基于改进 YOLOv7 的矿井人员检测算法

邵小强[1,2]，李　鑫[1,2*]，杨永德[1,2]，原泽文[1,2]，杨　涛[1,2]

（1. 西安科技大学 电气与控制工程学院，西安 710054；2. 西安市电气设备状态检测与供电安全重点实验室，西安 710054）

摘要　矿井人员的实时检测是建设智慧矿山必不可少的内容，通过视频监测井下人员，从而实现危险区域预警及联动控制，对于矿井安全生产具有重要意义。现阶段可见光图像识别技术针对井下昏暗环境中人员的辨识还有待完善。针对井下光照不均、煤尘干扰严重导致监控视频存在噪声多、图像模糊等问题，提出一种改进 YOLOv7 的矿井人员检测算法。该算法在 YOLOv7 的基础上进行改进，首先针对 ELAN 模块直接拼接形成通道隔离的问题，提出基于通道重组与特征关注的复杂场景检测方式：使用 ShuffleNetV2 作为主干网络，减少参数量，提升算法推理速度，促进通道间的信息流动；将 Swin Transformer 注意力机制引入 shuffle_block 中，提升图像中人员的关注度，抑制复杂环境对人员检测的干扰，同时 Transformer 优异的全局感受野有利于遮挡目标的检测；其次针对特征融合结果未侧重预期目标且模型缺乏针对性策略提升小目标检测性能，在颈部多尺度融合网络添加 ACmix 模块，兼顾全局特征和局部特征，提升了算法对小目标的检测能力；最后引入 Efficient IOU Loss 提升算法收敛速度的同时减小目标框及先验框高度和宽度的差值，实现更加精准的定位。通过公开行人数据集及自建矿井人员检测数据集验证表明：该算法较 YOLOv7 模型相比检测精度提升了 3.1%，达到 89.4%；召回率提升了 3.8%，达到 86.4%；速度提升了 15.8%，达到 68.8FPS；满足矿井人员实时检测的工作要求。

关键词　矿井人员检测；YOLOv7；Swin-Transformer；ACmix；Efficient IOU Loss

中图分类号　TD76　　**文献标志码**　A　　DOI　10.12178/1001-0548.2023163

Mine personnel detection algorithm based on improved YOLOv7

SHAO Xiaoqiang[1,2], LI Xin[1,2*], YANG Yongde[1,2], YUAN Zewen[1,2], and YANG Tao[1,2]

(1. College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China; 2. Xi'an Key Laboratory of Electrical Equipment Condition Monitoring and Power Supply Security. Xi'an 710054, China)

Abstract　Real-time detection of mine personnel is an essential part of the construction of intelligent mine. It is of great significance to realize early warning and linkage control of dangerous areas by monitoring underground personnel through video, which is of great significance for mine safety production. At present, the visible light image recognition technology needs to be improved for the identification of personnel in the dim environment of underground coal mine. Aiming at the problems of more noise and image blur in the monitoring video caused by uneven illumination and serious coal dust interference in the underground, this paper proposes an improved YOLOv7 mine personnel detection algorithm. Firstly, aiming at the problem of channel isolation caused by direct splicing of ELAN modules, a complex scene detection method based on channel reorganization and feature attention is proposed. The ShuffleNetV2 is used as the backbone network to reduce the number of parameters, improve the reasoning speed of the algorithm, and promote the information flow between channels. At the same time, the Swin Transformer attention mechanism is introduced into shuffle_block to improve the attention of people in the image and suppress the interference of complex environment on personnel detection. At the same time, the excellent global receptive field of Transformer is conducive to the detection of occlusion targets. Secondly, in view of the fact that the feature fusion results did not focus on the expected target and the model lacked targeted strategies to improve the detection performance of small targets, the ACmix module was added to the neck multi-scale fusion network to take into account both global features and local features, which improved the detection ability of the algorithm for small targets. Finally, Efficient IOU Loss is introduced to improve the convergence speed of the algorithm and reduce the difference between the height and width of the target frame and the prior frame to achieve more accurate positioning. Through the verification of public pedestrian data sets and self-built mine personnel detection data sets, it is shown that the detection accuracy of the proposed algorithm is

收稿日期：2023−06−07；修回日期：2023−07−27

基金项目：国家自然科学基金（52174198）

作者简介：邵小强，博士，副教授，主要从事深度学习、目标检测方面的研究。

*通信作者 E-mail: 1187751601@qq.com

3.1% higher than that of the YOLOv7 model, reaching 89.4%. The recall rate increased by 3.8% to 86.4%. A 15.8% speedup of 68.8FPS Meet the mine personnel real-time detection work requirements.

Key words mine personnel detection; YOLOv7; Swein-Transformer; ACmix; Efficient IOU Loss

目前全球煤矿开采正由传统机械化开采向智能化开采过渡，文献 [1] 提出智能化煤矿系统架构，认为矿井目标检测是煤矿智能化高速通信及信息获取的基础，且应以人员检测为主。由于矿井使用人工光源照明且井下煤尘干扰严重，导致监控图像存在光照不均、细节模糊等问题[2]。工作人员无法长时间有效对视频进行多场景监控，井下作业人员的实际位置、工作情况等无法及时反馈到控制室，因此井下作业存在很大的危险性。

