第 49 卷 第 6 期：2442-2457 高电压技术 Vol.49, No.6: 2442-2457
2023 年 6 月 30 日 High Voltage Engineering June 30, 2023
DOI: 10.13336/j.1003-6520.hve.20222039

面向电力系统现场作业的安全风险管控智能检测算法

何 敏[1]，秦 亮[1]，赵 峰[2]，余金沄[1]，刘浩锋[1]，王秋琳[3]，徐兴华[4]，刘开培[1]
（1. 武汉大学电气与自动化学院，武汉 430072；2. 国网信息通信产业集团有限公司，北京 102211；3. 福建亿榕信息技术有限公司，福州 35300；4. 海军工程大学舰船综合电力技术国防科技重点实验室，武汉 430034）

摘 要：针对电力现场作业下的安全管控存在场景复杂、目标多样且部分遮挡而导致智能安全监测困难的问题，提出一种基于 YOLOv7-Tiny 的改进算法。首先，搭建了 YOLOv7-Tiny 检测网络，并在该算法框架中融合通道重组的注意力机制，从而有效提升通道之间的交互能力，增强复杂场景下目标区域的显著度；其次，在特征融合阶段，构建基于残差跳连的多尺度特征融合结构 Res-PANet(Residual Path-Aggregation Network)来有效融合多尺度目标，提升场景中的多目标检测能力；同时，在模型的输出检测头中结合 Swin-Transformer 模块，提升模型的感受野，增强模型对特征图的全局感知，提高模型在部分遮挡情况下的检测能力；接着，在训练时采取改进的 Mosaic 数据增强方式，提升小目标的分布数量，达到丰富目标场景、提高模型泛化能力的目的；最后，以电力人员安全帽及安全服的穿戴、电力围栏以及电力警示牌为安全作业的监测对象进行改进算法的验证，同时采取基于 Score-CAM 的热力图分析进一步验证模型改进的有效性。实验结果表明：融合改进模型的平均检测精度达 90.1%，图像检测速度为 46 帧/s，在嵌入式硬件 Jetson NX 上测试推理延时为 75 ms，能有效满足电力安全现场检测精度和检测速度的要求。

关键词：电力安全；智能监测；YOLOv7-Tiny 网络；Shuffle-Attention 机制；目标遮挡检测

Intelligent Detection Algorithm of Security Risk Management and Control for Power System On-site Operation

HE Min[1], QIN Liang[1], ZHAO Feng[2], YU Jinyun[1], LIU Haofeng[1], WANG Qiulin[3], XU Xinghua[4], LIU Kaipei[1]
(1. School of Electrical and Automation, Wuhan University, Wuhan 430072, China; 2. State Grid Information & Telecommunication Group Co., Ltd., Beijing 102211, China; 3. Fujian Yirong Information Technology Co., Ltd., Fuzhou 35300, China; 4. National Key Laboratory of Science & Technology on Vessel Integrated Power System, Naval University of Engineering, Wuhan 430034, China)

Abstract: Aiming at the problem of complex scenes, diverse targets and difficulties in intelligent safety monitoring due to partial occlusion of safety control under power site operations, we propose an improved algorithm based on YOLOv7-Tiny. Firstly, the YOLOv7-Tiny detection network is built, and the attention mechanism of channel reorganization is fused in this algorithm framework to effectively improve the interaction ability between channels and to enhance the saliency of target regions in complex scenes. Secondly, in the feature fusion stage, Res-PANet, a multi-scale feature fusion structure based on residual hopping, is constructed to effectively fuse multi-scale targets and to improve the multi-target detection capability in the scene. At the same time, the Swin-Transformer module is combined in the output detection head of the model to enhance the perceptual field of the model, to achieve enhanced global perception of the feature map by the model, and to improve the detection ability of the model in the case of partial occlusion. Then, an improved Mosaic data enhancement is adopted during training to enhance the number of small target distributions, to achieve the purpose of enriching the target scenes, and to improve the generalization ability of the model. Finally, the wearing of safety helmets and safety clothing of electric personnel, electric fences and electric warning signs are taken as the monitoring objects of safety operations for the verification of the improvement algorithm, and the heat map analysis based on Score-CAM is also adopted to further verify the effectiveness of the model improvement. The experimental results show that the average detection accuracy of the fusion-improved model can reach 90.1%, the image detection speed

基金资助项目：国家重点研发计划 (2020YFB0905900)。
Project supported by National Key R&D Program of China (2020YFB0905900).

is 46 frame/s, and the test inference delay is 75 ms on the embedded hardware Jetson NX, which can effectively meet the requirements of power safety field detection accuracy and detection speed.

Key words: power security; intelligent monitoring; YOLOv7-Tiny model; Shuffle-Attention mechanism; object occlusion detection

0　引言

电力生产安全对电力系统的稳定运行至关重要，其中，由于电力生产工作中的不规范操作导致的安全事故占相当大的比重，严重威胁了生产人员的安全和电力系统的正常运行[1-2]。因此，准确快速地识别出现场作业的违规风险，提高电力现场作业的安全水平，对电力生产安全意义重大。

近年来，随着人工智能技术的发展，尤其是在计算机视觉领域，以 YOLO(You Only Look Once)系列模型为典型代表的目标检测算法发展到第 7 代，即 YOLOv7 模型，实现了模型在推理速度和精度上的进一步优化，更加有利于满足电力系统中目标场景下高精度与高效率检测的需求。在这一类电力视觉应用场景中既包含输电线路上设备故障的识别与定位[3-6]，同时在电力现场作业安全监测中也被有效利用：如安全帽的检测[7]，人脸的监测[8]、人体安全行为识别等[9-10]，极大地提升了电力系统的智能化管控能力。

在电力安全管控方面，相关研究对象主要集中在安全帽、安全手套、绝缘服和安全标识牌等静态的通用安全着装及防护装置上。其中，文献[11]针对电力生产影像分辨率低而导致检测效果不佳的问题，提出一种基于稀疏表示模型的 SSD(Single Shot Multi-Box Detector)目标检测方法，进行图像的超分辨率增强，实现了现场人员是否佩戴安全帽、安全带的检测。本文主要采取的图像分辨率为 352×288，对于分辨率较大的场景模型的泛化能力不够，同时文献[11]主要针对小目标检测的优化，没有进行多尺度特征融合，使得模型在多目标场景的检测效果较差；文献[12]采用 YOLOv3 目标检测算法，根据检测对象的特征来实现对安全帽、安全带的检测；文献[13]首先通过伽马校正的数据增强方式来提升图像的细节，并结合 K-means++聚类方式来提升边界框的定位准确率，最后用改进的 YOLOv3 算法实现安全帽，安全手套和绝缘靴的有效监测。在以上 2 个研究中，由于 YOLOv3 算法本身的检测缺陷并不能有效解决目标被部分遮挡的问题，且该模型相比较于后续的 YOLOv5、v6 等存在占用内存更多、参数量更大的缺点，不利于目标场景的高效检测；文献[14]采用 YOLOv4-Tiny 并结合卷积注意力模块(convolutional block attention module，CBAM)和有效的通道注意力模型(efficient channel attention module，ECA)的注意力机制有效升模型的准确率，实现对安全帽和安全服的检测。该文献主要针对结构被大量剪枝后的 YOLOv4 模型在检测精度上的优化，并没有针对场景复杂、多目标等主要的检测难点进行特定场景的分析与优化，因此该模型对本文任务场景并不具有较好适用性；文献[15]提出以 ResNet50(Residual Network)作为 YOLOv5 模型的骨干，来监测电网现场作业下的安全帽、安全服、杆塔等。残差结构能弥补特征提取过程中的信息丢失，而该文献中的场景检测多为背景单一的天空图像，对于高相似的伪目标和复杂场景并没有被有效分析与解决；文献[16]通过引入 YOLOX 目标检测模型，并采取 Mosaic 和 Mix-Up 的数据增强方式，提升模型在复杂作业环境下安全带穿戴检测的鲁棒性。过度的使用数据增强会导致数据的失真，使得模型训练的权重参数不利于自然场景的检测，针对此问题本文在后续具体实验中给出了解决方法。

综上所述，虽然电力安全作业检测已有一定程度的研究，然而当检测对象出现场景复杂、伪目标多、部分遮挡等情况时，上述方法往往难以实现对目标区域的有效定位，尤其是多目标的识别能力下降，导致误检、漏检问题严重。因此，本文聚焦以上问题，以 YOLOv7-Tiny 为基础，设计了以电力安全帽和电力安全服的穿戴与否、电力信号指示牌以及电力围栏为监测对象的目标检测算法。本文主要创新与贡献如下：

1）针对复杂背景难检测问题，采用特征关注的方式来学习特征中目标像素级的成对关系和通道之间的依赖性，以此减弱背景特征的比重。同时利用通道重组减轻多通道之间的信息隔离，进一步加强各通道的联系，整体实现模型在复杂环境中的特征提取能力。

2）针对多尺度目标存在误检漏检问题，采用

设计的 Res-PANet 结构，在原同尺度下融合的基础上，设计将几何信息表征能力强的高分辨率特征与语义信息表征能力强的低分辨率特征，通过残差连接的方式进行融合，保留原始信息同时提高模型的收敛能力，提升模型对于安全作业下的多目标检测能力。

3）针对目标被部分遮挡的问题，在输出检测头中，设计具有全局感知能力的 Swin-Transformer 模块，利用基于滑窗和偏移滑窗的多头自注意力机制去进行特征的相似度度量，挖掘特征潜能，以此有效提升半遮挡目标的检测效果。

4）在训练的过程中，采取改进的 Mosaic 数据增强方式，丰富电力安全作业场景的变化性，提升模型的鲁棒性。

1 YOLOv7-Tiny 模型及其改进算法

1.1 YOLOv7-Tiny 模型

2022 年 Wang 等人提出 YOLOv7-Tiny 算法模型，其精度相比 YOLOv3、YOLOv4、YOLOv5 等有了巨大提升，且检测速度更快。整体结构如图 1 所示，该模型结构主要包括主干特征提取网络(Backbone)、特征融合网络(Neck)以及特征检测输出层(Head)3 个部分构成。

