DOI: 10.16452/j.cnki.sdkjzk.2025.03.011 文章编号:1672-3767(2025)03-0107-12

基于改进 YOLOv3 的电力巡检照片分类命名方法

郑 高,郑恩辉,王桂荣
(中国计量大学 机电工程学院,浙江 杭州 310018)

摘 要:为实现电力巡检过程的智能化,解决拍摄的巡检照片质量不高、分类准确率低的问题,本研究提出一种基于改进 YOLOv3 的电力巡检照片分类命名方法。该方法利用融合通道洗牌和通道注意力提高模型的特征表征能力,结合多层感知器神经网络与后处理模块完成最终的分类命名。实验结果表明,改进后的 YOLOv3 模型平均准确率优于原 YOLOv3 模型,平均精度均值由 85.62%提高至 94.73%;与现有的主流分类模型相比,能更好地处理拍摄质量差的巡检照片,提升电力巡检效率。

关键词:电力巡检;深度学习;YOLOv3;目标检测;注意力机制
中图分类号:TP391.4 **文献标志码**:A

Classification and naming method for power inspection photos based on improved YOLOv3

ZHENG Gao, ZHENG Enhui, WANG Guirong
(College of Mechanical and Electrical Engineering, China Jiliang University, Hangzhou 310018, China)

Abstract: To achieve intelligent power inspection process and solve the problem of poor quality and low classification accuracy of inspection photos taken, a classification and naming method for inspection photos based on improved YOLOv3 was proposed. The modules of fused channel shuffling and channel attention were used to improve the model's feature characterization ability. Then, the final classification and naming task was completed by combining multiplayer perceptron neural network and post-processing module with the model. The experimental results show that the improved YOLOv3 model has better average accuracy than the original YOLOv3 algorithm, with the mean average precision YOLOv3 increasing from 85.62% to 94.73%. Compared to mainstream image classification models, the improved model can better handle poorly captured inspection photos, thus improving the efficiency of the power inspection process.

Key words: electric power inspection; deep learning; YOLOv3; object detection; attention mechanism

计算机视觉技术不断发展,在各行各业得到了广泛应用,其中目标检测是计算机视觉领域中的重要组成部分。Girshick 等[1]提出的 R-CNN(regions with convolutional neural network features)是最早成功使用 CNN 进行目标检测的方法之一,也是目标检测领域的一个重要里程碑。Girshick 等[2]结合 SPPnet[3]提出了 Fast R-CNN,其 Rol Pooling 层沿用了 SPPnet 中的 SPP(spatial pyramid pooling)层,有效提升了训练速度与检测精度。Ren 等[4]在 Fast R-CNN 的基础上使用 RPN(region proposal network)优化了候选框的提取过程。Redmon 等[5]提出 YOLO(you only look once)算法,与当时的主流算法不同,YOLO 算法将输入图像分为大小相同的正方形网格,目标物体中心点落在哪个网格,该网格就负责检测该目标并返回检测框,

收稿日期:2023-07-05
基金项目:国家自然科学基金项目(61203113);浙江省自然科学基金项目(LGG22F030001)
作者简介:郑 高(1997—),男,山东青岛人,硕士研究生,主要从事无人机与计算机视觉方面的研究.
E-mail:605700534@qq.com
郑恩辉(1975—),男,浙江杭州人,副教授,博士,主要从事无人机与计算机视觉方面的研究,本文通信作者.
E-mail:ehzheng@cjlu.edu.cn

其检测精度比 Faster R-CNN 低,但检测速度具有明显优势。Redmon 等[6]将 YOLO 算法一直更新到 v3 版本后不再更新,之后由其他开发者接手。

电力巡检照片的自动分类命名是实现电力巡检智能化的前提,通过无人机巡检获得的照片是未命名的,若不经分类命名,巡检人员无法确定含有缺陷的照片属于哪条线路、哪座电线杆塔、杆塔的哪一侧或哪一相,对巡检工作造成了极大的不便。为实现电力巡检智能化,国内外相关研究人员基于深度学习模型开展了大量研究。Zhou 等[7]提出类感知边缘辅助轻量级语义分割网络,引入类感知边缘检测,并使用区域注意力与区域引导边缘注意力模块提取更深层的区域特征,将巡检照片中的杆塔部位分割之后再做后续处理,但该方法更适用于城市场景,难以应对复杂的野外场景。王振[8]对点云数据进行抽稀处理,构建点云数据模型库对输电线路和电力杆塔进行三维建模,通过巡检照片拍摄点构成的三维杆塔模型进行巡检照片分类,但该方法工作量大,且对照片拍摄质量要求较高,拍摄位置产生偏移会导致分类准确率较低,并且无人机的巡检航线存在安全隐患。Vieira-E-Silva 等[9]发布了包含多个高压电力线组件高分辨率图像的数据集 STN PLAD,提出 MS-PAD(multi-size power line asset detection)的目标检测识别方法,该方法融合了 SSD(single shot multibox detector)与 Faster R-CNN 的思想,使用两种检测器分别检测大目标与小目标,但两阶段检测策略和更多的参数量导致训练模型体积较大,网络训练和检测速度较慢,更适用于待检测目标较多的场景。Bayasgalan 等[10]提出多种方案用于实时检测绝缘子缺陷,并通过对比所提方案的实验结果得出最佳的缺陷检测方案,但是并没有详细给出算法使用说明,无法验证其可靠性。Ohta 等[11]建立了一种无人机巡检系统,该系统可对电线杆塔的绝缘子进行图像采集、定位与识别,使用微型飞行器链路(micro air vehicle link,MAVLink)协议通信,控制软件通过机器人操作系统(robot operating system,ROS)实现,并使用 YOLOv3 作为实时目标检测网络,但相较于 YOLO 系列的最新版本,YOLOv3 的检测速度慢,不适用于实时检测,并且该方法没有应用缺陷识别,需要无人机操作者进行现场巡检,增加了巡检人员的工作量,未能实现巡检智能化。

