

基于改进 YoloV4 的电网变压器 油液渗漏检测方法

陆志欣, 田涵宁, 郭国伟

(广东电网 佛山供电局, 广东 佛山 528000)

摘要: 及时发现电网变压器油液渗漏问题对于电网的安全与稳定运行尤为重要; 传统电网变压器油液渗漏检测主要依赖于人工定期巡检, 但人工巡检无法实现全天候监测, 具有滞后性; 当前主流目标检测模型应用于电网变压器油液渗漏检测时, 存在检测速度较慢、准确率低和鲁棒性较差等问题, 无法满足实际应用; 为此提出一种改进 YoloV4 的变压器油液渗漏检测方法; 首先, 通过引入 Mobile-ViT 作为模型的骨干结构, 利用卷积和 Transformer 结构有效提取目标的局部和全局信息特征, 降低计算量; 其次, 提出多尺度特征融合层, 旨在实现局部和全局信息的多尺度特征融合, 增强上下文语义表达, 用以更好地实现电网变压器油液渗漏检测; 实验结果表明, 该方法在电网变压器油液渗漏数据集上检测精度达到了 95.3%, 检测速度达到了 50.6 fps; 相较于原生 YoloV4 方法检测精度提高了 2.6%, 检测速度提升了 2.6 fps; 经实际应用, 该方法部署在边缘设备上推理速度也达到了 43 fps, 满足了实际工程的需求。

关键词: 电网; 变压器; 目标检测; 油液渗漏检测; YoloV4; Mobile-ViT; 多尺度特征; 特征融合

Oil Leakage Detection Method for Power Grid Transformers Based on Improved Yolov4

LU Zhixin, TIAN Hanning, GUO Guowei

(Foshan Power Supply Bureau, Guangdong Power Grid Corporation, Foshan 528000, China)

Abstract: It is particularly important for the safe and stable operation of power grid to find the oil leakage problem of power grid transformers in time. Traditional power transformer oil leakage detection mainly depends on manual regular detection, but manual detection cannot achieve all-weather monitoring and has lag. When the current mainstream object detection model is applied to the oil leakage detection of power grid transformers, there are some problems such as slow detection speed, low accuracy and poor robustness. It cannot meet the practical application. For this reason, an improved you only look once version 4 (YoloV4) transformer oil leakage detection method is proposed. Firstly, by introducing Mobile Vision Transformer (Mobile-ViT) as the backbone structure of the model, the local and global information features of the object are effectively extracted by the convolution and Transformer structure, which reduces the computation. Secondly, a multi-scale feature fusion layer is proposed, which aims to realize the multi-scale feature fusion of local and global information and enhance the context semantic expression, so as to better realize the oil leakage detection of power grid transformers. The experimental results show that the detection accuracy of this method on the power grid transformer oil leakage data set reaches 95.3%, and the detection speed reaches 50.6 fps; Compared with the native YoloV4 method, the detection accuracy is improved by 2.6%, and the detection speed by 2.6 fps. After practical application, the reasoning speed of this method deployed on edge devices also reaches 43 fps, which meets the needs of practical engineering.

Keywords: power grid; transformer; object detection; oil leakage detection; YoloV4; Mobile-ViT; multi-scale features; feature fusion

0 引言

变压器作为电网“变电”和“配电”环节的核心电器设备, 保证其安全稳定运行意义重大^[1]。当前变压器主要分为两种: 干式变压器^[2]和油浸式变压器^[3], 油浸式变压器因制作工艺简单、造价成本低、环境适应性强、寿命长等优

点, 而被广泛应用, 尤其是在特高压线路和低压户外场景。油浸式变压器相较于干式变压器, 最大的区别就是油浸式变压器为实现散热和绝缘功能, 内部含有大量的介质液体“油”, 俗称变压器油, 是变压器的重要工作介质, 能起到冷却和灭弧等保护作用。

而在变压器的工作过程中因质量、安装运输、工作环

收稿日期: 2023-06-28; 修回日期: 2023-07-24。

基金项目: 南方电网公司科技项目(GDKJXM20220216)。

作者简介: 陆志欣(1983-), 男, 大学本科, 高级工程师。

通讯作者: 郭国伟(1980-), 男, 硕士, 高级工程师。

引用格式: 陆志欣, 田涵宁, 郭国伟. 基于改进 YoloV4 的电网变压器油液渗漏检测方法[J]. 计算机测量与控制, 2024, 32(2): 85-92.