在 Backbone 中，卷积小组件模块主要由基础卷积块 Conv2D、批量归一化层(batch normalization，BN)及带泄露线性整流激活函数层(leaky rectified linear units，Leaky-ReLU)构成；Max Pooling 是最大池化层，主要用于特征维度的降低；SPPCSPC 是融合了跨阶段局部网络(cross stage partial network, CSP-Net)的金字塔池化结构，如图 1 的右上角所示，其主要是将特征金字塔池化(spatial pyramid pooling，SPP)结合小残差结构时，还引入了大的残差边，将 SPP 前一层的输出经过卷积模块后直接与小残差结构的输出进行通道的拼接，从而提取出更多有用的特征信息。同时相比较以前的 YOLO 系列算法，YOLOv7-Tiny 最大的区别在于采用了高效层聚合网络(efficient layer aggregation networks，ELAN)结构，即高效的神经网络聚合结构，在不破坏原始梯度路径的情况下，提高了网络的特征提取效率[17]。

在 Neck 中，采用路径聚合网络(path aggregation network，PA-Net)结构，通过将浅层与深层特征进行融合达到增强特征信息的目的。

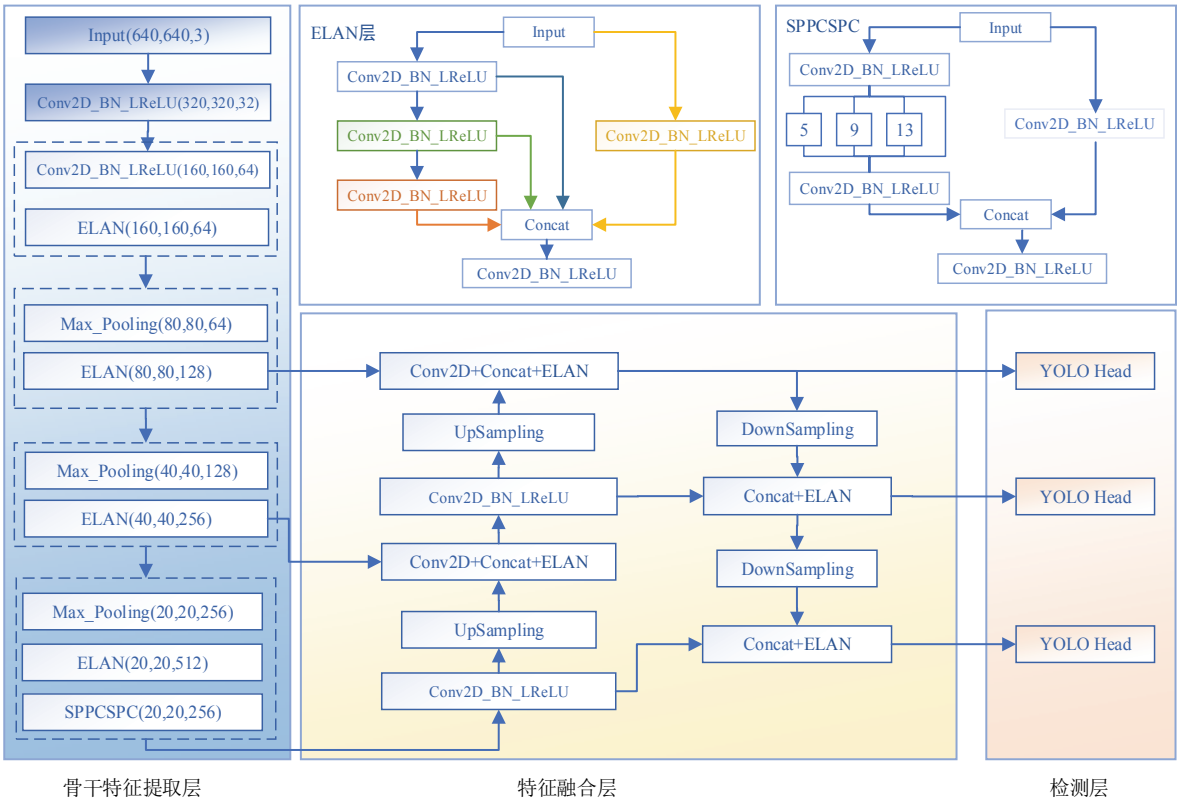


图 1 YOLOv7-Tiny 模型结构图

Fig.1 YOLOv7-Tiny model diagram

在 Head 中，通过 1×1 的卷积输出目标的定位信息、分类信息及置信度信息。并通过这 3 个信息的损失实现模型对目标区域在分类和定位上的优化，损失计算式为

$$L_{\text{cls}}=\sum_{t=0}^{S\times S}\sum_{j=0}^{n_B}I_{ij}^{\text{obj}}\sum_{c\in\text{classes}}(\hat{p}_i(c)\ln(p_i(c)))+\sum_{t=0}^{S\times S}\sum_{j=0}^{n_B}I_{ij}^{\text{obj}}\sum_{c\in\text{classes}}((1-\hat{p}_i(c))\ln(1-p_i(c))) \tag{1}$$

式(1)为模型的分类损失。$S\times S$ 为图像输入的尺寸 640×640；I_{ij}^{obj} 表示特征图；i 表示第 i 个方格，j 表示方格预测出的第 j 个预测框；n_B 为预测框数量；$c\in$classes 表示所属正确的类别；$p_i(c)$ 和 $\hat{p}_i(c)$ 分别表示预测目标置信度和实际目标置信度得分。图 2 为 CIoU 损失函数的计算图，图中 A 框为真实框，B 框为预测框。

式(2)为目标框的定位损失，主要采用的是 CIoU 损失函数。

$$\begin{cases}S_{\text{IoU}}=\dfrac{A\cap B}{A\cup B}\\ v=\dfrac{4}{\pi^2}(\arctan\dfrac{w^{gt}}{h^{gt}}-\arctan\dfrac{w}{h})^2\\ \alpha=\dfrac{v}{(1-S_{\text{IoU}})+v}\\ L_{\text{box}}=1-S_{\text{IoU}}+\dfrac{\rho^2(A,B)}{c^2}+\alpha v\end{cases} \tag{2}$$

式中：S_{IoU} 为预测框与真实框的交并比；M 框为包含 A 框和 B 框的最小外接矩形；α 为权重函数。其中 $\rho^2(A,B)$为真实框与预测框的中心点之间的欧氏距离，即图中 d 的长度；c 为包围 A、B 框的最小外围矩阵 M 框的对角线长度；w^{gt} 和 h^{gt} 为真实框 A 的宽和高；w 和 h 为预测框 B 的宽和高。通过 CIoU 损失中的惩罚项 v 能有效解决对于大小不同的框，在交并比值相同时，但重叠度却不同的情况即尺度敏感性的问题，能有效剔除差异性过大的负样本框。最后得到的 L_{box} 即为总的回归框损失函数

$$L_{\text{conf}}=\sum_{i=0}^{S\times S}\sum_{j=0}^{n_B}I_{ij}^{\text{nobj}}(\hat{C}_i\ln(C_i)+(1-\hat{C}_i)\ln(1-C_i))-\sum_{i=0}^{S\times S}\sum_{j=0}^{n_B}I_{ij}^{\text{obj}}(\hat{C}_i\ln(C_i)+(1-\hat{C}_i)\ln(1-C_i)) \tag{3}$$

式中：L_{conf} 为目标的置信度损失函数；obj 和 nobj 表示第 i 个方格中的目标有和无；C_i 和 $\hat{C}_i$ 表示预测框和真实框的类别。

总的损失由式(1)—式(3)构成，即

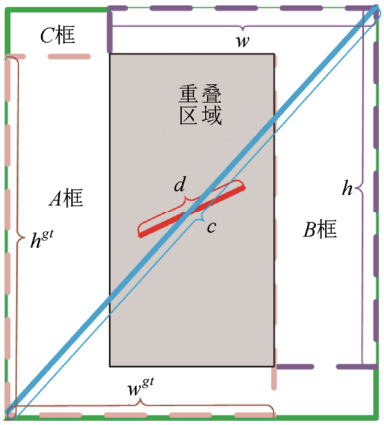


图 2　CIoU 的计算图

Fig.2　Calculation diagram of CIoU

$$L_{\text{total}}=0.5L_{\text{cls}}+0.05L_{\text{box}}+L_{\text{conf}} \tag{4}$$

在预测时，经过非极大值抑制等处理操作后消除大量冗余预测框，最后输出置信度得分最高的预测类别，并定位目标返回目标的坐标信息。

1.2　基础算法对比与改进算法设计

本文首先进行了常用目标检测算法对比实验，实验结果见表 1。表 1 中第 1 列为实验对比模型，第 2—7 列为实验检测 6 个目标的平均精度(average precision，AP)值，第 8 列为 6 个目标的平均 AP 值(mAP)，第 9 列为模型的大小对比、最后 1 列为各个模型的检测速度对比。其中 6 个目标的标签名分别为 aqm(佩戴安全帽)、wdaqm(未戴安全帽)、gzf(身穿工作服)、wcgzf(未穿工作服)、signal(信号指示牌)和 fence(电力围栏)。

1）在模型大小上，YOLOv7-Tiny 具有最小的模型体积，主要由于采取的 ELAN 结构相比较原始 YOLO 系列的 CSP-Net 的残差连接更加高效简洁。便于后续模型的应用和部署。

2）从检测速度上，SSD 算法具有更高的检测速度，主要原因是 SSD 没有进行多尺度的特征融合，使得推理速度相对更快，但是会导致模型对于多尺度目标的检测能力不够好，通过均值平均精度(mean average precision，mAP)值也可以看到该模型的检测精度最低。而 YOLOv7-Tiny 的检测速度仅比其低 4 帧/s，在检测速度上同样具有较好的检测优势。

3）在检测准确率上，YOLOv7-Tiny 具有最高的检测效果，其 mAP 为 86.9%，但可以发现 YOLOv7-Tiny 模型对于未戴安全帽(wdaqm)、电力信号指示牌(signal)以及电力围栏(fence)具有较低的检测准确率。

表 1 目标检测算法对比

Table 1 Comparison of target detection algorithms

模型	AP/%						mAP/%	模型大小 /MB	检测速度/(帧·s^{-1})
	aqm	wdaqm	gzf	wcgzf	signal	fence			
SSD	79.63	55.63	79.07	86.53	51.40	82.45	72.45	93.1	71.39
Faster R-CNN	79.81	55.34	82.73	90.24	55.46	80.34	73.99	108.0	9.78
YOLOv3	91.25	79.91	89.62	91.99	65.11	82.25	83.36	235.0	42.41
YOLOv4	93.18	81.96	89.00	89.74	70.78	83.90	84.76	244.0	36.50
YOLOv5s	92.27	77.12	87.66	90.31	70.43	80.18	82.99	27.2	49.45
YOLOXs	92.54	79.37	88.83	93.15	69.91	85.47	84.88	34.3	52.53
PP-YOLOv2s	82.13	73.26	78.44	88.13	70.92	82.35	79.21	16.2	56.43
YOLOR	95.60	86.40	91.10	94.60	68.30	82.40	86.40	105.5	38.94
YOLOv6s	95.20	84.50	90.50	91.20	72.50	84.60	86.60	37.9	54.89
YOLOv7-Tiny	94.60	82.80	92.40	94.40	72.40	84.60	86.90	12.0	67.88