综上,针对现有算法处理存在电力巡检照片的分类准确率低、场景适配性差、检测响应慢及鲁棒性差等问题。本研究提出一种基于改进 YOLOv3 巡检照片分类命名方法,通过融合通道洗牌[12]和通道注意力[13]的多通道注意力残差模块(multi-channel attention residual block,MCarB)提升特征表达能力,强化显著特征并抑制冗余特征,提高检测精度。

1 YOLOv3 网络结构与原理

1.1 YOLOv3 网络结构

YOLOv3 是 YOLO 目标检测系列算法的第 3 个版本,主要分为 3 部分:主干部分是特征提取模块,连接部分是多尺度检测模块,输出部分是边界框的检测模块。YOLOv3 网络结构如图 1 所示,Conv2D 代表卷积模块,Residual Block 代表残差模块,Concat 代表拼接模块,UpSampling 代表上采样模块。

相较于前两个版本,YOLOv3 在 3 个方面改进了网络结构:①采用 Darknet-53 的网络结构,包含 53 个卷积层和池化层,可以有效提高特征提取的准确性[14];②将 ResNet(residual network)[15]的残差块引入到网络中并进行了改进,改进后的残差块能够更好地保留图像的低级特征和高级特征,提高检测的准确性和鲁棒性;③引入特征金字塔网络(feature pyramid networks,FPN)机制,使用多个不同大小的特征图来进行多尺度检测工作,可以检测不同大小和形状的目标。

1.2 YOLOv3 原理

YOLOv3 是基于锚框的算法,即预先在图片上生成先验框,然后通过 YOLOv3 判断框内是否存在目标,若存在则根据结果对先验框的中心及其长、宽进行调整,最终确定检测框大小和位置。YOLOv3 网络的输入图片尺寸(像素)为 416×416,输出 13×13、26×26、52×52 等 3 种不同尺度的特征图,分别对应图 1 中 YOLOv3 的输出,将图像分成多个网格,每个网格上设有预先设定宽高的 3 个先验框,其中 13×13 的特征图用于检测大目标,26×26的特征图用于检测中等尺度的目标,52×52的特征图用于检测小目标,共返回

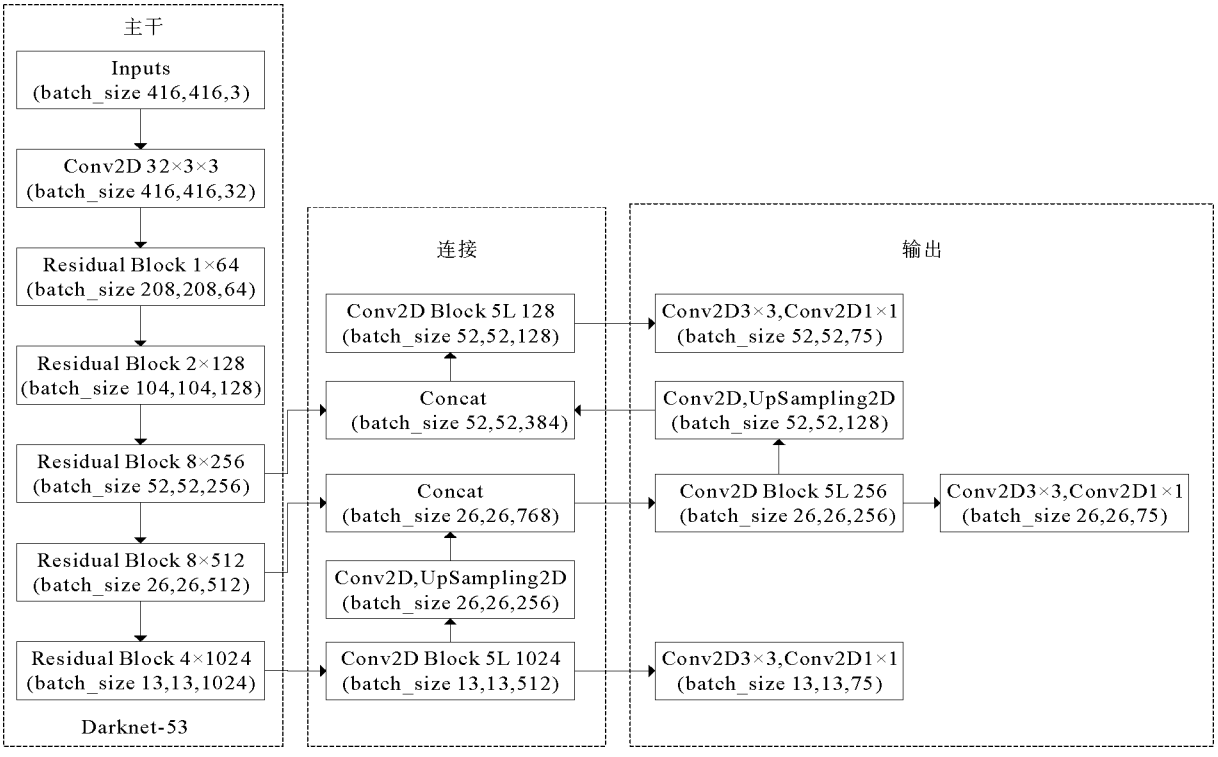


图 1　YOLOv3 网络结构

Fig. 1　YOLOv3 network structure

13×13×3+26×26×3+52×52×3 个框。物体中心点落在哪个边界框内,该边界框就负责识别这个物体。边界框的检测过程如图 2 所示,其中虚线框为先验框,红色实线框为网络输出的检测框。确定检测框的各参数计算式为:

$$b_x=\sigma(t_x)+c_x,b_y=\sigma(t_y)+c_y; \tag{1}$$

$$b_w=P_w e^{t_w},b_h=P_h e^{t_h}。 \tag{2}$$

式中:(b_x,b_y)为检测框的中心坐标,b_w 和 b_h 分别为检测框的宽度和高度,(b_x,b_y,b_w,b_h)为最终检测目标的边界框参数,t_x 和 t_y 为检测偏移量,t_w 和 t_h 为检测尺度,(t_x,t_y,t_w,t_h)为网络检测的边界框参数,(c_x,c_y)为特征图的左上角坐标,P_w 和 P_h 为先验框映射在特征层上的宽度和高度。$\sigma(x)$为 sigmoid 函数,目的是将检测偏移量缩放到[0,1],保证测得的边界框的中心坐标落在单元格内。

