. The average accuracy achieved 95.8%, which was 4.6 percentage points higher than that of the original model. Compared with the current mainstream YOLO series of object detection models, the improved model significantly reduced the computational and parameter complexity of the model, while also achieved higher detection accuracy. It

收稿日期:2023-11-24　修回日期:2023-12-29
基金项目:湖北省教育厅重点科研项目(D20211802)和湖北省科技厅重点研发计划项目(2022BEC008)
作者简介:翟亚红(1979—),女,副教授,主要从事计算机视觉和人工智能研究,E-mail:zhaiyh_dy@huat.edu.cn
通信作者:徐龙艳(1982—),男,高级实验师,主要从事智能网联和机器视觉研究,E-mail:xuly_dy@huat.edu.cn

was deployed on edge devices and met the standard of real-time detection. It can accurately and quickly locate and detect sheep, providing ideas and support for intelligent sheep breeding.

Key words: domesticated sheep; intelligent farming; behavior recognition; attention mechanism; YOLO v5n; sheep data

0 引言

我国羊类养殖遍布在全国 32 个省、市、自治区中[1]。然而,在养羊业蓬勃发展的今天,仍存在一些问题制约着羊产业的发展[2]。随着规模化的羊只养殖兴起,必然会出现各种疫病的发生,对所饲养的羊群健康造成一定的危害[3]。羊只生病时常常伴随着异常行为,如喂料时不上槽等。因此对羊群的进食以及运动行为进行实时检测有助于对羊群的健康进行判断。传统上,畜禽的行为检测一般采用人工观察法[4]。但是,人畜长时间接触有共患病风险,而且会受饲养员的主观经验影响,用摄像头观察虽然可以防止直接接触,但是需要长时间查看监控,工作时间及工作量并无明显减少。因此,设计一种实时、准确的绵羊行为识别算法对绵羊智能化养殖有着重要意义。

随着机器视觉技术的发展,目前,深度学习技术已经被应用到了动物目标检测中,其中行为检测是研究热点之一[5]。LI 等[6]提出了一种用于猪只行为识别的时空卷积神经网络,精度达到 97.6%。张宏鸣等[7]在卷积操作时利用填充(Padding)增强网络对目标边缘特征的提取能力,采用丢弃(Dropout)方法提高网络的泛化能力,实现了对牛进食行为的识别。杨灵星等[8]提出了一种基于 EfficientDet 的检测方法,实现了对猪排泄行为的识别。TU 等[9]开发了 YOLOX-S 和 YOLO v5s 检测器用于检测猪只行为,准确率分别达到 99.4% 和 98.43%。相较于牛和猪等其他畜牧动物而言,绵羊有较强的合群性,喜欢聚集一起[10],群聚时遮挡严重,对目标检测算法而言具有挑战。

针对上述问题,提出基于改进 YOLO v5n 的绵羊行为识别算法,区别于通用目标识别算法使用的基准数据集 coco、ImageNet、VOC、KITTI 等预训练网络参数,利用识别目标的唯一性,用绵羊数据集训练针对性更强并且区分度更大的绵羊行为识别模型。

1 数据采集与预处理

1.1 数据采集

实验数据采集于河南省沁阳市北盛牧业有限公司绵羊养殖场,采集时间为 2023 年 7—8 月,天气晴,光照较强,室内环境。羊圈大小不一,羊只数量普遍为 5 ~ 10 头不等,由于多数行为识别研究都是通过对拍摄的图像进行训练,可能存在采集行为不完整等情况。因此本研究使用摄像头长时间录制视频,从真实养殖场景出发,挑选养殖羊只数量较多、大小羊混养的羊圈来采集数据,构建绵羊行为数据集,尽可能提高数据多样性,提高模型泛化能力。采集的部分绵羊图像如图 1 所示。

1.2 数据集构建与分析

本研究按照以下 4 个步骤进行处理,得到最终的绵羊数据集。

(1)数据筛选。对视频进行筛选,选出羊群活动较多的 10 段视频,每段视频时长 60 s,利用视频分帧技术,每秒视频截取 2 幅图像,共得到视频图像 1 200 幅,分辨率为 1 920 像素 ×1 080 像素。后期人工对图像进行筛选,去掉运动模糊等图像共得到视频图像 1 185 幅。

(2)数据标注。使用 Labelme 数据标注工具对视频图像进行标注,同时保存标注信息为 json 格式。



图 1 部分绵羊图像

Fig. 1 Examples of sheep pictures

(3)数据增强。基于 Mosaic 数据增强方式，本文选择左右翻转、上下翻转、缩放、平移、不规则裁切、大小变换、角度变换以及色域变换等数据增强方法来扩大数据的多样性。最终绵羊数据集扩充为3 400 幅图像。单幅图像增强变换结果如图 2 所示。