4）总体来看，YOLO 系列的算法相比较于 SSD 和 Faster RCNN 具有检测精度上的优势，而对比 YOLO 系列算法，随着版本的创新，相较于 YOLOv3 和 YOLOv4，后续的 YOLO 算法，如 YOLOv5、YOLOv6 和 YOLOv7 等模型具有明显的模型大小和检测速度上的优势，其中 YOLOv7-Tiny 具有最好的检测精度和最高的检测速度，能为后续模型的进一步部署与优化提供借鉴方案。

同时进一步分析实验数据发现，主要由于图 3 中列出的数据样本情况导致的检测效果不佳。具体情形如下：存在检测对象与背景相似的场景，如图 3(a)安全帽和安全服与周围场景颜色相似导致漏检或误检的情况；图 3(b)存在电力信号指示牌、电力围栏等大小且形状不一的目标，使得目标的判定没有像安全帽一样统一的标准；图 3(c)同时在人员较多的场景存在部分遮挡的情况导致目标出现漏检；图 3(d)存在部分拍摄图像背景较为单一，模型泛化能力不够。

综合以上问题分析，本文的整体改进思路与章节安排如下：

1）首先在第 2 章中分析考虑到复杂背景主要是由于在特征提取时容易将目标特征与背景特征进行无差别提取，于是设计特征提取加权优化的方式，通过在特征层面和通道层面进行背景抑制，从而突出目标特征区域提高特征的可用性，从通道优化层面判断目标的有无，从特征层角度更加准确定位目标的位置，同时结合通道重组的方式描述空间和通道中的特征联合关系，在像素级上实现不同子特征之间的信息融合，提升目标特征的辨识度。

2）接着在第 3 章中分析针对电力信号指示牌、电力围栏多目标且大小不一的，场景检测容易出现目标特征丢失而漏检的问题，采取在多尺度特征融合层进行高分辨率特征的复用表达，由于高分辨率特征包含更多的原始信息，通过结合高分辨率上的特征弥补低分辨率上特征的丢失问题，实现多目标的有效检测与定位。



(a) 与背景相似的安全服 (b) 多目标类场景

(c) 密集遮挡目标 (d) 单一背景

图 3 数据集样本

Fig.3 Dataset samples

3）其次在第 4 章中分析针对潜在的密集遮挡目标，主要由于卷积的方式提取特征仅考虑局部感受野，而对全局特征信息没有充分的认识，使得被部分重叠的目标容易被前景目标的识别而漏检，于是设计了一种混合的特征提取结构用来增强检测头的输出效果，即利用 Swin-Transformer 支路进行全局信息定位与度量，利用卷积支路进行局部特征细化，将两者特征融合来有效挖掘潜在目标。

4）最后从数据集的角度出发，通过改进的数据增强方式，增加小目标数据集，丰富目标场景，提升模型的鲁棒性。

改进后的 YOLOv7-Tiny 算法整体结构如图 4 所示。图 4 中，首先采取改进后的 ELAN-SA 模块

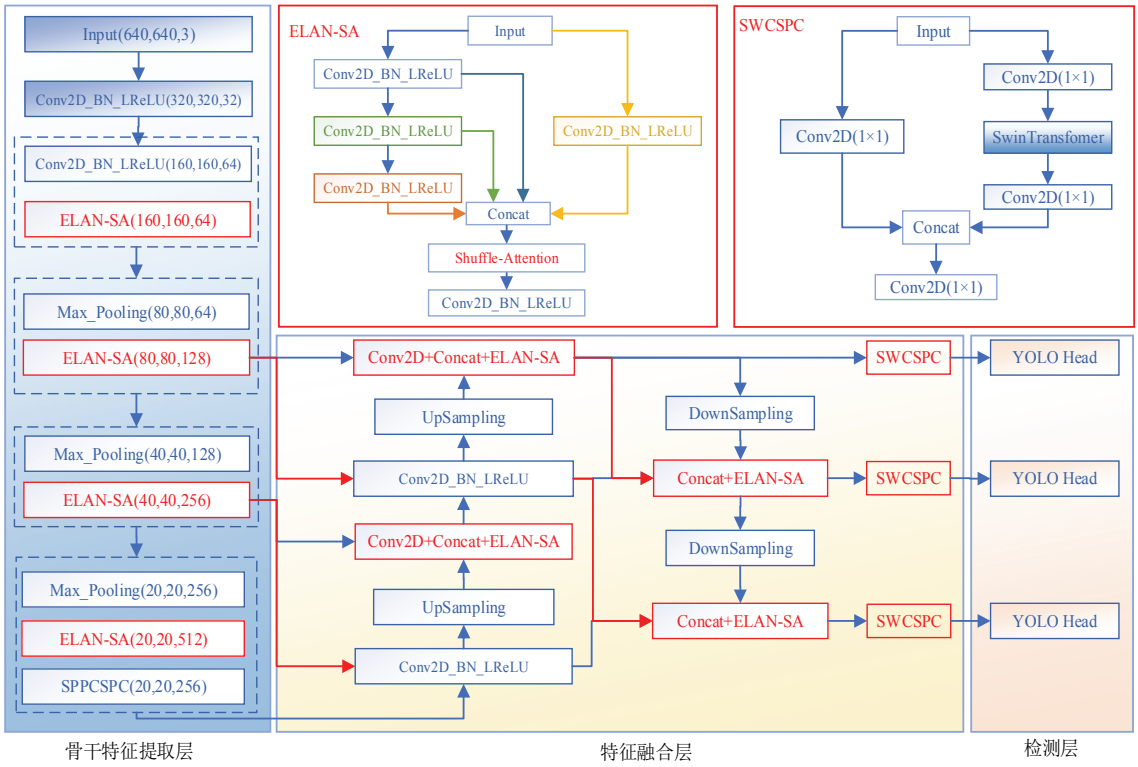


图 4　改进的 YOLOv7-Tiny 网络结构对比图

Fig.4　Improved YOLOv7-Tiny network structure comparison chart

代替原始的 ELAN 模块进行模型的特征提取，主要是将通道直连的模块进行通道的关注与重组。其次在特征融合区域针对原始的 PANet 结构，采取解码器端即上采样端前的特征融合，编码器端即下采样端后的特征融合，在特征融合过程中同样采取 ELAN-SA 模块进行特征的加强提取。最后是在输出层采取 Swin-Transformer 结构下的跨阶段局部网络进行特征的输出检测。

同时，本文所设计方法的整体计算流程如图 5 所示。图 5 中主要是将采集的数据进行模型的输入训练，并在训练过程中进行模型损失、权重参数的优化与更新以及模型性能的测试，最后训练完成保存最优结果。在模型泛化能力验证时，选取测试集上的数据进行验证，一方面得到测试集下模型的各项性能指标，另一方面进行图像检测速度的分析和检测结果的可视化。

2　基于特征关注与通道重组的复杂场景检测

YOLOv7-Tiny 在利用 ELAN 模块进行特征提取时，采取 4 个特征进行通道的拼接，如图 2 中的 ELAN 所示，由于每个通道都包含不同的特征信息，这种直接拼接的方式容易让各个卷积提取到的特征通道形成隔离，导致模型对于复杂场景的目标检测效果较差，主要是由于特征提取时并不能有效判定目标的有无和目标的区域定位。

为解决以上问题，本文首先采用分组卷积的方式来扩展特征块的计算基数，使得特征信息被有效表征，同时也能有效减少模型的计算量；其次建立针对目标及其位置信息的关注，一方面通过在通道之间建立特征映射关系，使网络充分利用这些通道信息赋予目标通道更高的权重，有效确定目标的有无；另一方面，在特征图层面，进行特征图区域的关注，捕获成对的像素级关系，更好地定位目标所在位置，以此实现复杂背景下的目标检测。最后，采取通道交换的方式，使得特征通道具有更强的鲁棒交互性[18]。

图 6 是具体计算结构图，首先会将输入的特征图($H\times W\times C$，其中 H、W 和 C 分别是特征的长、宽和通道数)按照通道维度进行分组操作，有利于通道的分散和模型计算量的减小。再将特征进行 g 组的划分后，对每个分组的内部进行基于 Shuffle-Attention 的小单元处理。在 Shuffle-Attention 小单元中，其中一部分用于通道注意力机制的加权，另一部分用于空间注意力机制的加权。其中，通道