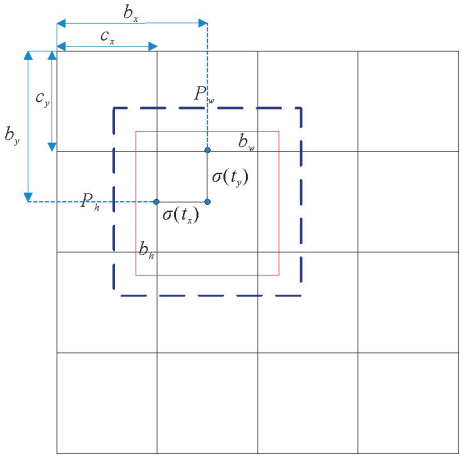


图 2　边界框检测过程

Fig. 2　Box detection process

2　基于改进 YOLOv3 的巡检照片分类方法

2.1　改进模型整体结构

电力巡检的实际场景中,拍摄巡检照片时,杆塔复杂的结构也会被同时拍摄到,影响目标检测模型对巡检照片中杆塔部位的检测效果。因此,本研究对原 YOLOv3 网络进行改进,达到过滤无关背景信息、精确分类电力巡检照片的目的。

改进的 YOLOv3 网络主要由特征提取主干网络(主干部分)、特征融合网络(连接部分)和分类与回归器(检测部分)组成,其整体结构如图 3 所示,其中 4-branch Conv 表示 4 分支的分组卷积模块,检测头表示网

络的分类器与回归器。

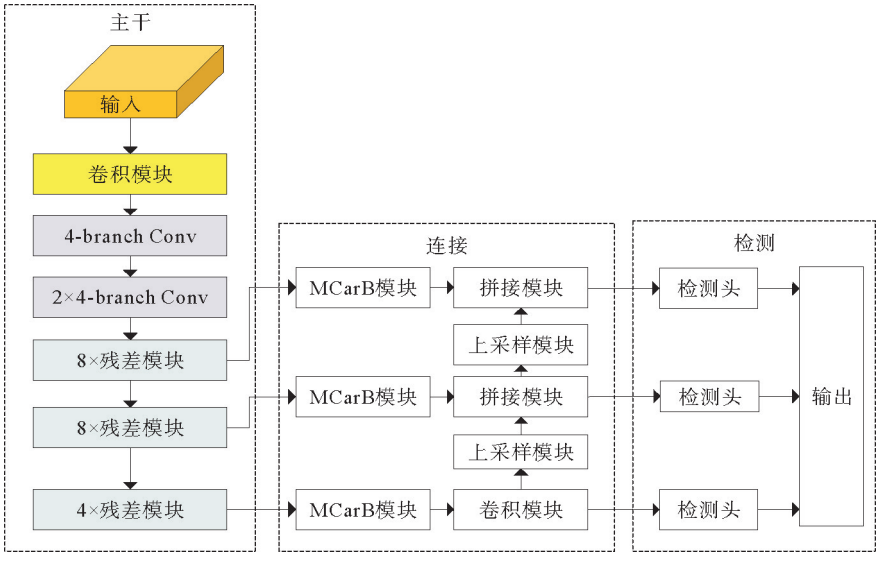


图 3 改进的 YOLOv3 模型结构

Fig. 3 Structure of improved YOLOv3 model

主干网络结构如图 4 所示,输入图像尺寸为 416×416,首先经过卷积模块将通道数扩充至 32,然后经过一个分组卷积模块输出尺寸为 208×208×64 的特征图,再经过两个分组卷积模块输出尺寸为 104×104×128 的特征图。接下来依次使用 8 个、8 个、4 个串联残差模块,输出尺寸为 52×52×256、26×26×512 和 13×13×1 024 的特征图,作为连接部分的输入。

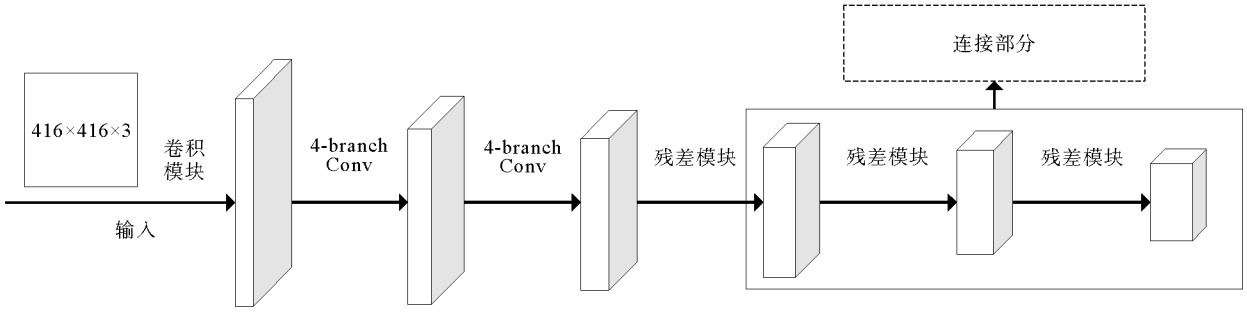


图 4 主干部分结构

Fig. 4 Structure ofbackbone

连接与检测部分结构如图 5 所示。该部分以主干部分输出的 3 张不同尺寸的特征图作为输入,分别通过 McarB 模块获得融合注意力权重后的特征图,尺寸不发生改变。最后输出的 3 个特征图尺寸分别为 52×52×21、26×26×21 和 13×13×21,以检测不同大小的目标。输出维度中包括检测框对应的 5+N 个值,其中 5 个值分别为检测框的中心坐标(x, y)、检测框的宽和高(w, h)以及检测框置信度,N 为检测类别个数,输出值为检测框各个类别的概率。