当前目标检测算法分为传统目标检测与深度神经网络[3]两大类，传统目标检测需要手工设计特征，使用滑动窗口的方式搜索图像，最终采用分类器进行分类。此类算法存在手工设计特征鲁棒性差，存在窗口冗余等问题，导致传统检测方法逐渐被深度神经网络所取代[4]。文献 [5] 使用 YOLOv4 针对矿井红外图像进行人员检测，通过迁移训练提升模型的泛化性，但需使用超分辨率卷积网络对红外图像进行预处理，导致模型整体参数增加，使得井下设备无法提供足够的计算量。文献 [6] 提出一种基于 Retinex 理论和小波多尺度的边缘检测算法，从低照度矿井图像中获取边缘图像，向矿用巡检机器人提供环境信息，该算法具有良好的实时性，但易受外界环境干扰，鲁棒性较差。文献 [7] 使用 YOLOv5 对矿井目标进行检测，采用轻量化主干网络加速模型的推理速度，使模型保持一定精度的同时达到实时检测标准，但是对于遮挡目标检测效果不佳。文献 [8] 使用 YOLOv5 和 Deepsort 算法实现井下人员检测及跟踪，引入跟踪模型增强了模型的抗遮挡能力，并且部署于嵌入式平台实现了矿井人员计数。但是该模型 ID 转换仍然存在，需要进一步的改进。上述方法专注于改进特征提取网络，适应井下图像特点，得到高精度的图像；或者为了满足参数轻量化，保证实时性，使用轻量化主干网络进行替换，使得模型存在一定的问题，无法保证模型检测精度与速率之间的均衡[9]。

针对上述问题，本文基于 YOLOv7[10]，提出一种可用于矿井人员实时检测的模型，首先采用 ShuffleNetV2[11] 作为模型的主干网络，加强 CPU 端的推理速度，同时在 shuffle_block 模块中引入 Swin Transformer[12] 注意力机制，增强感受野，提升模型的全局感知能力，优化模型在遮挡情况下的检测效果；其次在颈部多尺度融合网络添加 ACmix[13] 模块，通过模块内部的卷积通道和自注意力通道捕捉更多的特征，提升模型对小目标的敏感度[14]；最后引入 Efficient IOU Loss[15] 加速算法训练过程的收敛速度。

1 矿井人员检测模型

1.1 YOLOv7 检测模型

YOLO[16-20] 作为目标检测单阶段经典模型，由于其优异的运行速度和良好的精度被广泛应用于系统实时检测，并能够支持 GPU 设备以及边缘设备到云端的部署。而 YOLOv7 在 MSCOCO 和 ImageNet 数据集上检测显示，其准确率及速度均远超其余 YOLO 系列模型。相较于其前身 YOLOv5，YOLOv7 通过改进主干网络和特征融合的方式，进一步提升了检测精度；使用自适应卷积和 SPP-PANet 多尺度融合进行优化，使得模型保持较高精度的同时实现更快的检测速度；同时其架构相对简单，易于拓展和修改，还具有许多实用的工具和接口，使得用户能够快速进行模型训练和应用部署。因此本文选择 YOLOv7 作为矿井人员检测模型。

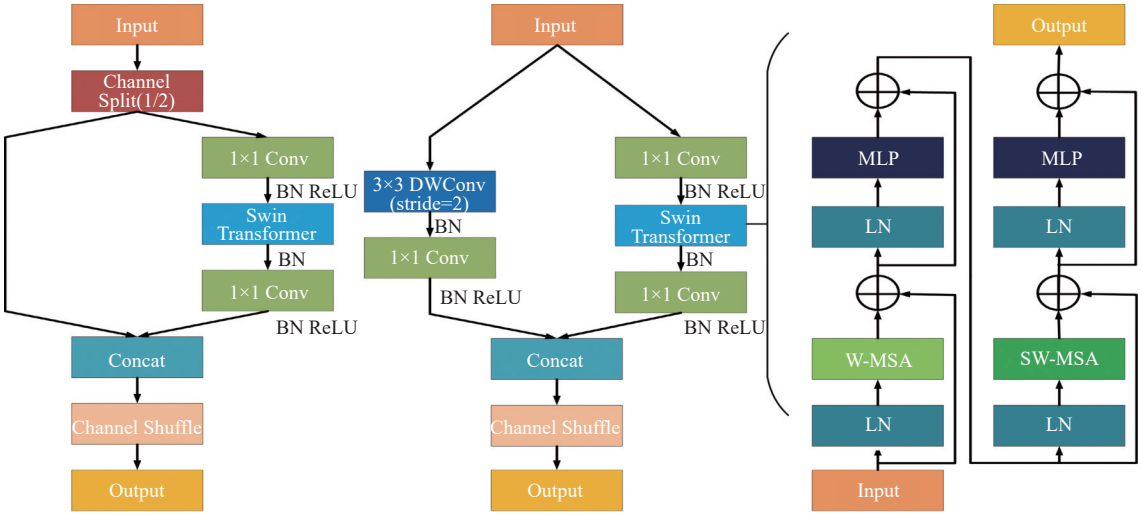
YOLOv7 主要包含了 Input、Backbone、Neck、Head 四部分。图像首先经过 Input 进行数据增强等预处理操作；然后送入 Backbone 进行特征提取；随后将所提取特征送到 Neck 进行融合得到三种不同尺度的特征，使得模型不仅能够在大规模、多类别的数据集上取得较好的表现，同时还能够在小样本、小类别的数据集上进行有效的训练和检测；最终在 Head 输出检测结果。其特有的高效聚合网络 ELAN，通过扩张、变换及融合基数的方式提升模型的学习能力，同时控制梯度路径加速模型的收敛；其设计的 Rep 重参数化卷积通过梯度流传播路径将重参数化的卷积和模型相结合，实现模型训练过程中速度与精度的均衡[21]。

1.2 基于通道重组与特征关注的复杂场景检测

YOLOv7 使用 ELAN 模块进行特征提取时，采用四个特征进行通道的拼接，但由于每个通道都包含着不同特征信息，这种直接拼接的方式易使各个卷积提取到的特征形成通道隔离，导致模型对于

复杂场景检测效果较差。

为了解决这一问题，本文引入高性能轻量化网络 ShuffleNetV2，其设计有两种模块：ShuffleNet-Unit1 和 ShuffleNet-Unit2，如图 1 所示。其核心操作是 Group convolution 和 Channel shuffle，利用分组卷积来扩展特征块的计算基数，使得特征信息被有效表达；同时减少参数量，降低网络过拟合的风险。但是分组卷积会限制网络特征信息的交流范围，影响网络的特征提取能力。因此在分组卷积后连接通道混洗操作，将不同组间的特征信息重新分组，促进通道间的信息流动，提升组间特征信息学习能力[22]。通道重组结构可以关注更多局部区域相关特征点所包含的信息，所以重组后的特征图相较原始特征图包含更多的语义信息。同时本文将原结构最大池化层采用深度可分离卷积进行替换，实现了通道与区域的分离，增强特征提取能力的同时降低参数量[23]；使用全局池化层替换原结构中的全连接层进行特征融合，保留前面卷积层提取的空间信息，提升网络的泛化能力。表 1 展示了改进 ShuffleNetV2 结构。