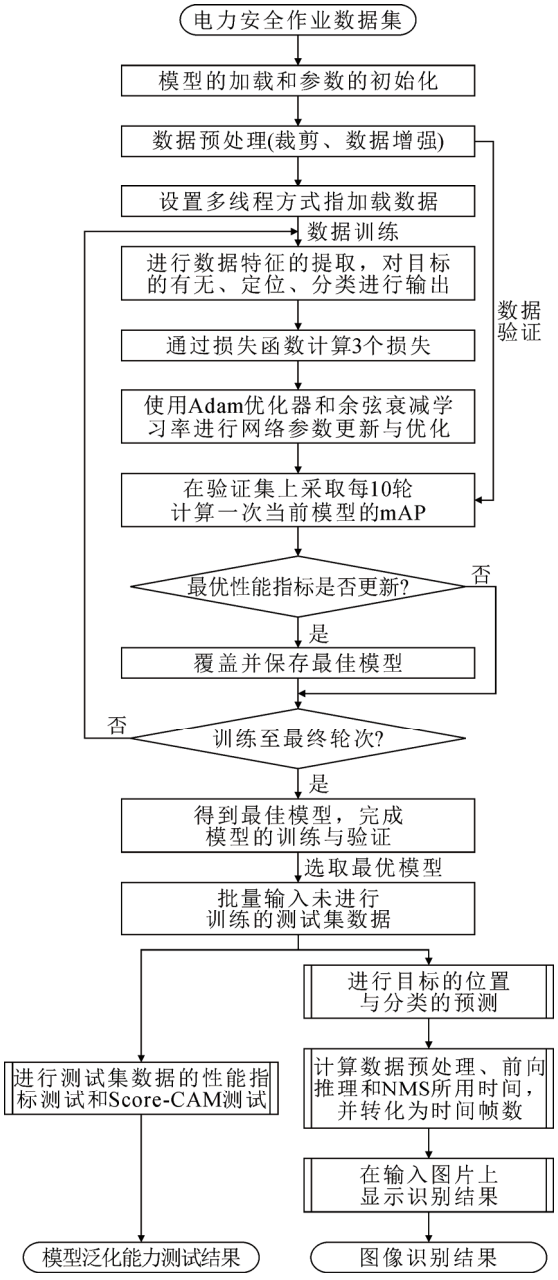


图 5 本文方法的整体计算流程

Fig.5 The overall calculation process of this method

注意力机制的加权方式如图 6 中间层上半部分的计算路线，首先采取自适应的平均池化方式基于通道维度的特征压缩，将压缩后的特征经过可学习参数 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 的线性变化，得到该特征的通道加权系数，最后将该系数与原始特征进行乘积即可得到通道特征加权后的特征，其计算式为

$$\begin{cases}\sigma(x)=\dfrac{1}{1+\mathrm{e}^{-x}}\\ \boldsymbol{F}_{\mathrm{out1}}=\boldsymbol{F}_{\mathrm{in1}}\sigma(\boldsymbol{w}\cdot\mathrm{Avg}(\boldsymbol{F}_{\mathrm{in1}})+\boldsymbol{b})\end{cases}\tag{5}$$

式中：$\boldsymbol{F}_{\mathrm{out1}}$ 为第 1 部分的输出特征；$\boldsymbol{F}_{\mathrm{in1}}$ 为第 1 部分的输入特征；Avg(*)为平均池化函数；$\boldsymbol{w}$ 和 $\boldsymbol{b}$ 为线性函数参数，通过该过程主要实现目标在通道层面上的优化与关注。

另一部分用于空间注意力机制的加权，如图 6 中间层下半部分的计算路线。主要采取组归一化(group normalize，GN)操作来获取空间统计数据，学习特征图中的像素分布，即潜在的目标分布情况。其主要是对每个分组进行数据归一化的均值和方差，由于 GN 层的计算与特征批量的大小无关，同时其准确性在很宽的批量大小范围内均比较稳定，能有效提升模型的训练效果[19]。GN 操作公式为

$$\begin{cases}S_i=\{k\mid k_N=i_N,[\dfrac{k_C}{C/G}]=[\dfrac{i_C}{C/G}]\}\\ \mu_i=\dfrac{1}{m}\sum\limits_{k\in S_i}x_k\\ \delta_i=\sqrt{\dfrac{1}{m}\sum\limits_{k\in S_i}(x_k-\mu_i)^2+\varepsilon}\\ \hat{x}_i=\dfrac{1}{\delta_i}(x_i-\mu_i)\\ y_i=\gamma\hat{x}_i+\beta\end{cases}\tag{6}$$

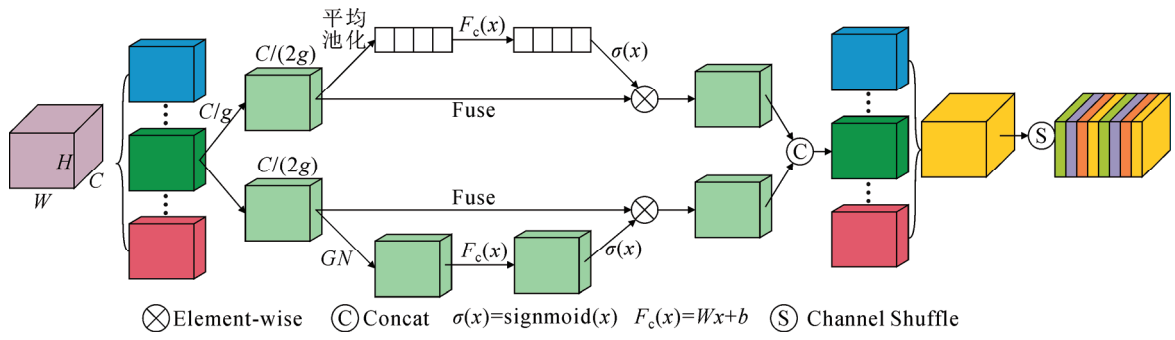


图 6 Shuffle-Attention 结构图

Fig.6 Shuffle-Attention structure diagram

式中：S_i 表示输入特征的像素的集合；k_N 和 i_N 分别表示 S_i 集合的长宽索引；C/G 表示分组卷积后的每组通道数；$\frac{k_C}{C/G}$ 和 $\frac{i_C}{C/G}$ 表示同一个通道下的长宽索引对应的值；μ_i 为 S_i 的均值；δ_i 为标准差；$\hat{x}_i$ 为归一化的特征，引入 γ 和 β 变量，对输入进来的数据进行缩放和平移。利用 γ 和 β 两个参数，让本文的网络可以学习恢复出原始网络所要学习的特征分布。y_i 为经过 GN 归一化的输出值。通过该过程主要实现模型对单通道特征层面上目标像素的分布。

后续操作与通道注意力机制一样，采取可学习参数 $\boldsymbol{w}$ 和 $\boldsymbol{b}$ 的线性变化，得到该特征的加权系数最后将该系数与原始特征进行乘积即可得到特征加权后的特征，其计算式为式(7)所示，GN(*)为式(6)的组归一化过程。

$$\boldsymbol{F}_{\text{out2}} = \boldsymbol{F}_{\text{in2}}\sigma(\boldsymbol{w}\cdot\text{GN}(\boldsymbol{F}_{\text{in2}})+\boldsymbol{b}) \tag{7}$$

最后，所有更新后的子特征会通过 Concat 方式被重新聚合。但此时的特征层中彼此之间并没有信息沟通，彼此独立，通道维与空间维的特征更新无法融合。因此，在输出前还需要以通道重组，如图 6 的最后部分，通过 Channel Shuffle 的方式去描述空间和通道中的特征联合关系，在像素级上实现不同子特征之间的信息融合，使得特征融合模块更加高效[20]。再结合前面的 ELAN 模块，构成了整个 ELAN-SA 模块，其结构如图 7 所示。

表 2 为改进的对比实验，第 1 列多个改进处的实验模型对比，第 2—7 列为 6 个检测目标的 AP 值，最后 1 列为 6 类目标的平均 AP 值对比。由表 2 可知，在 YOLOv7-Tiny 的 ELAN 模块中引入 SA 模块后，整体 mAP 相比较原始网络从 86.9%提升至 88.1%，该提升效果主要体现在工作服的穿戴情况，其中 gzf 的 mAP 提升 1.3%，wcgzf 的 mAP 提升 2.5%。主要是工作服相比较其他目标与周围环境的颜色相似度更高，改进后的模型能很好解决复杂背景下的伪目标检测问题。

3　基于残差跳连的 Res-PANet 多目标检测

特征金字塔结构被有效用于目标检测中的多尺度特征融合区域，在 YOLOv7-Tiny 模型中，采用 PANet 网络来进行目标检测中的多尺度特征融合。然而，随着图像的分辨率增大，且目标场景更加丰富时，通用的特征金字塔(feature pyramid networks，FPN)及 PANet 结构并不能有效发挥其潜力[21]。尤其

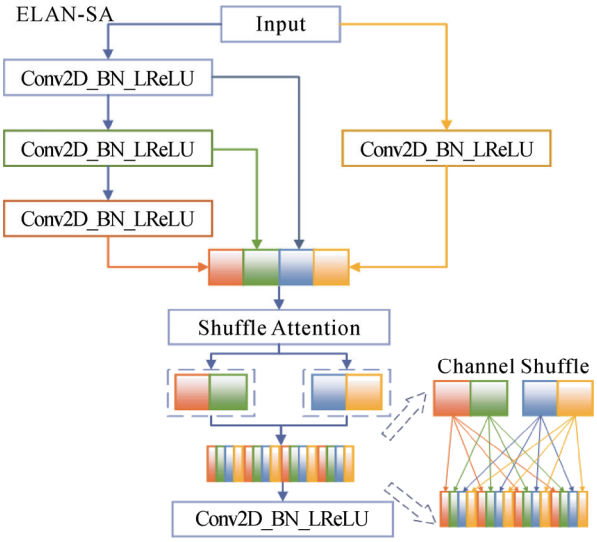


图 7　ELAN-SA 结构图

Fig.7　ELAN-SA structure diagram

表 2　模型每处改进的前后性能对比

Table 2　Before and after performance comparison of each improvement of the model

模型	AP/%						mAP/%
	aqm	wdaqm	gzf	wcgzf	signal	fence	
YOLOv7-Tiny	94.6	82.8	92.4	94.4	72.4	84.6	86.9
ELAN-SA	93.9	83.6	93.7	96.9	74.6	85.7	88.1
Res-PANet	94.7	86.0	91.8	96.1	72.8	85.9	87.9
SWCSP	94.6	84.4	93.5	94.5	74.1	85.5	87.7

是在特征提取至深层时，目标信息容易出现丢失而导致检测不到的问题。而对于特征融合结构部分目前已有的多目标检测方案是进行多支路下的特征加权[22]及增加多个骨干特征尺度输出进行特征融合[23]。该思路尽管使高分辨率下的特征信息的丢失问题被有效减轻，但由于多支路权重的分配，前者会导致梯度优化的减慢，由于检测尺度的增加，后者会导致模型计算量的增大。

基于此问题，本文受文献[21]针对高分辨率下多尺度融合的启发，提出一种 Res-PANet 的结构，该结构为多个残差跳跃连接的多分辨率特征金字塔网络，利用高分辨率下的特征信息获得更大和更好的特征信息。同时通过残差跳跃连接可减少反向传播期间梯度路径的长度，提高模型的收敛性[24]。且在特征融合阶段仍然保留 3 个尺度的模型计算，不会增大计算量。其总体设计思路是在金字塔网络中编码器和解码器之间增加横向跳跃连接，通过步长为 2 的卷积来进行特征尺度的统一，以提高空间精