相较于原 YOLOv3 模型,本研究提出的模型主要进行两方面的改进:

1) 巡检照片通常存在较多与检测目标无关的背景信息,这些信息会干扰网络提取目标的特征,导致原图像的特征提取不准确。因此,主干网络与 FPN 的连接处添加了 MCarB 模块。

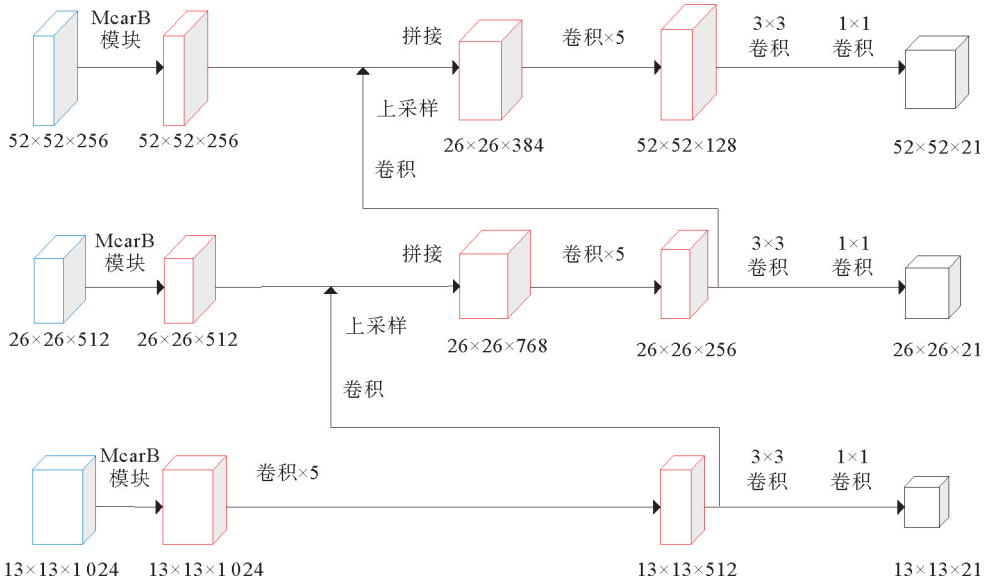


图 5　连接与检测部分结构

Fig. 5　Structure of connection and detection

2）添加的多通道注意力残差模块 McarB 使整个模型更加复杂，导致运算量增加，对模型训练有一定的影响。因此使用 ResNeXt[16] 网络中的分组卷积替换原网络图 1 中的前两个残差块中的卷积，从而提高模型的训练速度，避免模型训练过拟合。

2.2　多通道注意力残差模块

MCarB 模块结构如图 6 所示。图 6 中，FC_1 和 FC_2 是全连接层，融合是将得到的权重加权赋予每个通道特征。MCarB 模块整体采用残差结构，融合了通道洗牌与通道注意力的思想。通道洗牌是在组卷积的基础上改进，组卷积虽然减少了计算量，但是分组会导致原特征图不同组之间的特征信息不交互。通过通道洗牌可以重组输入特征图，使每一组卷积的输入来自不同的组，以保证信息流的连续性，有利于模型的特征提取。通过学习，通道注意力自动获取每个特征通道的权重占比，然后根据权重强化原图像中的显著特征并抑制冗余特征，从而过滤与检测目标无关的背景信息。该过程首先进行特征压缩，将一个通道中整个空间特征编码为一个全局描述特征，然后通过全连接层生成每个特征通道的权重，最后将所得权重逐个加权到之前的特征。

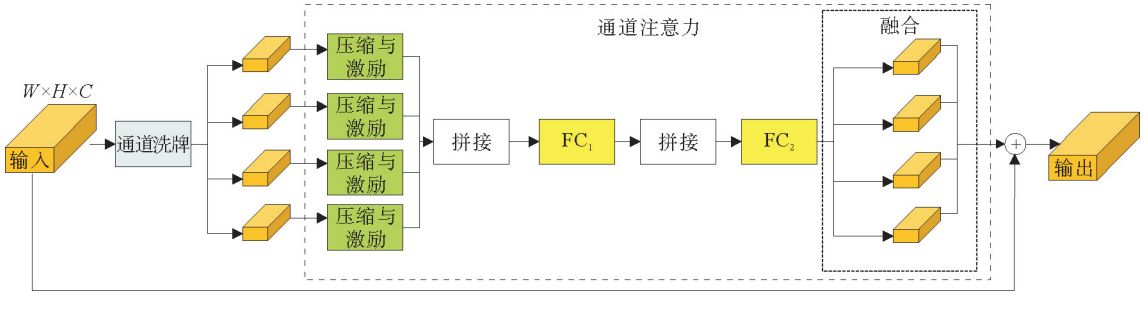


图 6　MCarB 模块结构

Fig. 6　Structure of MCarB module

MCarB 模块的输入为 $W\times H\times C$ 的特征图，W 和 H 分别代表特征图的宽和高，C 代表通道数。通过通道洗牌模块打乱特征图通道顺序后再将通道拼接成 4 个分支，通道洗牌的具体流程如下。