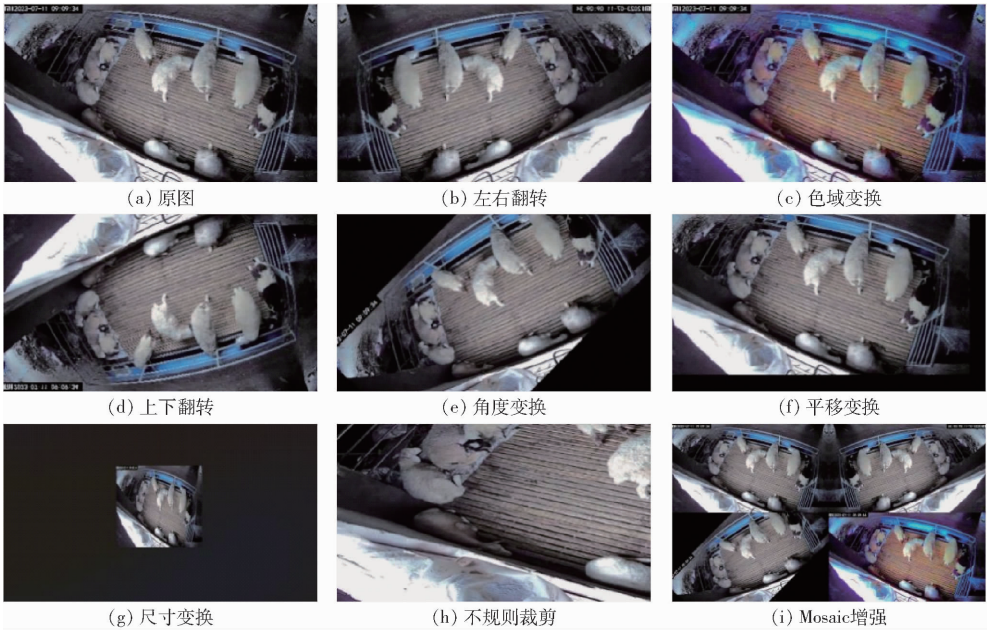


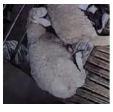
图 2　图像增强示例

Fig. 2　Image enhancement examples

(4)数据集划分。目前主流的小数据集通常都是按照 9∶1 来划分训练集和测试集，本文为了尽可能减少训练集过多而导致的过拟合现象，将步骤(3)中的数据按照 8∶2 划分为训练集和测试集。其中训练集 2 720 幅，测试集 680 幅。表 1 为绵羊行为判定标准和数据信息。

表 1　绵羊行为判定标准和数据信息

Tab. 1　Criteria and data information for determining sheep behavior

行为类别	行为图像	图像数量/幅	标签
站立		4 036	stand
躺卧		3 885	lie
进食		2 589	eat

2　研究方法

2.1　YOLO 目标检测算法

目前主流的目标检测模型分为两大类，一类是单阶段模型(one-stage)，一类是双阶段模型(two-stage)。其中，单阶段目标检测模型是直接在整个图像上预测物体类别和位置，减少了网络训练时间和模型复杂度，其特点是检测速度较快，但检测精度较低，具有代表性的网络模型有 YOLO 系列[11-13]、SSD[14]和 RetinaNet[15]等。双阶段目标检测模型先根据图像提取候选目标区域(Region proposal)，然后基于候选区域做二次修正得到检测结果，检测精度较高，但检测速度较慢，具有代表性的网络模型有 R-CNN[16]、Fast R-CNN[17]、Faster R-CNN[18]、FPN[19]和 Mask R-CNN[20]等。

YOLO v5 模型是一种基于 YOLO(You only look once)架构的单阶段目标检测模型，共有 5 个版本，模型内存占用量从小到大依次为 YOLO v5n、YOLO v5s、YOLO v5m、YOLO v5l、YOLO v5x，其中，YOLO v5n 模型是一种为了适应低功耗设备而设计的，考虑到需要实时检测等因素，本研究采用参数量较小、计算量较少的 YOLO v5n 模型作为基准模型。

YOLO v5n 模型主要包括输入层(Input)、主干网络(Backbone)、头部网络(Head)、颈部网络(Neck)4 部分。与 YOLO v4 相比，输入层设计了自适应锚框计算和自适应图像缩放两种方法，自适应锚框计算可以在每次训练开始之前，根据不同的数据集来自适应计算不同训练集中的最佳锚框值，自适应图像缩放方法能够自适应地添加最少的黑边到缩放之后的图像中，进一步提升推理速度；YOLO v5n 模型主干网络采用 CSP-Darknet53，作用是提取和融合图像特征；头部网络主要用于预测目标的类别、置信度和位置(bounding boxes)；颈部网络在

采用 CSPNet[21] 设计的 CSP2 结构和 Focus 模块后，加强了网络特征融合能力，降低了输入特征图通道数和尺寸，减少了计算量和内存消耗。YOLO v5n 网络模型结构如图 3 所示。