度，同时保持高效特征表示。所设计的 Res-PANet 结构如图 8 所示。

图 9 中，把由小尺度特征向大尺度上采样的过程称为解码部分，把由大尺度特征向小尺度特征下采样的部分称为编码部分。其中解码器残差部分如图 9(a)所示，本文进行采样前的融合，分析其原因：主要由于上采样属于特征图还原的部分，通过弥补原始信息的不足才能更好地还原特征信息，其具体计算过程如下：

$$\boldsymbol{F}_{\text{decode}}=\text{Concat}(\text{UP}(\text{Concat}(\boldsymbol{F}_{\text{in1}},\boldsymbol{F}_{\text{in2}})),\boldsymbol{F}_{\text{in3}}) \quad (8)$$

式中：Concat(*) 表示特征的拼接融合；UP(*) 表示特征的上采样。

如图 9(b)所示，编码器部分采取下采样后的融合，分析其原因主要由于下采样是信息减少的部分，通过残差结构的引入能有效弥补信息的丢失。

$$\boldsymbol{F}_{\text{encode}}=\text{Concat}(\text{Concat}(\text{Down}(\boldsymbol{F}_{\text{in1}}),\boldsymbol{F}_{\text{in3}}),\boldsymbol{F}_{\text{in2}}) \quad (9)$$

式中 Down(*) 表示特征的下采样。

由表 2 可知，在 YOLOv7-Tiny 特征融合层中设计的 Res-PANet，其整体 mAP 相比较原始网络从 86.9%提升至 87.9%，该提升效果主要体现在安全帽的穿戴情况，其中 aqm 的 mAP 提升 0.1%，wdaqm 的 mAP 提升 3.2%。主要由于安全帽属于小目标的类型，通过本文所设计的残差连接结构，有效弥补了特征提取过程中小目标信息的丢失，提升了其检测精度。

4 基于全局感受野增强的 Swin-Transformer 遮挡检测

当目标出现部分被遮挡时，会导致目标出现漏检的情况，主要原因是由于在输出检测时卷积仅有局部信息的认识，而对于全局感受较差，最后没有给予部分被遮挡目标较好的定位区域和高的置信度得分。由于 CNN 结构具有更好的局部效果，而 Transformer 类型的结构则会具有更高的全局感受效果，而传统的 Transformer 结构由于其计算量较大，本文采取轻量化的 Swin-Transformer 进行特征的提取，并借鉴 CSPNet 的跨阶段特征增强提取方式，设计了一种混合的特征提取结构用来增强检测头的输出效果，左侧支路保留卷积的局部特征感知，右侧支路则采取全局信息度量，通过这一特征细化模块用来增强部分被遮挡目标的检测效果，将其简写为 SWCSP(Swin-Transformer-CSPNet)。

图 10 中，左边支路为卷积结构的残差，右边支路为基于 Swin-Transformer 的瓶颈结构，最后进行 2 个支路的特征融合。在 Swin-Transformer 结构中，该模块主要通过移动窗口学习目标特征，同时通过多头自注意力机制，来进行特征图的相似度学习，提升特征的可用性[25]。其中 Swin-Transformer 的计算方式如下：

$$\begin{cases}\boldsymbol{F}_{\text{att}}=\text{LN}(\boldsymbol{F}_{\text{in}}+\text{W}-\text{MSA}(\boldsymbol{F}_{\text{in}}))\\ \boldsymbol{F}_{\text{mid}}=\text{LN}(\boldsymbol{F}_{\text{att}}+\text{MLP}(\boldsymbol{F}_{\text{att}}))\\ \boldsymbol{F}_{\text{attW-M}}=\text{LN}(\boldsymbol{F}_{\text{mid}}+\text{SW}-\text{MSA}(\boldsymbol{F}_{\text{mid}}))\\ \boldsymbol{F}_{\text{out}}=\text{LN}(\boldsymbol{F}_{\text{attW-M}}+\text{MLP}(\boldsymbol{F}_{\text{attW-M}}))\end{cases} \quad (10)$$

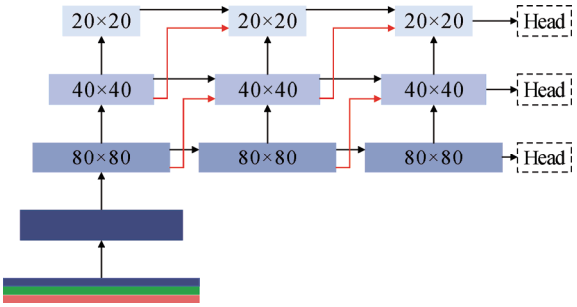


图 8 Res-PANet 结构

Fig.8 Res-PANet structure diagram

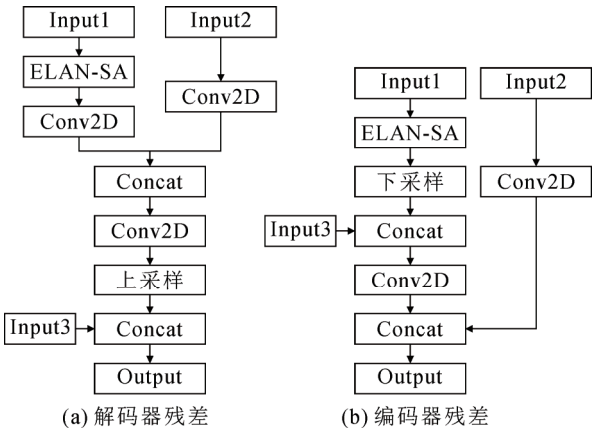


图 9 编解码结构中的残差跳连

Fig.9 Residual hopping in codec structure

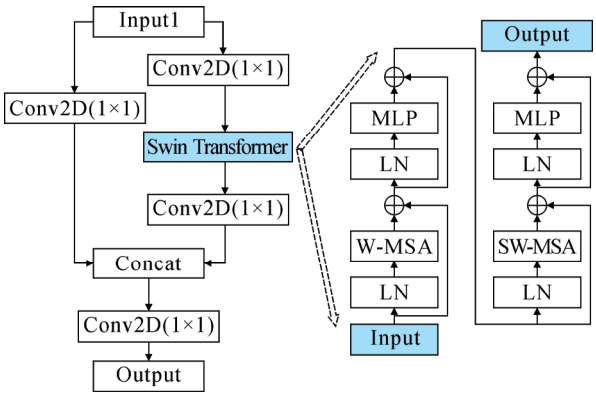


图 10 瓶颈结构的 Swin-Transformer 层

Fig.10 The Swin-Transformer layer of the bottleneck structure

式中：MLP(*)为全连接神经网络，是一种基础的特征增强方式；LN(*)为层归一化操作，用于数据特征的归一化，W-MSA(*)和 SW-MSA(*)分别表示滑窗注意力机制和窗口转换的滑窗注意力机制，主要计算如下：

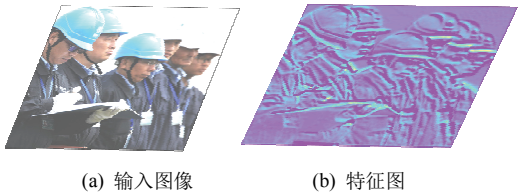
$$\mathrm{W-MSA}(\boldsymbol{q},\boldsymbol{k},\boldsymbol{v})=(\boldsymbol{F}_{\mathrm{in}}\boldsymbol{w}^{q},\boldsymbol{F}_{\mathrm{in}}\boldsymbol{w}^{k},\boldsymbol{F}_{\mathrm{in}}\boldsymbol{w}^{v}) \quad (11)$$

$$\mathrm{SW-MSA}(\boldsymbol{q},\boldsymbol{k},\boldsymbol{v})=(\boldsymbol{F}_{\mathrm{in}}\boldsymbol{w}^{q},\boldsymbol{F}_{\mathrm{in}}\boldsymbol{w}^{k},\boldsymbol{F}_{\mathrm{in}}\boldsymbol{w}^{v}) \quad (12)$$

$$A=\mathrm{Softmax}\left(\frac{\boldsymbol{q}\times\boldsymbol{k}^{\mathrm{T}}}{\sqrt{d_k}}\right)\times\boldsymbol{v} \quad (13)$$

式中：d_k为输入特征 $\boldsymbol{F}_{\mathrm{in}}$的通道维度；$\boldsymbol{q}$、$\boldsymbol{k}$ 和 $\boldsymbol{v}$ 是用特征 $\boldsymbol{F}_{\mathrm{in}}$得到的 3 个度量矩阵。$A$ 为滑窗中注意力机制的计算，通过除以该维度的均方值用于尺度的缩放。

图 11 中，输入的图像经过卷积特征提取后传入 Swin-Transformer，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进行单独的窗口内计算来有效挖掘每个窗口中的潜在目标，如图 11(c)所示，将特征图进行窗口的划分，通过滑窗的多头注意力机制 (window multi-heads self-attention，W-MSA)计算加强模型对特征的高效提取能力。由于采用 W-MSA 模块时，只会在每个窗口内进行自注意力计算(图 11(c)为 4 个窗口为例)，所以存在窗口与窗口之间是无法进行信息传递的，为了解决此问题，特征在被 W-MSA 提取后，会被传递到 SW-MSA 的提取中，用于加强各个独立窗口之间的联系，通过转换移动的窗口(shifted windows multi-heads self-attention，SW-MSA)操作可以使相邻 2 个窗口间进行交互，使上下层间建立跨窗口连接，从而达到全局建模的能力，如图 11(d)所示，根据左右两幅图对比能够发现窗口发生了偏移，窗口从左上角分别向右侧和下方各偏移，偏移后的窗口能弥补第 1 个窗口与剩下窗口的信息交流，同时由原来的 4 个窗口变成 9 个窗口，以此需要对每个窗口进行多头注意力机制 (multi-heads self-attention，MSA)的计算，由于在进行矩阵计算时采取的是并行计算，则会将 9 个窗口都转换为 4×4 的窗口进行计算，对于不足 4×4 的需要进行额外填充，因此带来计算量的增大，于是设计将图 11(d)中的滑窗模块进行滑窗特征重组，特征重组过程如图 12 所示，首先将图 12 左上角的红色区域的特征像素进行下移至图 12 右上角所示位置，接着将绿色区域的特征像素下移至图 12 右下角所示的位置，于是得到新的 4×4 窗口计算方式，这样能最大化的完整保留原始的 4×4 的窗口信息，同时又由于特征发生偏移，存在 2 个相邻块组成的窗口并不是连续的特征信息，于是采取 mask-MSA 的计算，主要是在 MSA 的基础上计算对于非邻近区域的值直接通过 0 值替代即可，即各个像素特征在计算注意力时仅与自身完整块的区域关联度高，而与同窗口下的邻近特征块关联度低，以此降低原始非接触特征区域的干扰。最后在计算完 MSA 后需要进行特征的还原得到经过邻近窗口信息交互后的加权特征，以此获得全局特征信息，挖掘潜在目标信息并有效发现部分被遮挡的区域。