1）将输入特征图分为 g 组，在 MCarB 模块的实际应用中分为 4 组，总通道数 $C=g\times n$。将原输入的

通道维度拆分为(G,N)两个维度,将整个特征图展开为(G, N, W, H)的四维矩阵。

2) 将矩阵沿着 G 轴和 N 轴进行转置,形成(N, G, W, H)的四维矩阵。

3) 将转置后的矩阵重塑成(G, N,W, H)的四维矩阵。

4) 将 G 轴平铺得到 4 个(N, W, H)的特征图,形成 4 个分支再进行卷积提取特征。

然后进行压缩与激励(squeeze and excitation,SE)的通道注意力机制操作。在进行通道洗牌后的每个分支中,先使用 1 个 1×1 卷积层将输入特征降维,再在各分支的每个通道中进行注意力压缩操作,将输入特征图压缩成 1 个 1×1×C 的特征向量 $\boldsymbol{Z}$,如式(3)所示。

$$z_c=\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}x_c(i,j)。\tag{3}$$

式中:z_c 为特征向量 $\boldsymbol{Z}$ 的第 c 个元素;x_c 为第 c 个输入特征图;c 的取值范围为[1,C],C 为输入特征的通道总数。

经过激励操作,得到每条通道的权重。拼接模块将 4 个分支输出的特征进行拼接,经过第 1 个全连接层和 ReLU 激活函数将特征降维,再经过 1 个全连接层将特征恢复至原先的维度,使用 sigmoid 激活函数将每个特征通道的权重值固定在[0,1]中,激励过程如式(4):

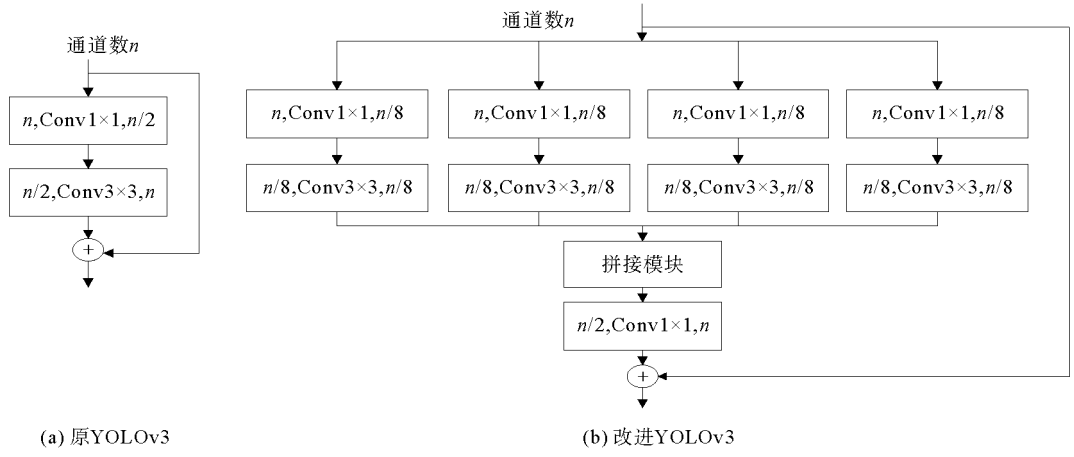
$$s=\delta(W_2\sigma(W_1z))。\tag{4}$$

式中:δ 为 ReLU 激活函数,$\delta(x)=\max(0,x)$;σ 为 sigmoid 激活函数,$\sigma(x)=\frac{1}{1+\exp(-x)}$;$W_1$ 为第 1 次全连接层操作;W_2 为第 2 次全连接层操作;s 是通道权重值,其中包含了 C 个通道权重值。

将激励操作中所得权重赋予每个特征通道并进行拼接,然后使用 1 个 1×1 卷积层将拼接后的特征通道维度调整为输入时的维度,用于细化输出特征,捕获图像中的深层特征。模块的输出为赋予每个通道权重后的特征图。

2.3 主干网络中的分组卷积

原 YOLOv3 网络中的残差结构如图 7(a)所示,原残差块由 1×1 和 3×3 两个卷积层组成。MCarB 模块虽然能提高模型的检测性能,但是在原网络结构基础上添加模块会使网络结构更加复杂。为避免改进时增加参数量,使改进后的模型获得更好的检测效果,将原残差块中的卷积用分组卷积代替,如图 7(b)所示,将原来的 1 条支路分为 4 个分支,每个分支卷积后的通道数变为原结构的 1/4,然后通过拼接模块进行特征拼接,再使用 1×1 卷积层恢复维度,将输入特征与经过卷积后的特征相加并输出分组卷积后的特征。图 7 中“+”表示残差块的输出叠加。



(a) 原YOLOv3 (b) 改进YOLOv3

图 7 主干残差块示意图

Fig. 7 Schematic diagram of the main residual block

参数量(P)和计算量(每秒浮点运算次数,F)是衡量模型复杂度的间接衡量标准,

$$P=\sum_{l=1}^{D}C_{l-1}\cdot K_l^2\cdot C_l, \tag{5}$$

$$F=\sum_{l=1}^{D}M_l^2\cdot C_{l-1}\cdot K_l^2\cdot C_l。 \tag{6}$$

式中:D 为卷积层数;M 为卷积核输出特征图的边长;K 为卷积核边长;C_l 表示第 l 个卷积层的输出通道数,也是该层的卷积核数。已知模块中所用两种卷积核边长为 1 和 3,输入输出通道数如图 7 所示,并且输出特征图边长相等。计算得出改进后结构参数量是改进前参数量的 31%,能有效提高模型的训练速度,同时降低模型训练过拟合的风险。

3　实验结果分析

3.1　训练数据集与实验环境

训练数据集选取 60 座双回直线杆塔,共有 1 320 张巡检照片及其对应的云台数据,图像分辨率为 5 472×3 078。为提高模型训练速度,将图像大小统一调整为 416×416。使用 LabelImg 工具制作相应的 XML 标签,并将数据集随机划分为 80%训练集和 20%验证集。改进 YOLOv3 算法所用数据集为其中的导线端、横担端巡检照片,这两类照片中每一类有 360 张,图像增强扩充数据集至 3 600 张,图 8 展示了某张图像通过不同数据增强方式后的对比。

(a) 原图像　(b) 高斯噪声和垂直翻转增强

(c) 对比度和水平翻转增强　(d) 色温和水平翻转增强

图 8　数据增强后的数据集图像

Fig. 8　Dataset images after data enhancement

实验使用 NVIDIAGeForce RTX 3060 显卡,在 Windows 11 操作系统、Python3.6 版本和 PyTorch 1.8.1 深度学习框架下进行训练。