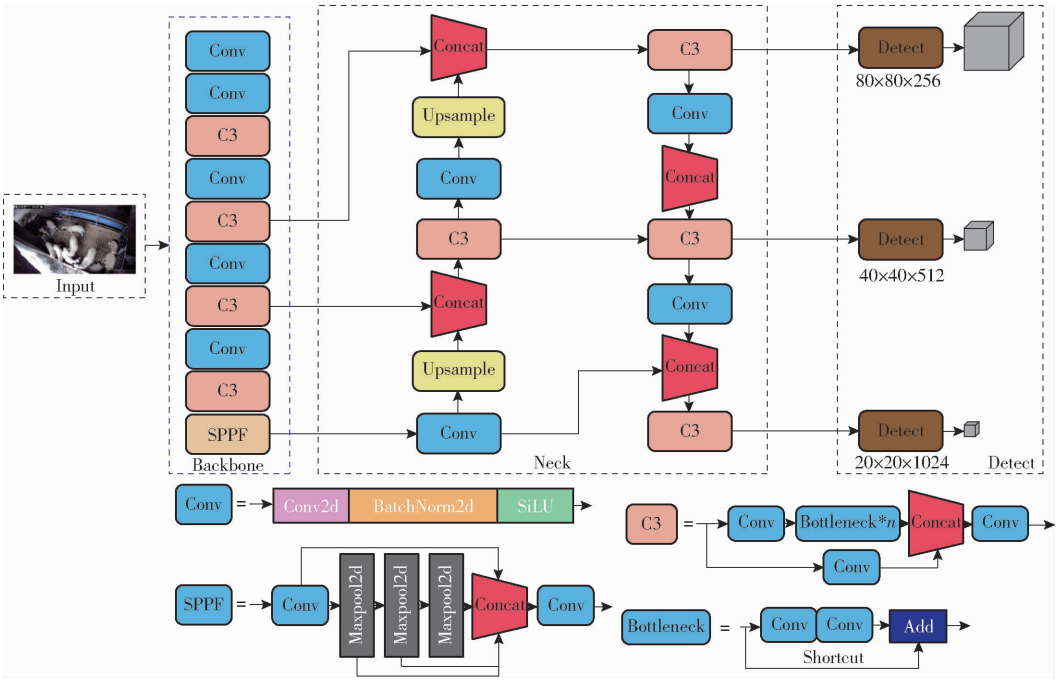


图 3 YOLO v5n 网络模型结构

Fig. 3 YOLO v5n network model structure

2.2 基于 YOLO v5n 的绵羊行为识别改进方法

作为当前流行的目标检测算法，YOLO v5n 已经整合了大量的优化策略，但是其提取的特征易受到噪声的影响。特征信息的丰富程度不足等问题使得 YOLO v5n 容易出现错检或漏检等问题，因此，本文设计了改进 YOLO v5n 的目标检测算法，能够更加有效地识别绵羊行为。

2.2.1 GIoU 损失函数

交并比（Intersection over union，IoU）是目标检测任务中常用的指标，通常用来衡量预测框（Predict box）和真实框（Ground-truth）的相似度，但是当 2 个框没有相交时，IoU 为 0，无法进行学习训练。

YOLO v5n 模型使用 CIoU 作为损失函数，但其边界框长宽比的衡量参数 v 计算相对复杂，导致训练过程计算开销较大，降低了模型收敛速度，且没有考虑难易样本平衡问题。因此本文引入 GIoU[22] 损失函数替代 CIoU 函数。

GIoU 于 2019 年提出，相较于 IoU 而言，GIoU 在其基础上增加了对非重合区域的考虑，使其能够更好地反映预测框与真实框之间的重合度。

2.2.2 SE 注意力机制

卷积神经网络（CNN）的核心构件是卷积算子，它使网络能够通过融合每一层局部感受野内的空间和通道信息来构建信息特征。而注意力机制可以有效地降低背景噪声带来的干扰，使得神经网络可以更加关注于目标物体的显著特征区域，因此得到了广泛的应用[23-26]。常见的注意力机制都是从特征空间域[27]或者特征图[28]的角度出发，获取目标在不同空间域或特征图上的重要性权值，但这些方法没有考虑通道维度上的注意力，无法有效地提取全局特征，并且需要更多的计算量。因此，本研究在 YOLO v5n 模型基础上引入 SE 注意力机制模块[29]（Squeeze-and-excitation networks，SENet）。SE 注意力机制模块可以通过学习自适应的通道权重，并根据任务的需要加权调整特征图中的通道贡献，使得模型更加关注有用的通道信息，提高特征的判别能力，从而提高模型性能和精度，同时可以有效地提取全局特征，使得靠近输入的层也可以获得全局感受野，并且 SE 注意力机制只增加少量参数和计算量，有着很好的灵活性和通用性。图 4 为 SE 注意力机制结构组成图。

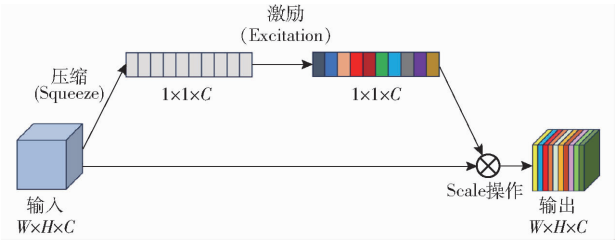


图 4 Squeeze-and-excitation 模块

Fig. 4 Squeeze-and-excitation module

首先，输入的特征图 $\boldsymbol{U} \in \mathbf{R}^{H \times W \times C}$ 通过 Squeeze

操作，将原始特征图维度从 $H \times W \times C$，压缩为 $1 \times 1 \times C$，生成一个通道描述符矩阵 $\boldsymbol{Z} \in \mathbf{R}^{1 \times 1 \times C}$，其中 H 为高度，W 为宽度，C 为通道数。相当于将每个二维特征通道（$H \times W$）变成一个实数（即变为一维）。将二维的特征通道 $H \times W$ 压缩成一维后，相当于这一维度获得了之前 $H \times W$ 全局的视野，感受野区域更广，得到了全局特征。$\boldsymbol{Z}$ 中第 c 个元素 z_c 计算公式为

$$z_c = F_{sq}(u_c) = \frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(i,j) \tag{1}$$

式中　$F_{sq}(\cdot)$——压缩函数

u_c——$\boldsymbol{U}$ 的第 c 个通道

为了提升泛化能力以及降低模型复杂度，在激活部分，由特定的激活函数 $F_{sq}(\cdot)$ 根据特征通道间的相关性来赋予一定的权值以突显网络认为更重要的部分通道，这里采用包含 2 个全连接层的 bottleneck 结构来捕获和通道相关的依赖性，其中第 1 个 FC 层起到降维作用，再使用 ReLU 函数进行激活，第 2 个 FC 层恢复到原始维度，将学习到的所有激活值都乘上原始特征，最后特征通过 Sigmoid 函数预测每个通道的重要性。该部分计算公式为

$$s = F_{ex}(\boldsymbol{Z}, W) = \sigma(W_2\delta(W_1\boldsymbol{Z})) \tag{2}$$

式中　s——每个通道对应的权重

$\boldsymbol{Z}$——经过 squeeze 操作后获得的全局特征描述符向量

σ——ReLU 函数

δ——Sigmoid 函数

$F_{ex}(\cdot)$——激励函数

W_1——第 1 层全连接层参数

W_2——第 2 层全连接层参数

最后的输出 $\widetilde{X}$ 是通过将通道间依赖关系的权重 $\boldsymbol{s}_c$ 应用到原始特征图 $\boldsymbol{u}_c$ 上进行缩放得到，可以表示为

$$\widetilde{X} = F_{scale}(\boldsymbol{u}_c, \boldsymbol{s}_c) = \boldsymbol{s}_c \times \boldsymbol{u}_c \tag{3}$$