境和负荷多方面原因会造成电网变压器油液渗漏^[4]，不仅影响了变压器散热效率，而且导致水分进入、绝缘强度下降，从而引起绝缘击穿、绕组烧毁等事故，将严重威胁电网的正常运行。

为减少电网变压器油液渗漏情况的出现，电网工作人员会定期对变压器进行巡检，但因人力成本的原因，巡检频率较低，一般为月检或季检，无法实现全天候地对变压器进行监测。

近年来，电网的重要设备周边都部署了监控摄像头，如何利用计算机视觉技术实时监测设备运行状态，减少巡检人力资源的消耗和保障巡检人员的安全成为电网的迫切需求。

目前常用于电网设备监测的计算机视觉方法主要分为两种：一是基于机器学习的图像处理方法，通过分析应用场景下图像的自有属性特征如颜色、轮廓、大小、纹理形状等人为设计特征^[5-7]，分析出图像的标志性的信息，标记和识别出图像内容，最后通过分类器分类。此类方法特征提取简单，计算量较小，在某些单一场景下表现优秀，但也受限于人工设计特征提取方法的局限性，在现实复杂场景下，易受复杂背景、光照亮度等影响和干扰，方法的鲁棒性较差，可用性不强；二是基于深度学习的目标检测算法，通过迁移当前广泛应用于通用领域具有优异性能的目标检测算法，如 SSD (Single Shot multi-boxes Detector)^[8]、YOLO (You Only Look Once)^[9-13]、Retina Net^[14]、Faster RCNN (Region Convolution Neural Network)^[15-16]、Mask RCNN^[17-18]等主流目标检测方法，将电网设备正常运行状态和异常状态作为两个不同目标来检测，但这类目标检测算法都采用卷积网络模型作为主干网络如 VGG^[19]、Residual Net^[20]、Dense Net^[21]等分类模型实现特征提取，其出发点是卷积结构在视觉图像处理中拥有较好的空间局部偏置归纳性，可以帮助模型较好地提取局部特征和模型学习，但是局部偏置归纳的特性也限制了模型的全局感知能力和性能。

随着 Transformer^[22] 的提出，Transformer 结构在自然语言处理领域有着广泛的应用，且表现非凡。ViT^[23] 将 Transformer 引入到图像处理领域，并取得了超越卷积神经网络的性能。Transformer 结构可以实现长距离跨度的相似度或关联度学习，实现全局的信息感知，可以有效地提升模型的表现。但图像的数据维度较高，进行高分辨率下的推理计算复杂度高，且优化较不稳定。为了加快推理速度，且进一步提升模型的性能，Mehta 等人^[24] 提出了 Mobile-ViT 提出一种轻量级的 ViT 分类网络模型，保证了较低计算量的同时也兼顾了卷积和 Transformer 的特性。

为了进一步实现对电网变压器油液渗漏快速且准确的检测，本文提出了一种改进 YoloV4 的电网变压器油液渗漏检测方法。本文贡献主要有以下几点：

1) 本文提出了一种改进 YoloV4 的网络模型，使用

Mobile-ViT 作为模型的骨干结构，利用卷积和 Transformer 有效地提取目标的局部和全局信息特征，降低了计算量。

2) 本文在改进 YoloV4 模型的基础上，针对性地提出了一种多尺度特征融合方法，实现了局部和全局信息的多尺度特征融合，增强了上下文语义表达，有效提升算法的检测精度。

3) 本文提出一种改进 YoloV4 的电网油液渗漏检测方法及整体系统架构，同时构建了电网油液渗漏相关数据集。

1 电网变压器油液渗漏检测系统及方法

1.1 系统总体结构

针对电网变压器油液渗漏实际应用需求，本文设计了一种油液渗漏检测系统，系统的总体结构如图 1 所示。主要包含有摄像终端，边缘推理，云端数据和运维中心等 4 个部分。

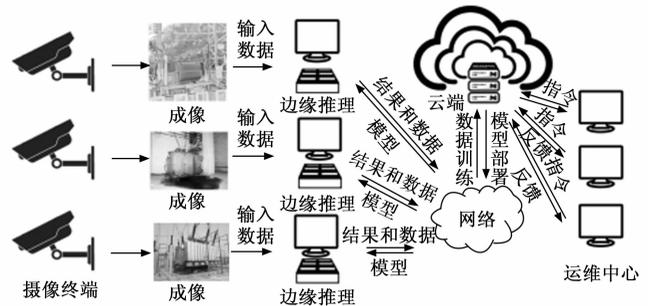


图 1 油液渗漏检测系统结构

其中摄像终端一般指包含有高清摄像头，通过摄像头实现对目标 24 小时的监控成像，获取高分辨率图像数据。

边缘推理设备包含有硬件部分和软件部分。其中硬件部分采用 Atlas 200DK，主板芯片为昇腾 310，内存规格为 LPD-DR4X 8G，支持 16 路 H.264 硬件解码，功耗为 20 W。软件部分包含有 Linux-18.04 操作系统，Mind Spore 推理框架。通过将图像进行模型处理后得到输出结果，将输出结果与图像通过网络同时输入到云端。

云端服务器包含有两个方向的信息处理：一方面为接收边缘设备的数据和预测结果。将获取的图像数据进行存储和进一步标注后的模型训练。云端服务器也可以将数据和指令传输给边缘服务器实现模型的远程部署和更新等相关任务。另一方面服务器可以接收和反馈运维中心的指令，运维中心通过下达指令可以调控和管理云端的数据存储、标注、模型训练及模型远程部署等相关任务。此外，云端服务器会将检测结果直接反馈给运维中心，并进行相应的处理结果提示。