(a) STR_shuffleNet-Unit1 (b) STR_shuffleNet-Unit2 (c) Swin-transformer

图 1 STR_shuffleNet 基本单元

表 1 改进 ShuffleNetV2 结构

层数	输出大小	核大小	步长	重复	通道数
Image	224×224				3
Conv1	112×112	3×3	2	1	24
DW conv	56×56	3×3	2	1	24
Stage2	28×28		2	1	116
Stage2	28×28		1	3	116
Stage3	14×14		2	1	232
Stage3	14×14		1	7	232
Stage4	7×7		2	1	464
Stage4	7×7		1	3	464
Conv5	7×7	1×1	1	1	1 024
Global pooling	1×1	7×7			

当目标出现部分遮挡时，易发生漏检现象，主要原因在于现有结构仅具有局部信息的感知能力，而对于全局信息的感知较差，对于被遮挡目标无法进行有效检测。由于 CNN 具有良好的局部感受野，而 Transformer[24] 具有优异的全局感受野，考虑传统 Transformer 结构参数量庞大，而 Swin-Transformer 采用小窗口将特征序列分为几部分，并在小窗口内计算自注意力机制，让计算复杂度随着输入图像的大小呈线性增长，大幅降低计算复杂度。故本文引入 Swin-Transformer 进行特征提取，尽管 Swin-Transformer 很高效，但其不擅长建立远程跨窗口来扩大感受野，因此将其引入 shuffle_block，在基于窗口的自注意力模块中引入空间 Shuffle 操作，以提供窗口之间的连接并增强系统的建模能力。

STR_shuffleNet 利用 shuffle_block 特有结构，左侧支路保留局部特征感知，右侧支路采取全局信息度量，通过特征细化来增强被遮挡目标的检测效果，同时引入自注意力机制，建立针对目标位置信息的关注，在通道间建立特征映射关系，使得网络充分利用这些通道信息赋予目标通道更高的权重，有效确定目标的存在；在特征图层面，对特征图区域进行关注，捕获成对的像素级关系，更好的定位目标位置。最后采用通道交换的方式，使得特征通道具有更强的鲁棒交互性[25]。基于此构建的分组卷

积及特征关注的结构有效解决了模型对于复杂场景及遮挡检测效果不佳的问题。

输入图像经过卷积特征提取后传入该模块，在已知全局特征信息的前提下，Swin-Transformer 采取滑动窗口获取特征图块并结合自注意力有效挖掘每个窗口中的潜在目标，发现部分被遮挡的目标。同时通过多头自注意力机制来进行特征相似度学习，提升检测性能的同时降低对运行速度的影响。Swin-Transformer 模型结构如图 1 所示，模型参数如表 2 所示。

表 2　Swin-Transformer 模型参数

	Output size	Swin-Transformer
		Concat 4×4, 96-d, LN
Stage1	4×(56×56)	$\begin{bmatrix} win.sz. & 7\times7 \\ \dim 96 & head3 \end{bmatrix}\times2$
		Concat 2×2, 192-d, LN
Stage2	8×(28×28)	$\begin{bmatrix} win.sz. & 7\times7 \\ \dim 192 & head6 \end{bmatrix}\times2$
		Concat 2×2, 384-d, LN
Stage3	16×(14×14)	$\begin{bmatrix} win.sz. & 7\times7 \\ \dim 384 & head12 \end{bmatrix}\times6$
		Concat 2×2, 768-d, LN
Stage4	32×(7×7)	$\begin{bmatrix} win.sz. & 7\times7 \\ \dim 768 & head24 \end{bmatrix}\times2$

表 2 中，win.sz. 7x7 代表使用窗口的尺寸，dim 代表特征图通道的深度，head 代表多头注意力模块中头部个数。计算方式如式 (1) 所示：

$$\begin{aligned} F_{attM} &= LN(F_{in} + W-MSA(F_{in})) \\ F_{mid} &= LN(F_{att} + MLP(F_{att})) \\ F_{attW-M} &= LN(F_{mid} + SW-MSA(F_{mid})) \\ F_{out} &= LN(F_{attW-M} + MLP(F_{attW-M})) \end{aligned} \tag{1}$$

式中，MLP(*) 为全连接神经网络，LN(*) 为特征归一化处理，W-MSA(*) 和 SW-MSA(*) 分别表示滑窗注意力机制和基于窗口转换的滑窗注意力机制，计算方式如式 (2) 所示：

$$\begin{aligned} W-MSA(q,k,v) &= (F_{in}w^q, F_{in}w^k, F_{in}w^v) \\ SW-MSA(q,k,v) &= (F_{in}w^q, F_{in}w^k, F_{in}w^v) \\ A &= soft\max\left(\frac{q\times k^T}{\sqrt{d_k}}\right)\times v \end{aligned} \tag{2}$$

式中，q是查询矩阵，k是键矩阵，v是值矩阵，d_k代表输入特征F_{in}的通道维度，qk^T代表不同输入矩阵间的注意力分数，缩放因子$1/\sqrt{d_k}$负责提高稳定性。

1.3　基于自注意力与卷积混合模块的小目标检测

YOLOv7 模型 Neck 使用固定权重的多尺度融合模块将特征图进行拼接，未对多尺度融合模块的权重进行调整，且张量拼接对相邻层特征信息融合不全面，导致对小目标特征信息关注度不够，易造成特征信息的丢失；Head 采用 IDetect[26] 衔接普通卷积，特征融合结果未侧重预期目标，缺乏针对性策略来提升小目标检测，导致模型对尺度相差较大目标及小目标检测效果不佳[27]。