式中　F_{scale}——权重向量 $\boldsymbol{s}_c$ 与特征映射 $\boldsymbol{u}_c$ 的逐通道乘法

2.2.3　C3Ghost 卷积模块

C3 卷积是一种结合了深度可分离卷积和空洞卷积的卷积模块，在通常情况下可以减少模型参数量和计算复杂度，但可能导致特征图分辨率下降，影响细节恢复和特征融合，增加模型不稳定性和训练难度。因此，本文使用一种更加优秀的结构 C3Ghost，它是一种轻量级的卷积神经网络结构，由 C3 结构和 GhostConv[30] 模块组成，相比于 C3 卷积，C3Ghost 在更加轻量化的同时，还提升了目标检测精度和速度。传统卷积模块如图 5 所示。

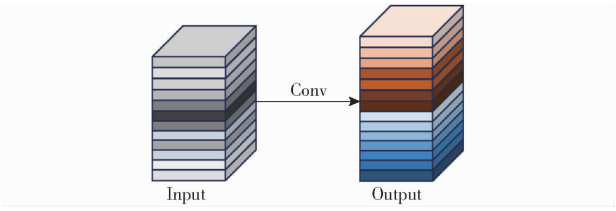


图 5　传统卷积模块

Fig. 5　Traditional convolution module

传统卷积计算方式为

$$Y = Xf + b \tag{4}$$

式中　X——输入特征图

f——n 个 c 通道 $k \times k$ 大小的卷积核

b——偏置

Y——通过卷积运算得到的输出特征图

GhostConv 模块如图 6 所示。

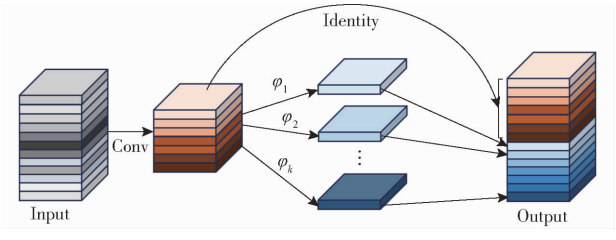


图 6　GhostConv 模块

Fig. 6　GhostConv module

GhostConv 计算公式为

$$Y' = Xf' \tag{5}$$

$$y_{i,j} = \phi_{i,j}(y'_i) \quad (i \in [1,m], j \in [1,s]) \tag{6}$$

式中　f'——卷积核，$f' \in \mathbf{R}^{m \times k \times k \times c}$

Y'——生成的 m 个原始特征图

y'_i——Y' 中的第 i 个通道特征图

$\phi_{i,j}$——第 j 个线性计算函数

$y_{i,j}$——由第 i 个原始特征图生成的第 j 个特征图

$\phi_{i,s}$ 为恒等映射，用于保留原始特征图。

2.3　基于 YOLO v5n 的绵羊行为识别模型

由于 GIoU 损失函数可以有效减少模型计算量和参数量，加快模型的收敛速度，同时还不会造成太多的精度下降，因此将 YOLO v5n 的 CIoU 损失函数替换为 GIoU 损失函数；为了弥补精度不足，在 YOLO v5n 的主干特征提取网络中引入 SE 注意力机制，通过学习自适应的通道权重，针对不同通道之间的相关性，提升网络对重要通道的关注强度，从而增强网络特征表示能力，在少量提升模型参数量的同时，提高模型检测精度；为进一步轻量化 YOLO v5n 模型，使用 C3Ghost 结构在 Backbone 层中替换 C3 模块，Ghost 模块可以对传统卷积神经网络模型特征图的冗余进行有效处理，在保证精确度的同时，

减少卷积数量,以少量的参数产生较多的特征信息。改进后的 GS - YOLO v5n 网络模型结构如图 7 所示。

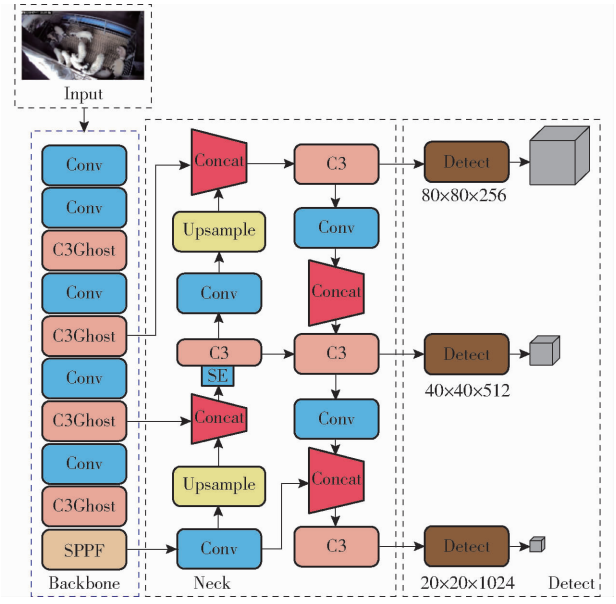


图 7 GS - YOLO v5n 网络模型结构

Fig. 7 GS - YOLO v5n network model structure

3 实验结果与分析

3.1 实验环境

实验采用的操作系统版本为 Ubuntu 18.04.5 LTS,CPU 采用 Intel Xeon Processor,GPU 采用 NVIDIA GeForce RTX 3060,CUDA 版本为 11.7,深度学习框架为 Pytorch 1.13.1,编译环境为 Python 3.7。