3.2　分类命名流程

分类命名流程如图 9 所示。输入待分类巡检照片以及照片对应的云台数据,先利用待分类云台数据中的海拔高度信息初步分类出地线、塔基,然后将每张巡检照片拍摄位置的经度、纬度和海拔高度数据以及无人机云台的偏航角数据,输入经典多层感知器(multilayer perceptron,MLP)神经网络,输出初始分类结果:杆塔的左侧上相、左侧中相、左侧下相、右侧上相、右侧中相、右侧下相,每个类别下有 3 张巡检照片。经过改

进 YOLOv3 的目标检测模型对每个类别下的 3 张巡检照片中的导线端与横担端进行检测,得到带有检测框的巡检照片。后处理模块的输入为通过改进 YOLOv3 模型获得的带有检测框的巡检照片,输出分类包括杆塔的导线端、横担端和绝缘子,并结合初始分类结果得到最终分类结果。最后,将巡检照片命名为最终分类结果中的名称。

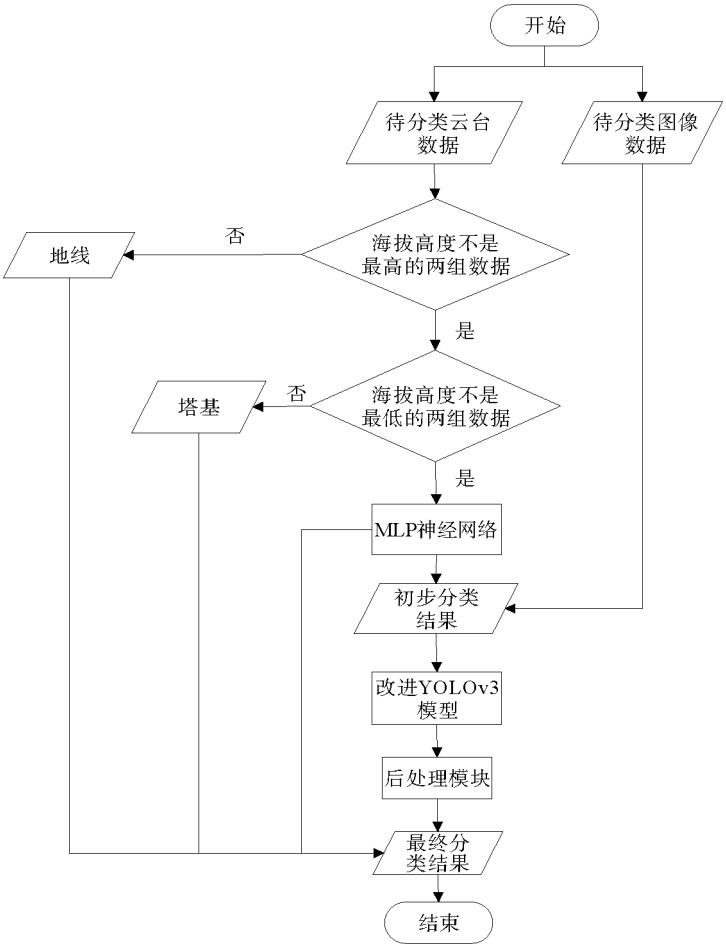


图 9 分类命名流程

Fig. 9 Classification and naming process

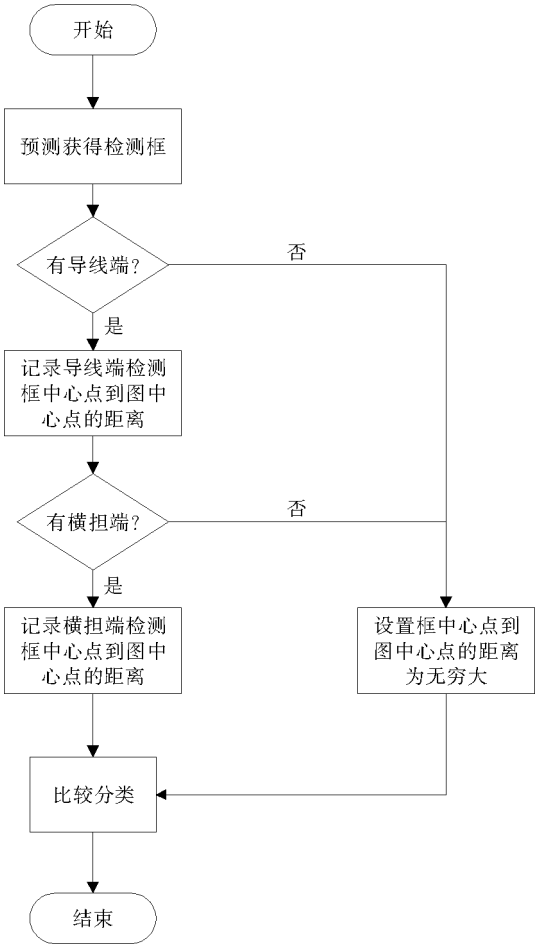


图 10 后处理流程

Fig. 10 Process of post-processing module

后处理模块的流程如图 10 所示,使用模型依次检测每一相中的 3 张未分类巡检照片,检测后保存带有检测框的巡检照片,通过后处理流程得到每张巡检照片中的各类别检测框中心点与照片中心点的欧式距离。最小欧式距离对应的巡检照片被分类为检测框对应类别,未得到分类的巡检照片被分类至绝缘子串类别。

完成整个分类流程后,结合初步分类结果和后处理模块得到的分类结果,得到每张巡检照片的最终分类结果,最后使用最终分类结果为巡检照片命名。图 11 为单座电线杆塔的 6 个相共 18 张巡检照片的最终命名结果。