卷积注意力多注重输入与输出间的关系，而自注意力则多注重输入与输入间的关系。针对上述问题，本文结合卷积注意力和自注意力两者的优势，在多尺度融合中、低层引入 ACmix 模块帮助模型关注小目标特征，从特征中学习规律，对其重新校准，聚焦位置，两种注意力兼顾全局特征和局部特征，提升了模型对小目标的检测能力，同时分组卷积及特征关注的结构弥补了 ELAN 模块对复杂场景检测效果不佳的缺点，如图 2 所示。

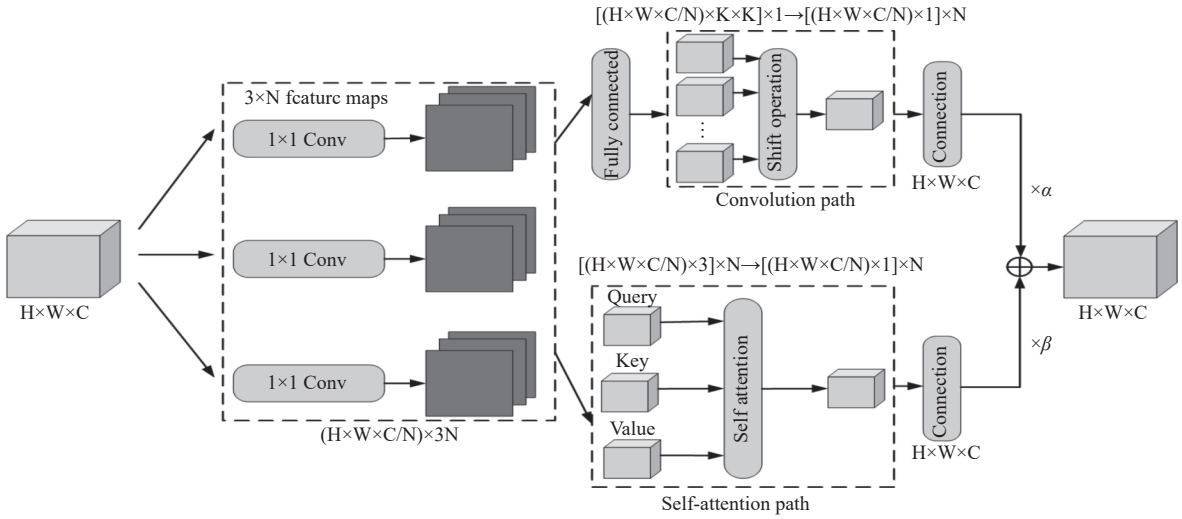


图 2　ACmix 结构

ACmix 模块由卷积注意力和自注意力两个模块并行组成。卷积注意力根据卷积的权值对特征图的局部感受野使用一个聚合函数，这些权值在整个特征图中共享，为图像处理带来不可或缺的归纳偏差，可以将标准卷积概况为两个阶段，第一阶段将输入的特征图沿某个位置的核权重进行线性投射，第二阶段将投影后的特征图依据核位置进行平移，最后进行聚合。计算方式如式 (3) 所示：

$$\begin{aligned} Stage1: \tilde{g}_{ij}^{(p,q)} &= K_{p,q} f_{ij}, \\ Stage2: g_{ij}{}^{(p,q)} &= Shift(\tilde{g}_{ij}^{(p,q)}, p-[k/2], q-[k/2]), \\ g_{ij} &= \sum_{p,q} g_{ij}{}^{(p,q)}. \end{aligned} \tag{3}$$

式中，K 代表核尺寸，F，G 分别代表输入、输出特征图，f_{ij}，g_{ij}分别代表 F 和 G 对应像素 (i,j) 的特征张量，(p,q) 代表核位置。

自注意力采用基于输入特征上下文的加权平均操作，通过相关像素对之间的相似函数动态计算注意力权重，这种灵活性使注意力模块能够自适应地关注不同的区域，获得更大的感受野和上下文信息，捕获更多特征，并易于区分背景与目标。自注意力也可以分解为两个阶段，第一阶段进行 1x1 卷积，将输入特征投影为查询、键和值，第二阶段包含注意力权重的计算及值矩阵的聚合。计算方式如式 (4) 所示：

$$\begin{aligned} &Stage1: q_{ij}^{(l)} = W_q^{(l)} f_{ij}, k_{ij}^{(l)} = W_k^{(l)} f_{ij}, v_{ij}^{(l)} = W_v^{(l)} f_{ij}, \\ &Stage2: g_{ij} = \Big\|_{l=1}^{N} \left(\sum_{a,b \in N_k(i,j)} A(q_{ij}{}^{(l)}, k_{ab}{}^{(l)}) v_{ab}^{(l)} \right). \end{aligned} \tag{4}$$

式中，$W_q^{(l)}$，$W_k^{(l)}$，$W_v^{(l)}$代表查询、键和值的投影矩阵，$N_k(i,j)$表示以 (i,j) 为中心的像素空间范围为 k 的局部区域，||为 N 个注意力头输出的串联。

最后将两分支的输出结果经过合并，兼顾全局特征和局部特征，从而提升了模型对于小目标的检测性能。

1.4 Efficient IOU Loss

YOLOv7 网络采用 CIOU[28] 作为坐标损失函数，其主要考虑目标的重叠面积、中心点距离和纵横比，定义 CIOU 损失如式 (5) 所示：

$$L_{CIOU} = 1 - \mathrm{IOU} + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \tag{5}$$

式中，b和b^{gt}分别代表目标框和预测框的中心点，c代表能够覆盖两个 Box 的最小外接框的对角线距离，ρ表示b和b^{gt}间的欧氏距离，α和v表示测量宽高比之间的差异。