输入的图像分辨率为 640 像素 × 640 像素,模型训练时批量尺寸(Batch size)设置为 24,训练 300 个迭代周期(epoch),使用自适应学习率(Adaptive moment estimation,Adam)优化器优化模型。

3.2 评价指标

采用精确率(Precision)、召回率(Recall)、参数量(Params)、平均精度均值(mAP@50)、浮点运算量(FLOPs)和帧速率(FPS)评价算法的识别效果。精确率表示算法检测到的正样本中真正正确的比例,精确率越高,表示算法对目标的检测准确性越高。

召回率表示检测到的正样本数量占真实正样本数量的比例,召回率越高表示算法对目标的覆盖范围越大。

平均精度均值(mAP)是目标检测中最常用的评估指标之一,mAP 是一个综合指标,mAP 越高,表示算法检测准确性越好。

采用浮点运算量和参数量来度量模型复杂度,用帧速率来度量推理速度,FPS 指的是每秒检测到的帧数,FPS 不只与算法的内存占用量有关,还与硬件设备有关。

3.3 实验结果分析

3.3.1 消融实验结果与分析

为了验证不同注意力机制对算法性能的影响,选取几种常见的注意力机制来进行对比实验,实验结果如表 2 所示。

表 2 不同注意力方法对比

Tab. 2 Comparison of different related attention methods

SimAM	ECA	Triplet	SEG	SE	mAP@50/%
√					90.7
	√				92.1
		√			91.7
			√		82.3
				√	92.9

注:“√”表示模型包含此模块。

从表 2 中可以看出,SE 注意力机制模块相较于其他注意力机制,mAP@50 最大。注意力可视化如图 8 所示。

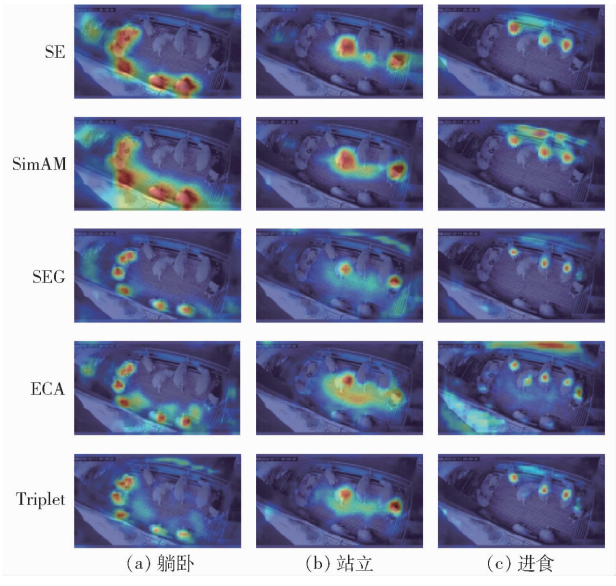


图 8 注意力可视化

Fig. 8 Attention visualization

为了验证改进方法的有效性以及单个分量方法对模型的影响,设计了消融实验来分析改进的结果,并对实验结果进行分析,实验结果如表 3 所示。

由于在真实的养殖环境中,设备算力往往较低,于是对原 YOLO v5n 网络的损失函数进行修改,使用 GIoU 损失函数,在参数量降低 13.5% 的同时浮点运算量降低了 10.6%,且精度有少量提升。这说明新的损失函数能够显著减少计算时间,提升整体性能表现。添加 SE 注意力机制后,相较于原始

表 3　消融实验结果

Tab. 3　Ablation experiment results

模型序号	GIoU	SEAttention	C3Ghost	参数量	浮点运算量	精确率/%	召回率/%	平均精度均值/%
1	–	–	–	1.77×10^{6}	4.7×10^{9}	90.5	86.4	91.2
2	√	–	–	1.53×10^{6}	4.2×10^{9}	88.9	87.7	91.6
3	–	√	–	1.87×10^{6}	4.2×10^{9}	92.4	87.9	92.9
4	–	–	√	1.47×10^{6}	3.3×10^{9}	91.2	87.2	92.3
5	√	√	–	1.60×10^{6}	4.2×10^{9}	91.6	88.1	93.8
6	√	–	√	1.48×10^{6}	3.4×10^{9}	89.1	87.9	92.4
7	–	√	√	1.49×10^{6}	3.3×10^{9}	91.9	89.8	93.6
8	√	√	√	1.52×10^{6}	3.3×10^{9}	92.4	91.1	95.8

注：实验以 YOLO v5n 模型为基准模型，“–”表示模型未包含此模块，“√”表示模型包含此模块。

网络显著提升了模型精确度、召回率和平均精度均值，且参数量并未增加太多，分析原因在于 SE 注意力机制能够辅助网络在特征提取过程中捕获目标信息，获得更多有效的特征细节。从消融实验结果看出，使用 C3Ghost 卷积模块取代骨干网络的 C3 卷积模块，参数量和浮点运算量大幅度降低，并且精确率、召回率和平均精度均值也有较大的提升。从消融实验结果可以看出，修改损失函数并改进 C3 模块相较于仅添加 SE 注意力机制，平均精度均值有小幅下降，而模型 5 和模型 7 的平均精度均值都有较大提升，分析原因为注意力机制能够以高权重聚焦重要信息，以低权重忽略不相关的信息，并且可以不断调整权重。