(a) 输入图像　(b) 特征图

(c) W-MSA 滑窗图　(d) SW-MSA 滑窗图

图 11　Swin-Transformer 计算图

Fig.11　Swin-Transformer calculation chart

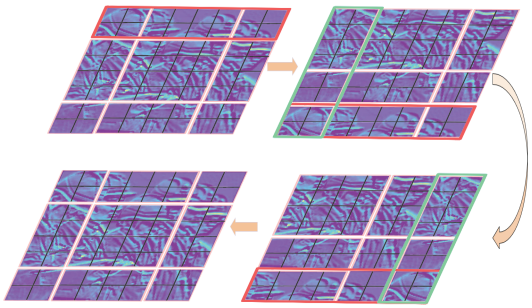


图 12　SW-MSA 滑窗过程

Fig.12　SW-MSA window sliding process

由表 2 可知，在 YOLOv7-Tiny 的检测头输出中设计的 SWCSP，其整体 mAP 相比较原始网络从 86.9%提升至 87.7%，该提升效果主要体现在对每一类目标都有较好的提升，尤其是 fence 类别，主要由于电力人员会在电力围栏的前方或者侧方对目标形成半遮挡的状态，本文改进的算法能有效提升这一检测精度。

5　实验与分析

5.1　数据集

为实现电力现场作业的风险识别，本文构建了

电力安全相关数据集，本数据集主要是从电力现场采集的实拍图片，共有 4023 张样本图片，每张图片包含一种或多种检测目标，且同一种的检测对象也可能有多个。目标标签包含 6 个类别：正确佩戴安全帽、未正确佩戴安全帽、正确穿着工作服、未正确穿着工作服、电力围栏、警告标志、对应在标签中分别为表 3 中的第 1 行 aqm、wdaqm、gzf、wcgzf、fence、signal。同时该数据集中各标签类别的个数如表 3 第 2 行所示，采用 labelimg 软件对样本图像进行标注，制作 VOC 格式的数据集，在每组类别的数量上要高于标准 VOC2017 数据集的数量，因此本数据集在样本容量上与标准数据集一样具有较好的训练能力。

将数据集随机划分为训练集、验证集和测试集，比例为 8:1:1。其中，训练集用于训练目标检测算法的参数，以获取该数据集的训练权重。验证集用于监测训练过程，防止训练过拟合。测试集用于检测训练效果和算法性能。本实验中，训练集共有 3258 张图片，验证集共有 362 张图片，测试集共有 403 张图片。

5.2 实验环境

本实验在 Ubuntu18.08 系统中，python 版本为 3.8.0、CUDA 版本为 11.2，基于 PyTorch1.8 环境下的深度学习框架进行训练和测试。训练使用配置为两块 NVIDIA GeForce RTX 3090-24G 显存的显卡，测试使用配置为本地设备 NVIDIA GeForce RTX 3060-6G Laptop GPU。

5.3 实验过程

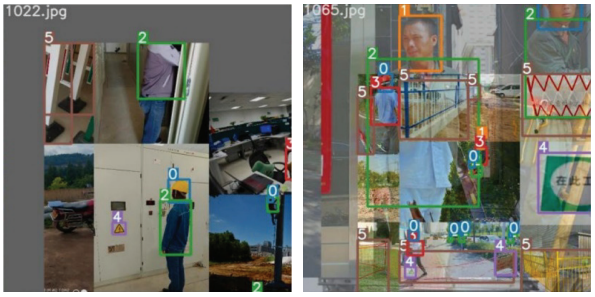
本实验训练过程采用了迁移学习的思想，使用 YOLOv7-Tiny 模型在 COCO 数据集下的预训练权重，将模型的预训练权重输入主干特征提取网络，提高网络训练效率，减少训练收敛时间。实验设置训练轮次为 200 轮，批次大小(batch size)为 16，设定初始学习率为 0.01，采用 Adam 优化器进行训练的优化。

在数据送入训练前，为提高电力安全作业检测的训练效率，避免数据集样本不足导致的过拟合问题，采取马赛克(Mosaic)数据增强是大部分目标检测算法进行数据增强的一种方式。由于原始马赛克数据增强方式是随机裁剪的，会有很大的概率将目标裁剪，使输入的模型样本中包含大量的背景图像，同时对于不足尺寸的区域采取黑色填充的方式，这影响模型的训练效果。于是本文提出一种改进的 Mosaic 数据扩增方式，如图 13 所示，首先将拼接

表 3 数据集标签类别分布表

Table 3 Dataset sample distribution table

目标对象	aqm	wdaqm	gzf	wcgzf	fence	signal
总目标个数	3694	1748	2957	1414	3542	1622



(a) Mosaic4 数据增强 (b) 改进的 Mosaic4 数据增强

图 13 改进 Mosaic 数据增强图像测试对比

Fig.13 Improved Mosaic data enhancement image test comparison

数量由 4 张扩张为 9 张，在按照包围图片的最小矩形面积进行裁剪，减少不必要的空白区域。最后通过缩放操作，达到模型的训练尺寸要求。同时为防止 Mosaic 生成过多标注框的干扰，以及模型训练的参数不利于真实场景的检测，本文采取在训练剩 10%的 epoch 时进行数据增强的关闭使用。

图 14 为各部分的训练损失曲线，改进后的模型(Ours)相比较改进前具有更低的损失，图 14(d)结果表明改进后的模型相比于原始模型在平均精确率方面有明显的提升，验证了改进策略的可行性。

5.4 消融实验结果

为验证改进算法的效果，在相同环境下进行了消融实验。具体如表 4 所示，表 4 中第 1 列为消融实验和原始实验模型对比，第 2—7 列为 6 个检测目标的 AP 值，最后 1 列为 6 类目标的平均 AP 值对比。由表 4 可知，以未做任何改进的 YOLOv7-Tiny 为基准平均精确率为 86.9%，“+”表示模块的混合。基本每一处的融合改进后平均精确率都有了一定提升。同时对于单独安全帽的检测中，会存在部分模型降低的情况，但都基本保持较小的波动内。综合来看 4 处融合改进效果最佳，在每个类别中都有了较高的检测准确率的提升，平均精确率相较于基准提高 3.2%。进一步验证了改进方案的可行性。

同时本文也将 YOLOv7 系列模型与改进的模型在检测效果和推理速度上进行了比较，由表 5 可见，改进的模型具有最高的检测准确率且比

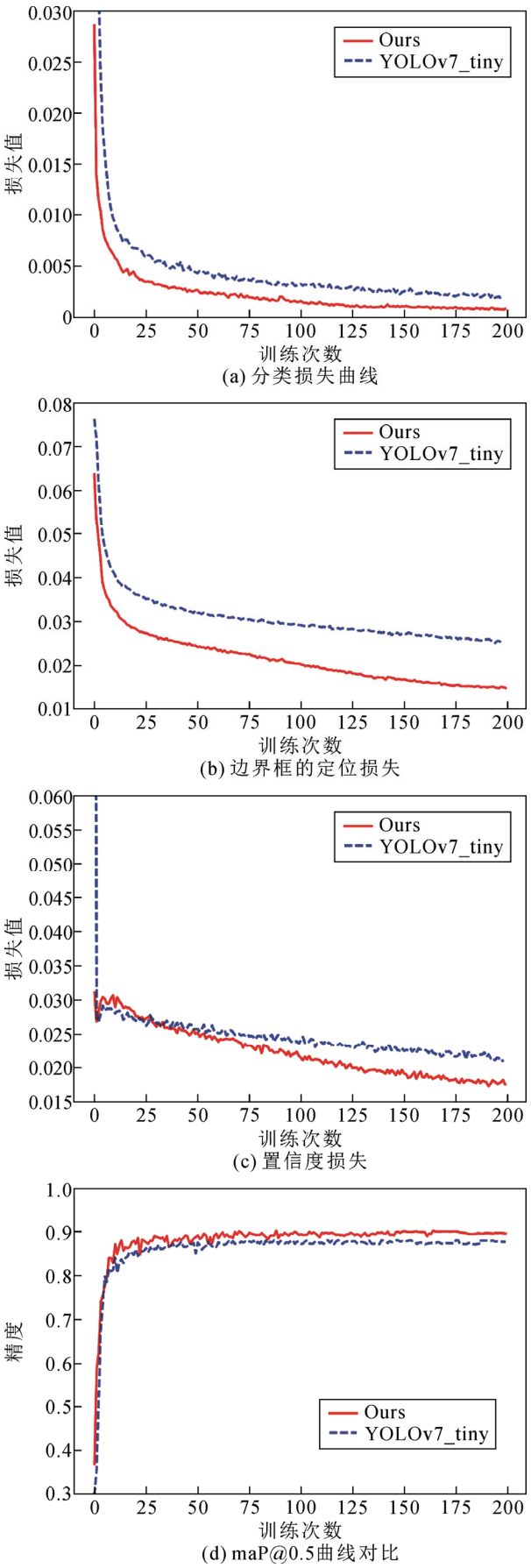


图 14　训练曲线对比图

Fig.14　Training curve comparison chart

表 4　消融实验对比表

Table 4　Comparison table of ablation experiments

模型	AP/%						mAP/%
	aqm	wdaqm	gzf	wcgzf	signal	fence	
YOLOv7-Tiny	94.6	82.8	92.4	94.4	72.4	84.6	86.9
YOLOv7	96.0	85.4	94.1	95.0	73.5	84.2	88.0
ELAN-SA+ResPANet	94.9	86.0	94.2	95.3	76.1	86.6	88.8
ELAN-SA+SWCSP	94.8	83.9	94.2	95.6	74.5	86.1	88.2
ELAN-SA+Mosaic9	94.5	82.9	93.2	95.5	76.5	85.5	88.0
ResPANet +SWCSP	94.4	85.5	93.8	94.5	73.2	86.0	87.9
ResPANet +Mosaic9	94.5	85.3	94.0	96.2	75.1	85.9	88.5
SWCSP+Mosaic9	94.4	83.3	92.8	96.8	77.4	86.5	88.5
ELAN-SA+ ResPANet + SWCSP	94.1	85.8	93.8	97.2	75.7	86.5	89.2
ELAN-SA+ ResPANet + Mosaic9	95.6	86.2	93.0	95.4	75.2	85.7	88.5
ResPANet +SWCSP+Mosaic9	94.2	83.4	92.9	96.2	76.1	85.0	87.9
ELAN-SA+ Res-PANet+SWCSP+Mosaic9	96.5	86.9	94.8	96.9	78.8	86.5	90.1