由于 Bounding Box 回归是决定目标定位性能的关键步骤，而 CIOU Loss 并不能有效地描述 Bounding Box 的回归，且 CIOU Loss 涉及到反三角函数，在计算过程中会消耗一定的算力，致使损失函数收敛较慢，同时纵横比描述是相对值，存在一定的模糊，CIOU Loss 也未考虑难易样本的平衡问题，导致检测精度也随之降低。因此，本文提出将基于焦点损失的回归损失应用于 YOLOv7，通过回归过程聚焦高质量的先验框来加速模型训练过程的收敛速度，EIOU Loss 在 CIOU Loss 的基础上分别计算宽高的差异值取代了纵横比，解决了纵横比的模糊定义，同时引入 Focal Loss 解决难易样本不平衡的问题，提升了模型的检测精度。EIOU Loss 定义如式 (6) 所示：

$$\begin{aligned} L_{EIOU} &= L_{IOU} + L_{dis} + L_{asp} \\ &= 1 - \mathrm{IOU} + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \end{aligned} \tag{6}$$

式中，损失函数分为 IOU 损失L_{IOU}，距离损失L_{dis}，方位损失L_{asp}，h^c代表最小包围框的高，EIOU Loss 在增加长宽比相似度的同时，也考虑到模型如何通过焦点损失的回归过程减少(w,h)和(w^{gt},h^{gt})之间的真实差异。因此，EIOU Loss 通过直接减小目标框和先验框高度及宽度的差值，实现了更加精准的定位[29]。

2 实验与分析

2.1 实验环境搭建

本文环境搭配采用 Windows10 系统，CPU 型号 Intel 酷睿 i7 12700H，GPU 型号 NVIDIA GeForce RTX 3070Ti，显卡内存 32G，编程语言 python3.6，模型框架 PyTorch1.7.1。设置实验初始学习率 0.001，采用随机梯度下降法更新网络参数，学习动量为 0.9，权重衰减率为 0.000 5。

2.2 数据集样本

Caltech Pedestrian Detection：此数据集为目前规模最大的街景行人数据集，具有人员遮挡、复杂背景、尺度变化大等多种场景，标注超 25 万帧，35 万个矩形框，2 300 个行人，同时注明了不同矩形框间的时间关系及遮挡情况。

INRIA Person Dataset：此数据集为目前常见的静态人员检测数据集，其中人员均身处不同光照条

件及地点，以站姿为主且高度均超过 100 个像素点，图片来自谷歌，故清晰度较高。

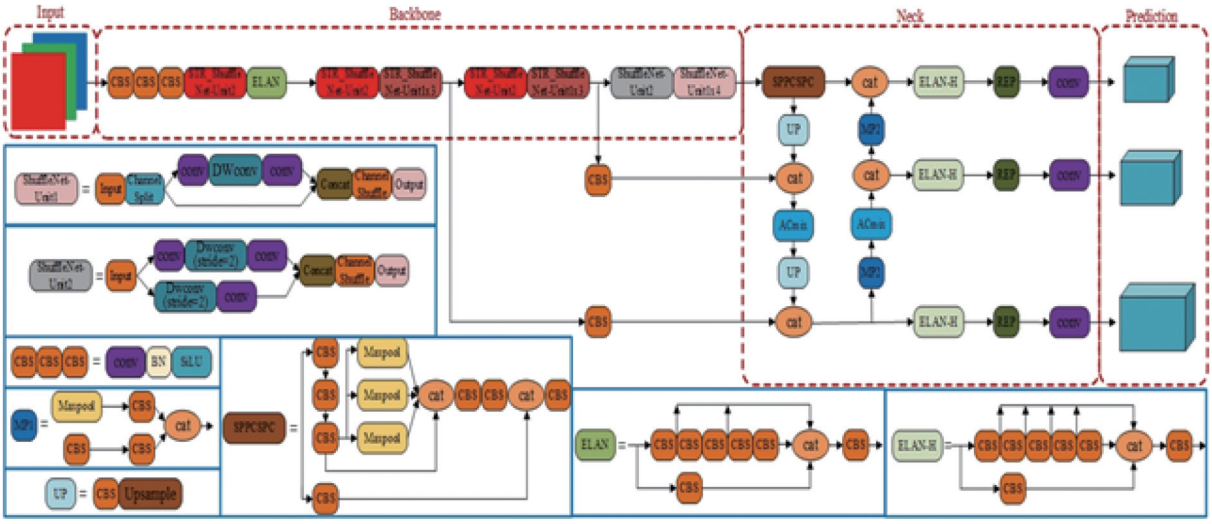


图 3　本文目标检测算法框架

自建矿井人员检测数据集：挑选矿井监控拍摄的两万帧图像构建数据集。首先使用 ffmpeg 工具将图像按帧切为图片，其次采用 Python 编写的 Labelimg 对图片中人员进行标注，最终转为适用于 YOLO 系列的 txt 文件。数据集涵盖井下多种环境：光照不均 2 267 张、煤尘干扰 1 568 张、目标遮挡 3 891 张、其余环境 1 800 张。

2.3　评价指标

本文采用模型参数量、每秒帧率 FPS、准确率 M_p、召回率 M_r、漏检率 M_m、误检率 M_f 及精确率均值 mAP@0.5 作为算法的评价指标。

$$M_p = \frac{T_p}{T_p + F_p} \tag{7}$$

$$M_r = \frac{T_p}{T_p + F_N} \tag{8}$$

$$M_m = \frac{F_N}{F_N + T_P} \tag{9}$$

$$M_f = \frac{F_p}{F_p + T_N} \tag{10}$$

$$mAP = \frac{T_p + T_n}{T_p + T_n + F_p} \tag{11}$$

式中，T_p 代表被正确检测出的人员，F_N 代表未被检测到的人员，F_p 代表被误检的人员，T_N 代表未被误检的人员。

2.4　目标检测实验结果与分析

为验证 YOLOv7 模型在矿井人员检测方面性能的优越性，本文分别将 Faster-CNN[30]、SDD[31]、YOLOv4、YOLOv5s、YOLOv5m、YOLOv5l 及 YOLOv7 七种模型在自建矿井人员检测数据集上进行训练。输入图像大小为 640x640，迭代次数为 150 次，批次大小为 16，实验结果如表 3 所示。