综合消融实验，在平衡模型轻量化和识别精度的前提下，采用 GIoU、SEAttention、C3Ghost 共 3 种方法改进 YOLO v5n 模型，并将改进后的模型命名为 GS－YOLO v5n。相较于 YOLO v5n 模型，改进模型参数量降低了 15%，浮点运算量仅为原始模型的 70%，并且精确率、召回率和平均精度均值分别达到 92.4%、91.1% 和 95.8%，能够在实现轻量化目的的同时提高模型的识别精度。

3.3.2　不同模型实验结果对比

为了对比 GS－YOLO v5n 模型与其他模型的性能，选取当前主流的单阶段目标检测算法，YOLO v3、YOLO v4s、YOLO v5n、YOLO v6n、YOLO v7-tiny、YOLO v8n 和 GS－YOLO v5n 模型进行对比实验，实验结果如表 4 所示。

由表 4 可以看出，YOLO v3 网络模型参数量和计算量最大，平均精度均值最小且帧率较低，分析原因为 YOLO v3 不适合处理密集目标；YOLO v4s 网络模型的平均精度均值仅次于本文模型，但浮点运算量和参数量较大，且帧速率较低，也不适用于数据集；YOLO v6n 网络模型帧率较低且识别精度较差，不适用于实时检测场景；YOLO v7-tiny 和 YOLO v8n 网络模型为 YOLO 系列近年提出的算法，其中

表 4　不同模型对比实验结果

Tab. 4　Comparison experiment results of different models

模型	平均精度均值/%	帧速率/(f·s^{-1})	参数量	浮点运算量
YOLO v3	87.8	23.6	6.17×10^{7}	1.553×10^{13}
YOLO v4s	94.5	32.9	7.62×10^{7}	1.69×10^{11}
YOLO v5n	91.2	37.1	1.77×10^{6}	4.70×10^{9}
YOLO v6n	88.1	24.8	4.50×10^{6}	1.14×10^{11}
YOLO v7-tiny	89.6	42.3	6.10×10^{6}	1.36×10^{11}
YOLO v8n	92.7	44.7	3.00×10^{6}	8.20×10^{9}
GS－YOLO v5n	95.8	48.5	1.52×10^{6}	3.30×10^{9}

YOLO v8n 在拥有较少浮点运算量和参数量的前提下还获得了较高精度，但相较于 GS－YOLO v5n 模型还存在精度较差的问题；此外，一般实时检测的要求为每秒检测图像数量大于 24，从表 4 中可以看出，改进后的 GS－YOLO v5n 模型帧速率为 48.5 f/s，远大于最低要求的 24 f/s。

从实验结果上可以看出，改进后的 GS－YOLO v5n 模型实现了在轻量化的同时拥有最高的识别准确度，且检测速度较高，能够满足实时检测需求。

绵羊的不同日常行为识别结果可以反映出模型对目标特征的识别能力，因此将本文模型 GS－YOLO v5n 分别与 YOLO v3、YOLO v4s、YOLO v5n、YOLO v6n、YOLO v7-tiny、YOLO v8n 模型对每类绵羊行为识别平均精度均值进行对比，结果如表 5 所示。

由表 5 可以看出，GS－YOLO v5n 模型对绵羊躺卧行为识别精度最高，达到 97.1%，对站立行为识别精度较低，为 94.8%，分析原因为在某些角度，站立行为和躺卧行为表现相似，模型容易产生混淆，误将站立行为识别为躺卧行为。GS－YOLO v5n 模型对躺卧和进食两种行为的平均精度均值均高于其他模型，最少提高 0.5 个百分点，最高提高 8.1 个百分点。YOLO v4s 对站立的平均精度均值略微高于

表 5 不同行为识别平均精度均值对比

Tab. 5 Comparison of mean average accuracy of different behavior recognition methods %

模型	站立	躺卧	进食
YOLO v3	83.7	92.2	87.5
YOLO v4s	95.6	93.4	94.5
YOLO v5n	88.4	90.1	95.1
YOLO v6n	85.2	91.5	87.6
YOLO v7-tiny	87.2	93.6	88.0
YOLO v8n	93.5	89.7	94.9
GS-YOLO v5n	94.8	97.1	95.6

GS-YOLO v5n,分析原因为 YOLO v4s 模型内存占用量较大。综合来看,GS-YOLO v5n 模型在绵羊躺卧行为识别任务上表现最好,在站立和进食行为检测任务上也有较优的表现。总体而言,GS-YOLO v5n 模型对绵羊日常行为识别具有一定的优势。

3.3.3 边缘设备算法部署

选定 Core-3399Pro-JD4 核心板作为嵌入式平台,配备了 ARM 双核 Cortex-A72 和四核 Cortex-A53 处理器架构,并支持 8/16 位操作系统的 NPU 设备,与 GPU 相比其功耗更低,约是 GPU 所需功耗的 1%,更适合边缘计算。

在对模型进行训练后,将原模型转换为 rknn 格式,引入帧速率作为评估指标,对比结果如表 6 所示。

表 6 边缘设备不同模型帧速率对比

Tab. 6 Frame per second comparison of different models of edge devices

模型	帧速率/(f·s^{-1})
YOLO v4s	15
YOLO v5s	7
YOLO v5n	18
YOLO v6n	17
YOLO v7-tiny	28
GS-YOLO v5n	32

由表 6 可见,在算力较低的边缘设备上,大模型的帧速率较低,尤其是 YOLO v5s,帧速率仅为 7 f/s,无法达到实时检测这一要求,经本文改进后的 GS-YOLO v5n 模型在平台端部署时的推理速度达 32 f/s,是 YOLO v5n 模型的 1.78 倍,可达到实时检测的需求。