表 5　模型大小和检测速度的比较

Table 5　Comparison of model size and detection speed

模型	mAP/%	模型大小/MB	检测速度(帧·s^{-1})
YOLOv7-Tiny	86.9	12.0	67.88
YOLOv7	88.0	72.0	26.53
Ours	90.1	15.4	46.25

YOLOv7 模型提高 2.1%，而模型大小仅为其 21%，检测速度高于 YOLOv7 接近 20 帧。

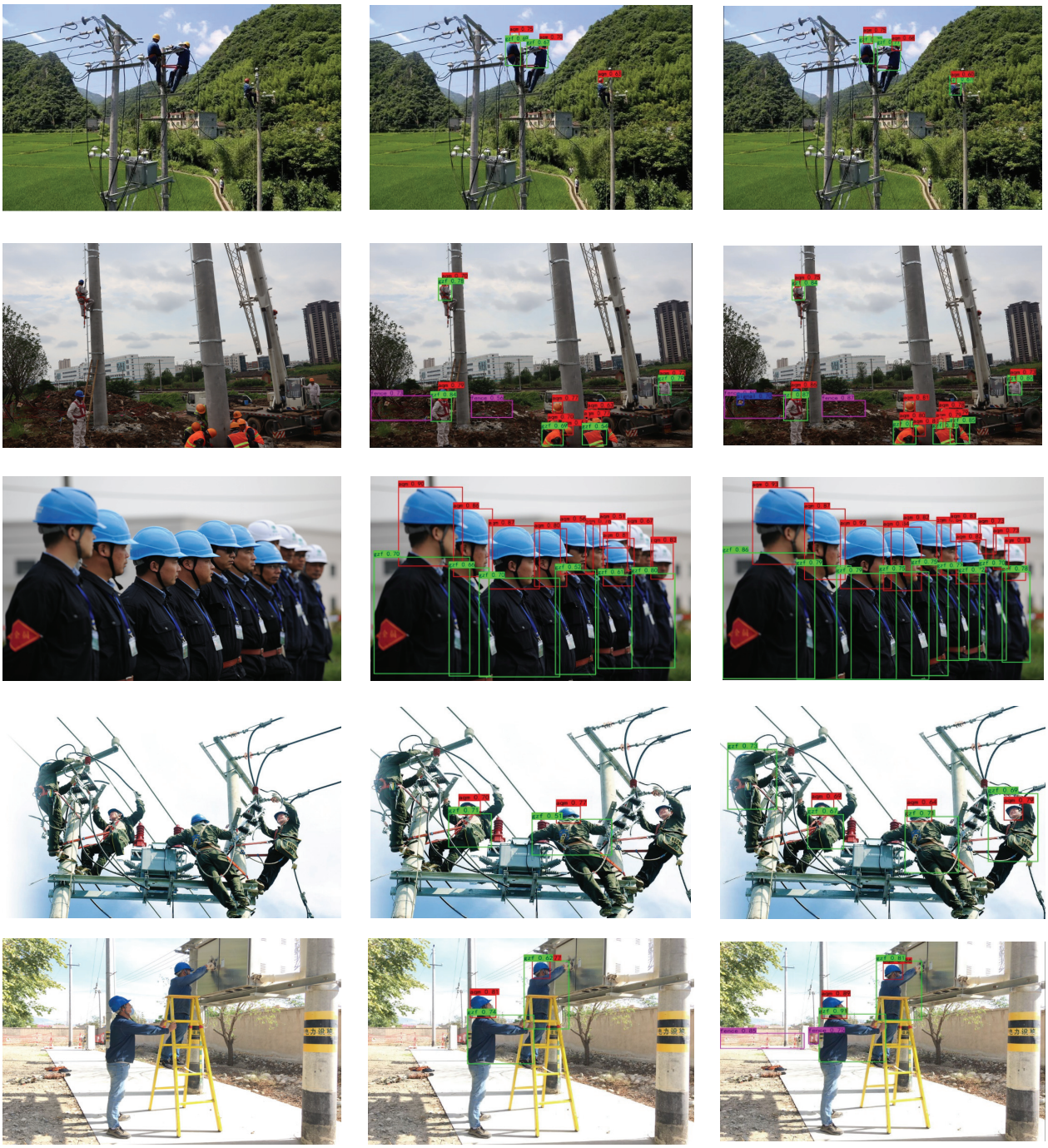
5.5　检测结果对比

为验证图像的实际检测效果，表 6 举例了改进前后算法对测试集的检测结果，其中第 1 行图像是典型的复杂背景下的图像，融合改进模型(Ours)能检测到复杂背景下的漏检安全服；第 2 行图像是典型的多目标图像，融合改进模型(Ours)能正确检测到左下角电力信号标识和人群中的漏检目标；第 3 行图像是典型的重叠目标检测，融合改进模型(Ours)能有效检测到存在部分重叠的最后 3 个人的安全帽和安全服；第 4 行图像是背景较为单一的多目标检测，融合改进模型(Ours)能有效检测到所有目标；第 5 行图像中，融合改进模型(Ours)能有效检测搭配远处的大小电力围栏目标，再一次验证了本模型改进的有效性。

最后本文通过 Score-cam 热力图进行图像的关注区域可视化对比，其中 Score-cam 的原理主要是用目标区域的得分进行特征图的加权，从而摆脱对

表 6 测试集图片检测结果

Table 6 Test set image detection results

原始图像	YOLOv7-Tiny 模型	Ours 模型
		

特征梯度的依赖(由于梯度信息比较复杂，且对于 Sigmoid 以及 ReLU 等激活函数会存在梯度消失的问题)，最终的结果是采取权重和激活图的线性组合获得。其计算原理为

$$\begin{cases} \boldsymbol{H}_l^n=\sigma(\mathrm{UP}(\boldsymbol{A}_l^n)) \\ \boldsymbol{C}_l^n=f(\boldsymbol{X}*\boldsymbol{H}_l^n)-f(\boldsymbol{X}) \end{cases} \tag{14}$$

式中：$\boldsymbol{A}_l^n$ 表示输出特征图的大小；l 表示输出的特征层级；n 表示每个特征层对应的通道数；$\sigma(*)$ 为 sigmoid 激活函数，用来归一化特征图的数值区间；$\boldsymbol{H}_l^n$ 表示原始输出图像尺寸，由于输出特征图有 3 个，本文采取将 3 个特征图进行融合输出。$f(\boldsymbol{X})$为输入特征图；$f(\boldsymbol{X}*\boldsymbol{H}_l^n)$ 为输入的特征图经过加权的结果。$\boldsymbol{C}_l^n$ 即为得到模型对于输入图像的关注区域[26]。

通过热力图可视化可知道，热力图关注的即为目标检测框的中心位置，红色为中心重点关注区域，向外扩散的关注比重降低，表明模型所关注的区域正是所需要检测的区域，同时用红色圈出改进前后的检测对比。表 7 中，第 1 行图像中对于左上角的安全帽小目标的检测，改进的模型能有效关注到；第 2 行图像中对于右上角重叠的安全帽，改进的模型也能有效关注；第 3 行图像中对于多个目标的关注，改进的模型都能有效检测到，进一步验证了模型改进的有效性。

5.6　移动硬件端测试

为更好地适用于边缘端下的变电站安全作业监测，将算法进行基于移动端硬件的测试，硬件移植对象采用 Jetson Xavier NX，如图 15 所示，通过调用硬件摄像头进行实时数据采集，采集图像分辨率大小为 1920×1080[27]。

表 8 中，本文主要测试了移动硬件端对于视频实时检测时所需要的时间，主要由第 2 列推理时间和第 3 列 NMS 候选框的筛选时间构成。在推理时间上，改进后的模型相较于原始 YOLOv7-Tiny 慢约 25 ms，比 YOLOv7 快约 60 ms，其检测速度在通过帧数转化约为 12.5(1000/(75+5))帧/s，具有较好的实用价值。

6　结论

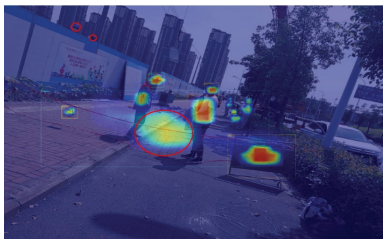
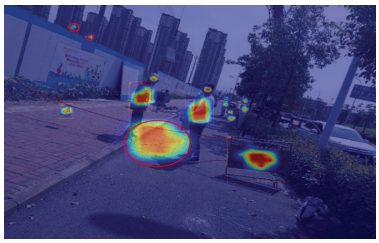
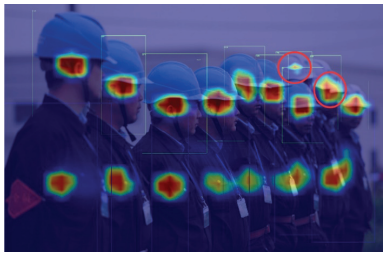
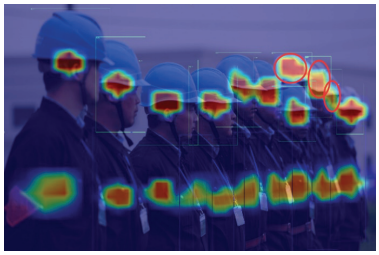
本文提出了一种基于改进 YOLOv7-Tiny 的电力现场作业安全监测算法，通过构建电力安全监督数据集，并分析对比相关因素对目标检测效果的影响，可得到以下结论：

1）背景复杂、多尺度目标和部分遮挡目标是影响模型检测的主要因素。通过通道重组与特征增强的 ELAN-SA 模块来提升模型在复杂场景中对目标的辨识，均值检测精度提升 1.2%；融合高分辨率特征图的 Res-PANet 残差跳连能弥补小特征层信息丢失的问题，提高多目标检测能力，均值检测精度提升 1%；基于 Swin-Transformer 模块检测输出头能捕获全局感受野，提升部分遮挡目标的检测效果，均值检测精度提升 0.8%。