表 3　常见检测算法实验结果

Model	Parameter/M	FPS	M_s	M_r	$M_\pm$	M_f	mAP@0.5
Faster CNN	68.2	41.6	0.842	0.809	0.101	0.133	0.833
SDD	82.1	32.9	0.850	0.855	0.214	0.125	0.864
YOLOv4	24.7	55.1	0.829	0.797	0.232	0.095	0.822
YOLOv5s	13.9	86.4	0.791	0.754	0.355	0.114	0.799
YOLOv5m	19.3	67.3	0.819	0.821	0.217	0.094	0.811
YOLOv5l	28.7	56.4	0.833	0.865	0.157	0.078	0.827
YOLOv7	26.3	59.4	0.857	0.832	0.179	0.097	0.849

在相同迭代次数内，由于 YOLOv7 模型参数的增加，模型检测速度低于 YOLOv5s 和 YOLOv5m，但 59.4 帧每秒的速度已经满足矿井人员实时检测的需求；且 YOLOv7 模型在自建矿井人员数据集

上其余评价指标基本优于其他模型，综合考虑下本文选择 YOLOv7 模型，并在此基础上进行改进以增强其在复杂环境中对矿井人员检测的各评价指标。

为验证本文算法改进的有效性及轻量化主干选择的合理性，将本文算法与 YOLOv7、YOLOv7-EfficientNetV3、YOLOv7-GhostNet、YOLOv7-ShuffleNetV2 通过自建矿井人员检测数据集按照相同参数设置进行训练，结果如图 4 和图 5 所示。

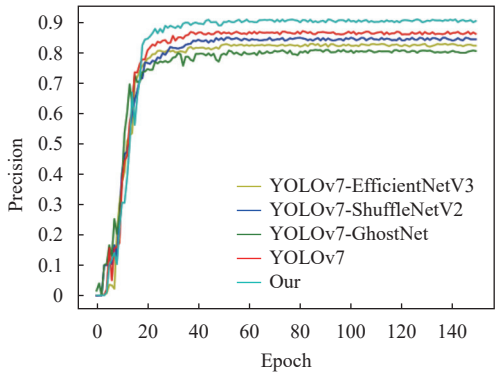


图 4　准确率曲线

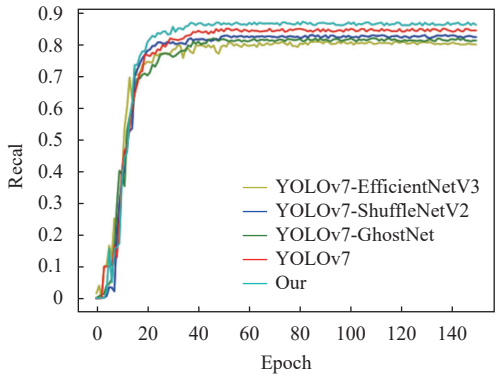


图 5　召回率曲线

原始 YOLOv7 迭代到 30 次时，准确率上升到 0.85 左右，最终收敛到 0.86 左右；YOLOv7-EfficientNetV3 迭代到 35 次时，准确率上升到 0.79 左右，最终收敛到 0.80 左右；YOLOv7-ShuffleNetV2 算法迭代到 40 次时，准确率上升到 0.83 左右，最终收敛到 0.84 左右；YOLOv7-GhostNet 算法迭代到 35 次时，准确率上升到 0.75 左右，最终收敛到 0.78 左右；本文算法迭代到 25 次时，准确率上升到 0.88 左右，最终收敛到 0.89 左右。

原始 YOLOv7 算法迭代到 40 次时，召回率上升到 0.82 左右，最终收敛到 0.83 左右；YOLOv7-EfficientNetV3 迭代到 45 次时，召回率上升到 0.77 左右，最终收敛到 0.78 左右；YOLOv7-ShuffleNetV2 算法迭代到 40 次时，召回率上升到 0.80 左右，最终收敛到 0.81 左右；YOLOv7-GhostNet 算法迭代到 50 次时，召回率上升到 0.78 左右，最终收敛到 0.79 左右；本文算法迭代到 35 次时，召回率上升到 0.85 左右，最终收敛到 0.86 左右。因此，本文算法的改进较 YOLOv7 相比具有显著的准确率及召回率提升；结合表 4 得知轻量化主干 ShuffleNetV2 可以保持较高精度及召回率的同时提升模型检测速度。

将本文算法通过自建矿井人员检测数据集进行训练，训练损失变化如图 6 所示，可以看出模型分类损失 cls_loss、定位损失 box_loss、置信度损失 obj_loss 随着训练进程均成下降趋势，在 30 次迭代训练后，模型趋于稳定，分类损失趋于拟合达到 0.16 左右；定位损失趋于拟合达到 0.19 左右；置信度损失趋于拟合达到 0.21 左右，均达到整体最优解，表明本文算法改进损失函数后模型具有良好的收敛能力及鲁棒性。

表 4　消融结果

Model	ShuffleNetV2	Swin-Transformer	ACmix	EIOU	Precision	Recall	mAP	FPS
1					0.867	0.832	0.849	59.4
2	√				0.841	0.812	0.824	79.9
3	√	√			0.859	0.829	0.833	70.3
4	√		√		0.853	0.845	0.831	69.9
5	√			√	0.842	0.843	0.826	81.1
6	√	√	√		0.889	0.853	0.877	65.9
7	√	√	√	√	0.894	0.864	0.882	68.8

综上所述，本文选取的轻量化 ShuffleNetV2 主干使得模型保持一定精度的同时降低了计算量；主干的替换、注意力机制的引入、多尺度融合的改进、损失函数的优化对于目标检测性能有着明显的提升。因此，本文算法对于井下复杂环境中的人员检测具有良好的精度及鲁棒性。