3.3.4 识别结果分析

不同目标检测模型对绵羊行为识别效果如图 9 所示。由图 9 可看出,对于不同光线情况,以及拥挤和遮挡等情况下,改进后的绵羊行为识别算法的识别效果优于其他目标检测模型,能够准确识别绵羊的行为,且置信度较高,不存在漏检等情况;YOLO v3 模型置信度较差,并且不能完全识别出图像中所有的绵羊,存在漏检、误检等情况,因此不适用于绵羊的行为检测;YOLO v4s 虽然置信度较高,但是存在漏检情况,对于被遮挡的羊只识别效果较差,不能对绵羊进行精确的识别;YOLO v6n 存在漏检和误检情况,并且帧率较低,不能满足实时检测的要求;YOLO v7-tiny 和 YOLO v8n 网络模型对绵羊行为的识别效果弱于 YOLO v5n 以及改进后的 GS-YOLO v5n 模型,分析原因在于 YOLO v7 系列主要用于小目标物体检测,YOLO v8 网络模型对锚框的改进导致目标定位精度降低,因此使用以 YOLO v5n 为基础的模型效果更优。

根据对比实验和不同网络模型的识别结果可以看出,基于 YOLO v5n 改进的绵羊行为识别目标检测模型不仅拥有最小的计算量和参数量,并且在识别绵羊行为时效果最好,拥有最高的平均精度均值,能够准确检测出被遮挡的羊只,不存在漏检、误检等情况,且检测速度完全可以满足实时检测的需求。综上所述,改进后的模型能够较好地满足识别绵羊行为的需求,更适合在绵羊养殖场进行部署。

4 结论

(1)从真实养殖场景出发,针对计算设备算力

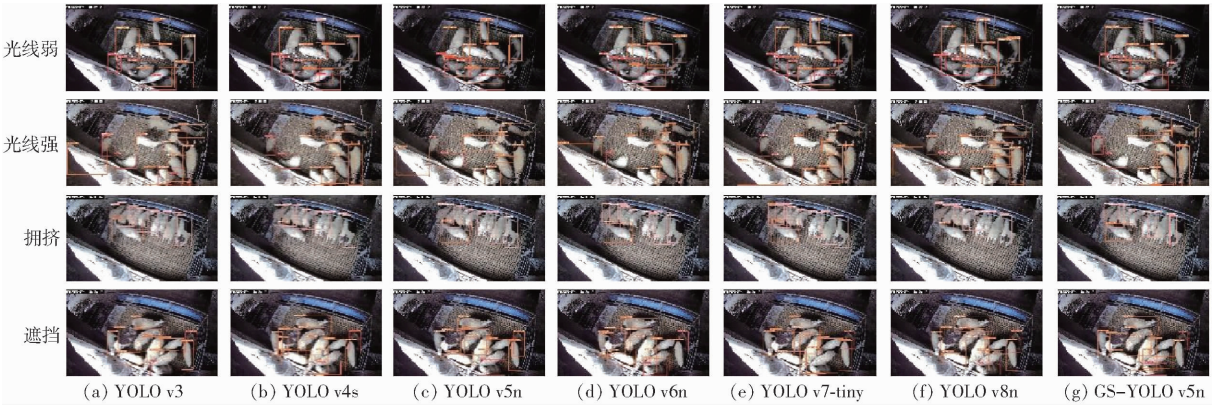


图 9 不同模型对绵羊行为的识别效果

Fig. 9 Recognition effect of different models on sheep behavior

不足以及对实时性要求高等情况，在 C3 模块的基础上融合了 GhostNet 网络，添加 SE 注意力机制模块，并对目标框 CIoU 损失函数进行了改进，提出了绵羊行为识别模型。该模型参数量比 YOLO v5n 模型降低了 15%，浮点运算量降低了 30%，并且平均精度均值达到了 95.8%，相比 YOLO v5n 模型提高了 4.6 个百分点。说明改进模型在更加轻量化的同时也具有较高的识别准确率。

（2）改进后的绵羊行为识别模型检测效果优于 YOLO v6、YOLO v7 以及 YOLO v8 等更先进的网络模型。相比于目前主流的目标检测模型，改进后的绵羊行为识别模型识别效果更好，能在羊群拥挤、光线不足等情况下精准识别绵羊行为，且速度较快，在算力较低的边缘设备上也能够满足实时检测的需求。

参 考 文 献

[1] 刘宏伟. 浅谈我国养羊业发展状况[J]. 吉林畜牧兽医，2021，42(6)：74.

[2] 王立波. 养羊业现状分析及解决对策[J]. 兽医导刊，2021(1)：100.

[3] 孙月辉. 肉羊场疫病科学防控问题[J]. 吉林畜牧兽医，2022，43(4)：101－102.

[4] 郝福明，田建艳，郑晟，等. 基于微惯性传感器的生猪运动监测系统[J]. 畜牧与兽医，2018，50(11)：23－27.
HAO Fuming，TIAN Jianyan，ZHENG Sheng，et al. Monitoring system of pig motion based on the micro inertial sensor[J]. Animal Husbandry & Veterinary Medicine，2018，50(11)：23－27.（in Chinese）

[5] 张宏鸣，孙扬，赵春平，等. 反刍家畜典型行为监测与生理状况识别方法研究综述[J]. 农业机械学报，2023，54(3)：1－21.
ZHANG Hongming，SUN Yang，ZHAO Chunping，et al. Review on typical behavior monitoring and physiological condition identification methods for ruminant livestock[J]. Transactions of the Chinese Society for Agricultural Machinery，2023，54(3)：1－21.（in Chinese）

[6] LI D，ZHANG K，LI Z，et al. A spatiotemporal convolutional network for multi-behavior recognition of pigs[J]. Sensors，2020，20(8)：2381.