3.3 超参数设置与评价指标

实验训练参数如表 1 所示,为加快训练速度,提高模型的可解释性,训练过程采用冻结训练与解冻训练策略,使用改进 YOLOv3 在 VOC2007 数据集上预训练的权重。在前 50 次迭代中冻结主干网络,并微调其他网络层,在冻结训练阶段,输入的批量大小设置为 8,学习率设置为 0.001。后 50 次迭代中解冻主干网络,在解冻训练阶段,输入的批量大小设置为 4,学习率设置为 0.000 1。

图 11　最终命名结果

Fig. 11　Final naming results

训练得到目标检测模型后,根据给出的 3 个模型评价指标:平均精确率均值(m_{AP})、网络每秒传输的帧数(frame per second,FPS)以及单张照片的检测时间对该模型的性能进行评价。其中,m_{AP} 用于评估模型的检测精度,m_{AP} 值越大,说明模型性能越好,计算式为:

$$m_{AP}=\frac{\sum_{i}^{N}A_i}{N}。\tag{7}$$

式中:N 为数据集中目标类别个数,i 为类别编号,A_i 为第 i 个类别的平均准确率。A_i 通过精确率 p 与召回率 r 获得,是 p-r 曲线下的面积,$p(r)$表示纵轴为 p、横轴为 r 的函数,如式(8)~(10)所示。

表 1　实验训练参数

Table 1　Experimental training parameters

训练参数	参数值
冻结训练迭代次数	50
解冻训练迭代次数	100
冻结训练输入批量大小	8
解冻训练输入批量大小	4
冻结训练学习率	10^{-3}
解冻训练学习率	10^{-4}
Cuda 调用	True

$$p=\frac{T_P}{T_P+F_P},\tag{8}$$

$$r=\frac{T_P}{T_P+F_N},\tag{9}$$

$$A_i=\int_0^1 p(r)\mathrm{d}r。\tag{10}$$

式中:T_P 表示为正样本被正确检测为正样本的数量,F_P 表示为负样本被错误检测为正样本的数量,F_N 表示为正样本被错误检测为负样本的数量。FPS 数值越大,神经网络的处理速度越快,FPS 和单张照片的检测时间常用于评估模型的检测效率。

3.4　实验对比与分析

1) 改进前后对比

改进 YOLOv3 在训练过程中的损失值变化曲线如图 12 所示,其中横坐标代表训练的迭代次数,纵坐标表示损失值大小。训练开始时初始损失值为 307.498,经过 5 轮迭代损失值迅速下降到 7.758,经过前 50 轮迭代损失值下降到 2.920。在第 51 次训练迭代时损失值升至 3.521,之后损失值不断下降逐渐趋向稳定,具有较好的收敛,最后损失值收敛至 2.150 左右。

选择损失值最小时对应的权重文件作为检测时的权重,对测试集进行测试,得到各个类别检测精度,并使用同样的训练策略在YOLOv3模型上进行测试。改进前后YOLOv3模型各类别检测精度如表2所示,

可以看出,改进后 YOLOv3 模型对横担端的检测精度提高 9.22%,对导线端的检测精度提高 9.01%,m_{AP} 值提高 9.11%。因此改进后 YOLOv3 模型具有更好的检测精度,能够满足实际应用需要。

2) 模块消融实验对比

模块消融实验结果如表 3 所示。相比原网络,两种模块对模型检测的平均精度分别提升 0.26%和 8.71%。同时,分组卷积提高了模型的检测效率,MCarB 模块使模型整体的检测速度略微减慢。从模型整体来看,改进后 YOLOv3 模型精度比原 YOLOv3 模型提升较大,并且检测效率基本相同。

3) 输入图像尺寸性能影响

低分辨率图像会对模型的性能产生一定的影响,但在计算资源不足或者实际应用中有运行效率的指标时,使用低分辨率图像会更符合实际需求。输入不同尺寸图像对模型性能的影响结果如表 4 所示,将输入图像分辨率从 416×416 像素逐渐增大至 608×608 像素,本模型的性能提升不大。增大输入图像的尺寸会增加计算量,降低模型的检测效率,且输入图像的尺寸过大,其训练过程需要更多的内存和显存来存储和处理图像,限制模型的可扩展性和可移植性。因此,选用 416×416 像素作为输入尺寸。

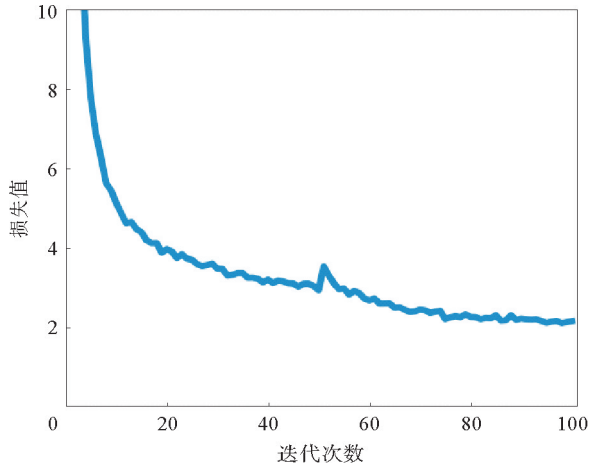


图 12 改进 YOLOv3 在训练过程中的损失值变化曲线

Fig. 12 Loss value curve of improved YOLOv3 during the training process

表 2 模型对比实验结果

Table 2 Modelcomparison experimental results %

模型	导线端检测精度	横担端检测精度	m_{AP}
原 YOLOv3	87.40	83.83	85.62
改进 YOLOv3	96.41	93.05	94.73

表 3 模块消融实验结果

Table 3 Module ablation experimental results

算法	m_{AP}/%	FPS	单张图片的检测时间/ms
YOLOv3	85.62	47.75	20.9
YOLOv3+分组卷积	85.88	48.29	20.7
YOLOv3+ McarB	94.33	45.74	21.8
YOLOv3+McarB+分组卷积	94.73	46.03	21.7