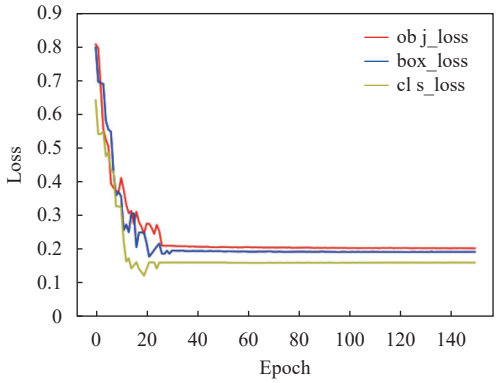


图 6 损失值曲线

本文在 YOLOv7 的基础上进行了主干网络的替换、注意力机制的引入、多尺度融合的改进、损失函数的优化。为验证各部分改进的有效性，以 YOLOv7 算法为基准，使用消融实验保持实验平台参数设置一致进行训练，消融实验结果如表 4 所示。

YOLOv7 的主干网络替换后，准确率下降了 3.1%，召回率下降了 2.4%，速度提升了 34.5%；在模型 2 的基础上添加 Swin Transformer 模块后，准确率提升了 2.1%，召回率提升了 1.9%，速度下降了 12.0%；在模型 2 的基础上添加 ACmix 模块后，精度提升了 1.4%，召回率提升了 3.9%，速度下降了 12.5%；在模型 2 的基础上，改进损失函数后，精度几乎不变，召回率提升了 3.8%，速度提升了 1.5%；在模型 2 的基础上同时添加 Swin Transformer 模块和 ACmix 模块，准确率提升了 5.7%，召回率提升了 5.0%，速度下降了 17.5%；而本文算法较模型 2 相比，准确率提升了 6.3%，召回率提升了 6.4%，速度下降了 13.8%；由此可知，分别对模型 2 进行注意力机制的引入、多尺度融合的改进，矿井人员检测性能提升有限，而将两种改进组合添加时，矿井人员检测性能获得了显著的提升，而损失函数的作用在于加速模型收敛的同时减少与目标框重叠较少的锚框，对边界回归进行优化，从而提升了对遮挡目标的检测精度。最终本文算法相较 YOLOv7 算法，准确率提升了 3.1%，召回率提升了 3.8%，速度提升了 15.8%。综上所述，通过消融实验验证了模型各部分改进的有效性，并且本文算法在自建矿井人员检测数据集上较原始 YOLOv7 算法具有显著的优势。

为了进一步验证本文算法的泛化性能及在小目标检测方面的提升，将三种数据集中目标 GT 框按照 [0, 32x32], [32x32, 96x96], [96x96, 640x640] 区间分为大中小三类进行验证，并分别统计检测结果的精度，性能指标对比如表 5 所示。

表 5 公开数据集实验结果

数据集	性能指标	YOLOv7				Our			
		大	中	小	总	大	中	小	总
Caltech Pedestrian Detection	Precision	0.884	0.882	0.865	0.877	0.908	0.903	0.901	0.904
	Recall	0.849	0.846	0.837	0.844	0.890	0.870	0.874	0.878
	mAP	0.873	0.875	0.847	0.865	0.884	0.882	0.880	0.882
INRIA Person Dataset	Precision	0.814	0.810	0.788	0.804	0.856	0.853	0.850	0.853
	Recall	0.850	0.844	0.829	0.841	0.864	0.865	0.860	0.863
	mAP	0.803	0.813	0.775	0.797	0.852	0.852	0.846	0.850
自建矿井人员检测数据集	Precision	0.879	0.869	0.853	0.867	0.895	0.894	0.893	0.894
	Recall	0.831	0.839	0.826	0.832	0.869	0.860	0.863	0.864
	mAP	0.863	0.853	0.831	0.849	0.883	0.883	0.880	0.882

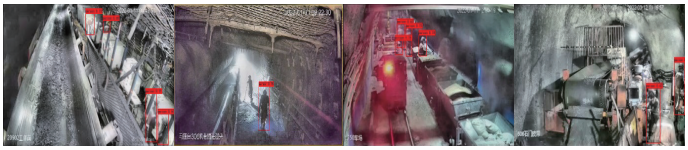
通过定量分析表 5 中三组数据集中 YOLOv7 算法与本文算法的性能指标，得出 YOLOv7 算法小目标检测性能均明显低于对大中目标的检测性能，而本文算法针对小目标检测方面的改进有着显著的效果，缩短了三种尺寸检测性能之间的差距，且检测精度能够基本持平；本文算法整体性能在不同数据集上验证也均优于 YOLOv7 算法。综上所述，YOLOv7 模型能实现对大、中目标的检测，但检测尺度不够宽广，特征提取不够精细，不适合矿井小目标及遮挡情况的检测，而本文算法进行针对性改进后不仅适用于矿井人员小目标检测，在多尺度变化、目标形变，雾霾干扰、光照剧烈、部分遮挡等场景中检测效果也均优于 YOLOv7 算法，因此，本文算法具有良好的泛化性与小目标检测性能。

为了更加直观的展示本文算法的检测效果，将本文算法与当前四种主流算法在矿井不同场景中进行检测测试，检测结果如图 7 所示。

(a) Master map

(b) Faster-RCNN

(c) SSD



(d) YOLOv5

(e) YOLOv7

(f) Our

图 7　主流算法检测结果

从第一张图片中观察到光照不均现象严重，Faster-RCNN、SSD、YOLOv7 均出现漏检现象，而本文算法引入 Swin-Transformer 基于全局感受野进行增强，对于遮挡现象达到了良好的检测效果；从第二张图片中观察到煤尘干扰现象严重，Faster-RCNN、YOLOv5、YOLOv7 均出现漏检现象，其中 YOLOv7 出现了误检情况，而本文注意力模块的添加，提升了目标在复杂环境中的对比度，有效的抑制了煤尘的干扰。从第三张图片中观察到目标较小，Faster-RCNN 出现漏检现象，SSD、YOLOv5、YOLOv7 表现良好，本文算法从增强小目标注意力出发，在多尺度融合阶段中层和低层引入 ACmix 模块帮助网络重点关注小目标特征，兼顾输入与输入之间的关系及输入与输出之间的关系，减少了漏检情况的发生。从第 4 张图片中观察到目标遮挡情况严重，SSD、YOLOv5 出现漏检现象，YOLOv7 出现误检情况，而本文算法检测效果良好，并且本文 EIOU 的引入减少了目标框与先验框的宽度和高度的差值，使得定位更加精准。综上所述，本文算法在井下多种复杂环境中检测效果良好，与当前主流目标检测算法相比更适用于矿井人员检测。