[7] 张宏鸣，武杰，李永恒，等. 多目标肉牛进食行为识别方法研究[J]. 农业机械学报，2020，51(10)：259－267.
ZHANG Hongming，WU Jie，LI Yongheng，et al. Recognition methond of feeding behavior of multi-target beef cattle[J]. Transactions of the Chinese Society for Agricultural Machinery，2020，51(10)：259－267.（in Chinese）

[8] 杨灵星，彭小雪，邹亮，等. 基于 EfficientDet 的生猪排泄行为识别研究[J]. 河北农业大学学报，2023，46(4)：125－130.
YANG Lingxing，PENG Xiaoxue，ZOU Liang，et al. Research on pig excretion behavior recognition based on EfficientDet[J]. Journal of Hebei Agricultural University，2023，46(4)：125－130.（in Chinese）

[9] TU S，ZENG Q，LIANG Y，et al. Automated behavior recognition and tracking of group-housed pigs with an improved DeepSORT method[J]. Agriculture，2022，12(11)：1907.

[10] 逯国富. 绵羊的生活习性、分布范围与生产类型[J]. 养殖技术顾问，2014(10)：41.

[11] BOCHKOVSKIY A，WANG C Y，LIAO H Y M. YOLO v4：optimal speed and accuracy of object detection[J]. arXiv preprint arXiv：2004.10934，2020.

[12] LI C，LI L，JIANG H，et al. YOLO v6：a single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv：2209.02976，2022.

[13] WANG C Y，BOCHKOVSKIY A，LIAO H Y M. YOLO v7：trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition，2023：7464－7475.

[14] LIU W，ANGUELOV D，ERHAN D，et al. SSD：single shot multibox detector[C]// Computer Vision-ECCV 2016 14th European Conference，2016：21－37.

[15] LIN T Y，GOYAL P，GIRSHICK R，et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision，2017：2980－2988.

[16] 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.

[17] GIRSHICK R. Fast R－CNN[C]// Proceedings of the IEEE International Conference on Computer Vision，2015：1440－1448.

[18] REN S，HE K，GIRSHICK R，et al. Faster R－CNN：towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2017，39(6)：1137－1149.

[19] LIN T Y，DOLLAR P，GIRSHICK R，et al. Feature pyramid networks for object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：2117－2125.

[20] HE K，GKIOXARI G，DOLLAR P，et al. Mask R－CNN[C]// Proceedings of the IEEE International Conference on Computer Vision，2017：2961－2969.

[21] WANG C Y，LIAO H Y M，WU Y H，et al. CSPNet：a new backbone that can enhance learning capability of CNN[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition workshops，2020：390－391.

[22] REZATOFIGHI H，TSOI N，GWAK J Y，et al. Generalized intersection over union：a metric and a loss for bounding box

regression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658 - 666.
[23] CAO Y, XU J, LIN S, et al. Gcnet: non-local networks meet squeeze-excitation networks and beyond[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 1971 - 1980.
[24] LI X, WANG W, HU X, et al. Selective kernel networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 510 - 519.
[25] LI Y, YAO T, PAN Y, et al. Contextual transformer networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1489 - 1500.
[26] 方鹏,郝宏运,李腾飞,等.基于注意力机制和可变形卷积的鸡只图像实例分割提取[J].农业机械学报,2021,52(4):257 - 265.
FANG Peng, HAO Hongyun, LI Tengfei, et al. Instance segmentation of broiler image based on attention mechanism and deformable convolution[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 257 - 265. (in Chinese)
[27] 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.
[28] YANG L, ZHANG R Y, LI L, et al. Simam: a simple, parameter-free attention module for convolutional neural networks[C]// International Conference on Machine Learning, 2021: 11863 - 11874.
[29] 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.
[30] 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.

(上接第 220 页)

[16] RZANNY M, MÄDER P, DEGGELMANN A, et al. Flowers, leaves or both? How to obtain suitable images for automated plant identification[J]. Plant Methods, 2019, 15(1): 77.
[17] 罗娟,蔡骋.多线索植物种类识别[J].计算机工程与应用,2020, 56(5):160 - 165.
LUO Juan, CAI Cheng. Multi-cue plant species identification[J]. Computer Engineering and Applications, 2020, 56(5): 160 - 165. (in Chinese)
[18] LEE S H, CHAN C S, MAYO S J, et al. How deep learning extracts and learns leaf features for plant classification[J]. Pattern Recognition,2017, 71: 1 - 13.
[19] 孟惜,苑迎春,刘博,等.融合枣果与叶片信息的枣品种识别方法研究[J].河北农业大学学报,2022,45(2):107 - 113.
MENG Xi, YUAN Yingchun, LIU Bo, et al. Classification of jujube by integrating fruit with leaf information[J]. Journal of Hebei Agricultural University, 2022, 45(2): 107 - 113. (in Chinese)
[20] REDMON J, FARHADI A. YOLO v3: an incremental improvement[J]. arXiv Preprint arXiv:1804.02767, 2018.
[21] TAN M, LE Q. EfficientNetV2: smaller models and faster training[C]// International Conference on Machine Learning. PMLR, 2021: 10096 - 10106.
[22] DENG J, DONG W, SOCHER R, et al. Imagenet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 248 - 255.
[23] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning. PMLR, 2015: 448 - 456.
[24] SIFRE L. Rigid-motion scattering for image classification[D]. Paris: Ecole Polytechnique, 2014.
[25] 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.
[26] HUANG G, SUN Y, LIU Z, et al. Deep networks with stochastic depth[C]// Computer Vision-ECCV 2016: 14th European Conference, 2016: 646 - 661.
[27] ALEX K, ILYA S, GEOFFREY E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017,60(6):84 - 90.
[28] SIMONYAN K, ZISSERMA A. Very deep convolutional networks for large-scale image recognition[J]. arXiv Preprint arXiv: 1409.1556, 2014.
[29] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 - 778.
[30] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 1 - 9.