表 4 图像输入尺寸对模型性能的影响

Table 4 Effect of image input size on model performance

输入图像尺寸	m_{AP}/%	FPS	单张图片的检测时间/ms
416×416	94.73	46.03	21.7
512×512	94.82	43.69	22.9
608×608	94.87	40.25	24.8

4) 算法对比

改进 YOLOv3 算法与目标检测经典算法的性能指标对比结果如表 5 所示。从表 5 中可以看出,Faster R-CNN 相较于其他算法检测性能较差,m_{AP} 值为 88.68%,且模型较大,在检测效率上不如本算法。相较于 RetinaNet,改进 YOLOv3 算法在精度和效率两方面上都有提高,其中检测精度提高 3.81%,单张照片的检测时间减少 2.2 ms。

目前现有的巡检照片目标检测算法有 YOLOv3+SE 与 MS-PAD 两种,本算法的检测类别置信度为 0.95,较其他两种算法更高。从实验数据来看,与 YOLOv3+SE 算法相比,本算法在检测精度上提升 2.45%,并且单张巡检照片的检测时间减少 0.6 ms。与 MS-PAD 算法相比,本算法的各项指标均优于 MS-PAD。更高版本的 YOLOv4、YOLOv5 系列模型更加轻量并且具有更快的检测速度,但对比检测精度,本算法比 YOLOv4 算法高 2.5%,比 YOLOv5s 高 2.86%,比 YOLOv5m 高 0.34%。由于巡检照片分类对检测

精度的要求更高,故在巡检照片分类任务中本算法更适用。YOLOv5 系列算法由于模型更加轻量,更适用于无人机的实时缺陷检测任务。

表 5　算法对比实验结果

Table 5　Results of algorithm comparison experiment

算法	主干	m_{AP}/%	FPS	单张图片的检测时间/ms
Faster R-CNN[4]	VGG	88.68	29.94	33.4
RetinaNet[17]	ResNet50	90.92	41.83	23.9
YOLOv3+SE[11]	Darknet53	92.28	44.78	22.3
MS-PAD[9]	VGG	89.27	33.23	30.1
YOLOv4[18]	Modified CSP	92.23	52.11	19.2
YOLOv5s	CSPDarknet	91.87	107.14	9.3
YOLOv5m	CSPDarknet	93.39	88.13	11.3
本算法	本算法主干	94.73	46.03	21.7

4　结论

本研究提出一种基于改进 YOLOv3 模型的巡检照片分类命名方法,融合通道洗牌与通道注意力机制,提高了模型的特征提取能力;使用分组卷积代替主干网络中的部分卷积,减少了运算量,提高了模型的训练效率。实验表明,相比于改进前 YOLOv3 算法,虽然本方法模型检测时间增加了 0.8 ms,但是平均精度提高了 9.11%,在巡检照片分类任务中更具优势。相比于其他图像分类方法,本方法可以帮助巡检人员更有效率地筛查出质量差的巡检照片,能够满足电力巡检过程中巡检照片分类的需要。

参考文献:

[1] 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. Piscataway:IEEE Press,2014:580-587.

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

[3] HE K,ZHANG X,REN S,et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(9):1904-1916.

[4] REN S,HE K,GIRSHICK R,et al. FasterR-CNN:Towards real-time object detection with region proposal networks[J]. Advances in Neural Information Processing Systems,2015,28(2):23-30.

[5] 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. Piscataway:IEEE Press,2016:779-788.

[6] ZHAO Z Q,ZHENG P,XU S,et al. Object detection with deep learning:A review[J]. IEEE Transactions on Neural Networks and Learning Systems,2019,30(11):3212-3232.

[7] ZHOU Q,LI Q,XU C,et al. Class-aware edge-assisted lightweight semantic segmentation network for power transmission line inspection[J]. Applied Intelligence,2023,53(6):6826-6843.

[8] 王振.输电线路激光点云数据挖掘与应用[D].济南:山东大学,2020.
WANG Zhen. Laser point cloud data mining and application for transmission lines[D]. Jinan:Shandong University,2020.

[9] VIEIRA-E-SILVA A L B,DE CASTRO FELIX H,DE MENEZES CHAVES T,et al. STNplad:A dataset for multi-size power line assets detection in high-resolution uav images[C]//34th SIBGRAPI Conference on Graphics,Patterns and Images (SIBGRAPI). Piscataway:IEEE Press,2021:215-222.

[10] BAYASGALAN Z,SUKHBAATAR A,TUDEVDAGVA U,et al. Top inspection and monitoring of HV power line tow-

ers damage by UAV[C]//2021 International Conference on Technology and Policy in Energy and Electric Power (ICTPEP). Piscataway:IEEE Press,2021:248-252.

[11] OHTA H,SATO Y,MORI T,et al. Image acquisition of power line transmission towers using UAV and deep learning technique for insulators localization and recognition[C]//23rd International Conference on System Theory,Control and Computing (ICSTCC). Piscataway:IEEE Press,2019:125-130.

[12] ZHANG X,ZHOU X,LIN M,et al. Shufflenet:An extremely efficient convolutional neural network for mobile devices[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2018:6848-6856.

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

[14] 邵延华,张铎,楚红雨,等. 基于深度学习的 YOLO 目标检测综述[J]. 电子与信息学报,2022,44(10):3697-3708.
SHAO Yanhua,ZHANG Duo,CHU Hongyu,et al. A review of YOLO object detection based on deep learning[J] Journal of Electronics and Information Technology,2022,44(10):3697-3708

[15] 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. Piscataway:IEEE Press,2016:770-778.

[16] XIE S,GIRSHICK R,DOLLÁR P,et al. Aggregated residual transformations for deep neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2017:1492-1500.

[17] 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. Piscataway:IEEE Press,2017:2980-2988.

[18] WANG C Y,BOCHKOVSKIY A,LIAO H Y M. Scaled-yolov4:Scaling cross stage partial network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE Press,2021:13029-13038.

(责任编辑:傅 游)