3　结束语

本文提出了一种改进 YOLOv7 的矿井人员检测算法，在 YOLOv7 的基础上，使用 ShuffleNetV2 轻量化主干，同时在 shuffle_block 中引入 Swin Transformer 注意力机制，保持一定精度的同时降低了模型计算量；在多尺度融合阶段中层和低层引入 ACmix 模块帮助网络关注小目标特征，提升模

型对小目标的检测能力；引入 EIOU 损失减小目标框和先验框高度与宽度的差值，实现更加精准的定位。

利用自建矿井人员检测数据集对本文算法进行验证，结果表明，本文算法准确率达 89.4%。检测速率达 68.8FPS，满足井下人员实时检测的要求，为矿井安全生产提供了良好的保障，对于煤矿开采向智能化开采过渡具有重要意义。

参 考 文 献

[1] ZHANG K, KANG L, CHEN X, et al. A review of intelli gent unmanned mining current situation and development trend[J]. Energies, 2022, 15(2): 513.

[2] 单鹏飞, 李晨炜, 来兴平等. 模拟暗湿工况下煤矸混合体态势热敏图像精准辨识实验[J/OL]. 煤炭学报: 1-12 [2023-04-14]. SHAN Pengfei, LI Chenwei, LAI Xingping, et al. Experiment on Accurate identification of thermal image of coal-gangue mixture under a simulated dusky and wet condition[J/OL]. Journal of china Coal Society, 1-12[2023-04-14].

[3] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2980-2988.

[4] ZHU H G. An efficient lane line detection method based on computer vision[J]. Journal of Physics:Conference Series, 2021, 1802(3): 032006.

[5] LI X Y, WANG S, LIU B, et al. Improved YOLOv4 network using infrared images for personnel detection in coal mines[J]. Journal of Electronic Imaging, 2022, 31(1): 013017.

[6] DU Y X, TONG M M, ZHOU L L, et al. Edge detection based on Retinex theory and wavelet multiscale product for mine images[J]. Applied Optics, 2016, 55(34): 9625-9637.

[7] 寇发荣, 肖伟, 何海洋等. 基于改进 YOLOv5 的煤矿井下目标检测研究[J]. 电子与信息学报, 2023, 45(07): 2642-2649.
KOU Farong, XIAO Wei, HE Haiyang, et al. Research on tarage detection in underground coal mines based on improved YOLOv5[J]. Journal of Electronics & Information Technology, 2023, 45(07): 2642-2649.

[8] 邵小强, 李鑫, 杨涛等. 改进 YOLOv5s 和 DeepSORT 的井下人员检测及跟踪算法[J]. 煤炭科学技术, 2023, 51(10): 291-301.
SHAO Xiaoqiang, LI Xin, YANG Tao, et al. Underground personnel detection and tracking based on improved YOLOv5s and DeepSORT[J/OL]. Coal Science and Technology, 2023, 51(10): 291-301.

[9] 李江昀, 赵义凯, 薛卓尔, 等. 深度神经网络模型压缩综述[J]. 工程科学学报, 2019, 41(10): 1229-1239.
LI J Y, ZHAO Y K, XUE Z E, et al. A survey of model compression for deep neural networks[J]. Chinese Journal of Engineering, 2019, 41(10): 1229-1239.

[10] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[EB/OL]. [2023-06-01]. http://arxiv.org/abs/2207.02696.

[11] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//European Conference on Computer Vision. Cham: Springer, 2018: 122-138.

[12] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021: 9992-10002.

[13] PAN X R, GE C J, LU R, et al. On the integration of self-attention and convolution[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2022: 805-815.

[14] CHEN C Y, LIU M Y, TUZEL O, et al. R-CNN for small object detection[C]//Asian Conference on Computer Vision. Cham: Springer, 2017: 214-230.

[15] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506(C): 146-157.

[16] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 779-788.

[17] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 6517-6525.

[18] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. [2018-04-08] https://arxiv.org/abs/1804.02767

[19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. [2023-06-01]. http://arxiv.org/abs/2004.10934.

[20] SONG Q S, LI S B, BAI Q, et al. Object detection method for grasping robot based on improved YOLOv5[J]. Micromachines, 2021, 12(11): 1273.

[21] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-style ConvNets great again[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 13728-13737.

[22] 张旭, 周云成, 刘忠颖, 等. 基于改进 ShuffleNet V2 模型的苹果叶部病害识别及应用[J]. 沈阳农业大学学报, 2022, 53(1): 110-118.
ZHANG X, ZHOU Y C, LIU Z Y, et al. Identification and application of apple leaf diseases based on improved ShuffleNet V2 model[J]. Journal of Shenyang Agricultural University, 2022, 53(1): 110-118.

[23] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 4510-4520.

[24] XIA Z F, PAN X R, SONG S J, et al. Vision transformer with deformable attention[EB/OL]. [2023-06-01]. http://arxiv.org/abs/2201.00520.
[25] NIU Z Y, ZHONG G Q, YU H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.
[26] WANG C Y, YEH I H, LIAO H Y M. You only learn one representation: Unified network for multiple tasks [EB/OL]. [2023-06-01] . http://arxiv.org/abs/2105.04206.
[27] RISHAV, SCHUSTER R, BATTRAWY R, et al. ResFPN: Residual skip connections in multi-resolution feature pyramid networks for accurate dense pixel matching [C]//Proceedings of the 25th International Conference on Pattern Recognition. New York: IEEE, 2021: 180-187.
[28] ZHENG Z H, WANG P, REN D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2022, 52(8): 8574-8586.
[29] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506(C): 146-157.
[30] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. New York: IEEE, 2017: 1137-1149.
[31] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

编 辑 张 莉