2）消融实验结果表明，融合所有改进的模型在安全帽上的 AP 值提升 1.9%、未戴安全帽的 AP

表 7　基于 Score-cam 的热力图可视化结果

Table 7　Heat map visualization results based on Score-cam

原始图像	YOLOv7-Tiny 模型	Ours 模型
		
		
		

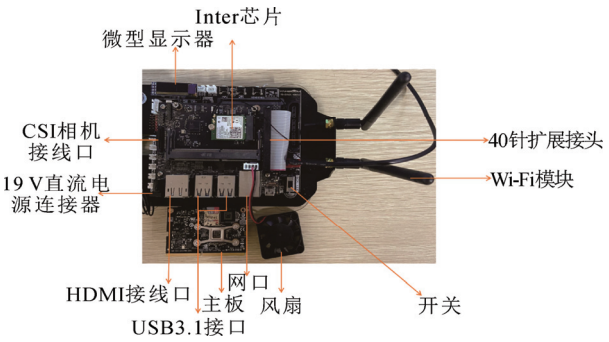


图 15 Jetson Xavier NX 结构图

Fig.15 The structure of Jetson Xavier NX

表 8 模型在硬件端测试对比

Table 8 Comparison of models tested on the hardware side

算法	推理延时/ms	NMS 候选框筛选时间/ms	mAP/%
YOLOv7-Tiny	50±4	4.5±1.5	86.9
YOLOv7	135±5	5.3±2	88.0
Ours	74.5±4	4.8±1.2	90.1

值提升 4.1%、工作服的 AP 值提升 2.4%、未穿工作服的 AP 值提升 2.5%、电力信号标识的 AP 值提升 6.4%、电力围栏标识的 AP 值提升 1.9%。

3）将融合实验模型在电脑端和移动硬件端测试，结果表明，在仅增加 3.4 MB 模型下，融合改进的模型平均检测精度由初始的 86.9%提升至 90.1%。检测速度在电脑端为 46 帧/s，在移动硬件端的推理延时约为 75 ms，其检测结果能为电力系统自动化的智能安全监测起到一定的借鉴作用。

参考文献 References

[1] 马富齐，王 波，董旭柱，等. 电力工业安全影像解译：基本概念与技术框架[J]. 中国电机工程学报，2022，42(2)：458-474.
MA Fuqi, WANG Bo, DONG Xuzhu, et al. Safety image interpretation of power industry: basic concepts and technical framework[J]. Proceedings of the CSEE, 2022, 42(2): 458-474.

[2] 熊卫红，毛兴华，李景禄，等. 小电阻接地方式对人身安全的影响及智能电阻接地方式研究[J]. 电力系统保护与控制，2019，47(14)：166-172.
XIONG Weihong, MAO Xinghua, LI Jinglu, et al. Influence of small resistance grounding mode on personal safety and research of intelligent resistance grounding mode[J]. Power System Protection and Control, 2019, 47(14): 166-172.

[3] 律方成，牛雷雷，王胜辉，等. 基于紫外成像和改进 YOLOv3 的瓷悬式绝缘子放电严重程度评估[J]. 高电压技术，2021，47(2)：377-386.
LÜ Fangcheng, NIU Leilei, WANG Shenghui, et al. Discharge severity assessment of porcelain suspension insulators based on UV images and improved YOLOv3[J]. High Voltage Engineering, 2021, 47(2): 377-386.

[4] 赵振兵，金超熊，戚银城，等. 基于动态监督知识蒸馏的输电线路螺栓缺陷图像分类[J]. 高电压技术，2021，47(2)：406-414.
ZHAO Zhenbing, JIN Chaoxiong, QI Yincheng, et al. Image classification of transmission line bolt defects based on dynamic supervision knowledge distillation[J]. High Voltage Engineering, 2021, 47(2): 406-414.

[5] 罗 鹏，王 波，马恒瑞，等. 基于组合式目标检测框架的低漏报率缺陷识别方法[J]. 高电压技术，2021，47(2)：454-462.
LUO Peng, WANG Bo, MA Hengrui, et al. Defect recognition method with low false negative rate based on combined target detection framework[J]. High Voltage Engineering, 2021, 47(2): 454-462.

[6] 刘开培，李博强，秦 亮，等. 深度学习目标检测算法在架空输电线路绝缘子缺陷检测中的应用研究综述[J/OL]. 高电压技术，2022：1-12[2022-09-23]. https://doi.org/10.13336/j.1003-6520.hve. 20220273.
LIU Kaipei, LI Boqiang, QIN Liang, et al. Overview of application research of deep learning target detection algorithm in insulator defect detection of overhead transmission line[J/OL]. High Voltage Engineering, 2022: 1-12[2022-09-23]. https://doi.org/10.13336/j.1003-6520.hve. 20220273.

[7] 赵 睿，刘 辉，刘沛霖，等. 基于改进 YOLOv5s 的安全帽检测算法[J/OL]. 北京航空航天大学学报，2021：1-16[2022-09-06]. https://doi. org/10.13700/j. bh. 1001-5965.2021.0595.
ZHAO Rui, LIU Hui, LIU Peilin, et al. Helmet detection algorithm based on improved YOLOv5s[J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 2021: 1-16[2022-09-06]. https: //doi.org/10.13700/j. bh. 1001-5965.2021.0595.

[8] 赵 洋，许 军. 基于 MobileNetV2 与树莓派的人脸识别系统[J]. 计算机系统应用，2021，30(8)：67-72.
ZHAO Yang, XU Jun. Face recognition system based on MobileNetV2 and raspberry Pi[J]. Computer Systems Applications, 2021, 30(8): 67-72.

[9] 姚宗强，陈宏利，闫文虎，等. 基于视频图像的电力基建施工违章行为自动识别系统设计[J]. 自动化与仪器仪表，2021(3)：143-145，149.
YAO Zongqiang, CHEN Hongli, YAN Wenhu, et al. Design of automatic identification system based on video images for electrical infrastructure construction violations[J]. Automation & Instrumentation, 2021(3): 143-145, 149.

[10] 常政威，彭 倩，陈 缨. 基于机器学习和图像识别的电力作业现场安全监督方法[J]. 中国电力，2020，53(4)：155-160.
CHANG Zhengwei, PENG Qian, CHEN Ying. Safety supervision method for power operation site based on machine learning and image recognition[J]. Electric Power, 2020, 53(4): 155-160.

[11] 郭敬东，李晓林. 面向电力智能安监的极低分辨率目标检测算法[J]. 计算机工程与设计，2020，41(11)：3188-3192.
GUO Jingdong, LI Xiaolin. Very low-resolution object detection algorithms for electric intelligent safety supervision[J]. Computer Engineering and Design, 2020, 41(11): 3188-3192.

[12] 袁 斌，张皓维，崔萌萌. 基于深度学习的电力基建现场安全管控系统[J]. 电网与清洁能源，2020，36(9)：30-36.
YUAN Bin, ZHANG Haowei, CUI Mengmeng. Deep learning-based safety control system for power infrastructure sites[J]. Power System and Clean Energy, 2020, 36(9): 30-36.

[13] ZHAO B N, LAN H J, NIU Z W, et al. Detection and location of safety protective wear in power substation operation using wear-enhanced YOLOv3 algorithm[J]. IEEE Access, 2021, 9: 125540-125549.

[14] LI K X, QIN L, LI Q, et al. Improved edge lightweight YOLOv4 and its application in on-site power system work[J]. Global Energy Interconnection, 2022, 5(2): 168-180.

[15] PENG G Y, LEI Y, LI H L, et al. CORY-Net: contrastive res-YOLOv5 network for intelligent safety monitoring on power grid construction sites[J]. IEEE Access, 2021, 9: 160461-160470.

[16] 曹　捷，郭志彬，潘立志，等. 高空作业场景下的安全带穿戴检测[J]. 湖南科技大学学报(自然科学版)，2022，37(1)：92-99.
CAO Jie, GUO Zhibin, PAN Lizhi, et al. Detection of safety belt wearing in aerial work[J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2022, 37(1): 92-99.

[17] 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[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S. l.]: IEEE, 2023: 7464-7475.

[18] ZHANG Q L, YANG Y B. SA-Net: shuffle attention for deep convolutional neural networks[C]//ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, Canada: IEEE, 2021: 2235-2239.

[19] WU Y X, HE K M. Group normalization[C]// Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018: 3-19.

[20] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 6848-6856.

[21] 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 (ICPR). Milan, Italy: IEEE, 2021: 180-187.

[22] 郝　帅，杨　磊，马　旭，等. 基于注意力机制与跨尺度特征融合的 YOLOv5 输电线路故障检测[J]. 中国电机工程学报，2023，43(6)：2319-2330.
HAO Shuai, YANG Lei, MA Xu, et al. YOLOv5 transmission line fault detection based on attention mechanism and cross-scale feature fusion[J]. Proceedings of the CSEE, 2023, 43(6): 2319-2330.

[23] 颜宏文，万俊杰，潘志敏，等. 基于改进 Yolov5-Lite 轻量级的配电组件缺陷识别[J/OL]. 高电压技术，2022：1-10[2022-10-12]. https://doi. org/10.13336/j. 1003-6520. hve. 20220387.
YAN Hongwen, WAN Junjie, PAN Zhimin, et al. Defect identification of distribution components based on improved Yolov5 Lite light weight[J/OL]. High Voltage Engineering, 2022: 1-10[2022-10-12]. https: //doi. org/ 10.13336/j. 1003-6520. hve. 20220387.

[24] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 770-778.

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

[26] WANG H F, WANG Z F, DU M N, et al. Score-CAM: score-weighted visual explanations for convolutional neural networks[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, USA: IEEE, 2020: 111-119.

[27] HAN G J, HE M, ZHAO F, et al. Insulator detection and damage identification based on improved lightweight YOLOv4 network[J]. Energy Reports, 2021, 7(Supplement 7): 187-197.



何　敏
1995—，男，博士生
研究方向为人工智能在电力系统中的应用，计算机视觉技术
E-mail: whuhemin@whu.edu.cn

HE Min
Ph.D. candidate



秦　亮(通信作者)
1980—，男，博士，副教授，博导
主要从事电力电子化电力系统的分析与控住、柔性直流系统，人工智能技术在电力系统中的应用、可再生能源发电方面的工作
E-mail: qinliang@whu.edu.cn

QIN Liang
Ph.D.
Associate professor
Corresponding author

收稿日期　2022-12-13　修回日期　2023-04-06　　编辑　卫李静