通过基于深度学习的智能监控系统可以有效地实现变压器油液渗漏的 24 小时的远程全天候的监管、训练和升级，这大大地提升了电网的智能化水平，有效地降低工作量，降低人力成本，提高系统的工作效率。本系统的核心部分为基于深度学习的模型检测算法。

1.2 YoloV4 算法模型

YoloV4 目标检测算法结构主要可分为 3 部分, 主干网络、颈部结构和预测层 3 部分构成, 如图 2 所示。

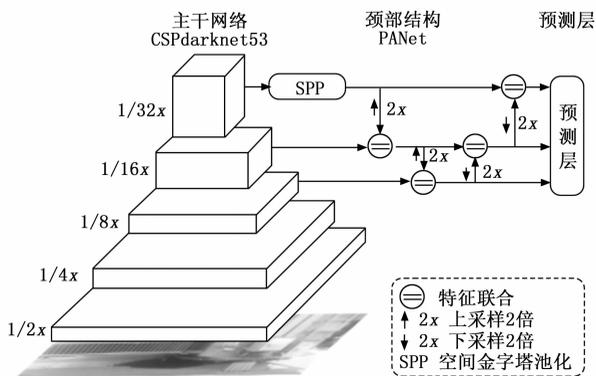


图 2 YOLOV4 模型结构

其主干网络采用 CSPdarknet-53 网络模型作为特征提取模型; 从主干网络中获取 3 个特征尺度经过颈部层进行信息的整合, 预测层在整合后的特征上进行目标预测。CSPdarknet-53 是一个纯卷积的网络结构, 其提取的特征信息具备较强的空间归纳偏置; 此外, CSPdarknet-53 在深层尾部结构中增加了 SPP (spatial pyramid pooling) 结构, 一方面, 有利于扩大对全局信息的感知。尽管如此, 但卷积结构缺乏足够的全局信息获取以及不同位置之间的相似关联度感知用以目标预测, 导致基于 YoloV4 的目标检测算法对液面和阴影的检测效果不佳。此外, CSPdarknet-53 网络结构的计算复杂度相对较高, 在边缘设备上推理速度较慢。

颈部结构采用 PA Net 结构, 通过多尺度聚合不同尺度的特征用以融合多尺度上下文信息, 对增强特征表达具有一定的效果。但是此结构特征层次串联度严重, 耦合程度较高, 浅层的特征需深层网络依次上采样后融合处理。再将获取到的融合特征送入预测层进行特征预测。此结构不利于并行计算, 拖慢了检测推理速度。

预测层包含了分类、坐标回归和置信度 3 种预测分支用于颈部结构输出特征的检测与分类。

1.3 改进的 YoloV4 算法模型

针对上述问题和实际需求, 为进一步提升 YoloV4 算法在电网变压器油液渗漏的检测性能。本文设计了一种改进 YoloV4 的油液检测模型, 其结构如图 3 所示。本部分将从主干网络、颈部结构、预测层 3 部分详细介绍其结构。

1.3.1 主干网络

为了改善原有 CSPdarknet-53 卷积模型对全局信息感知的不足和加快推理速度。改进的 YoloV4 模型引入 Mobile-ViT 结构代替原有 CSPdarknet-53 作为主干网络。Mobile-ViT 模型由 MobileNetV2 区块和 Mobile-ViT 区块堆叠构成, 去除分类任务带有的 1×1 卷积层、全局平均池化层和全连接层。其具体结构如表 1 所示。

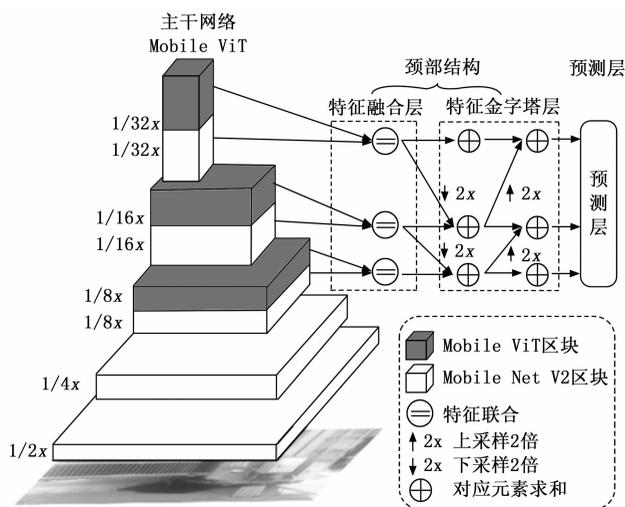


图 3 改进 YOLOV4 模型结构。

表 1 主干网络(Mobile-ViT)结构表

序号	结构层	下采倍数	重复次数	输出通道
1	Image	1		
2	Conv3x3	2	1	16
3	MV2-S1	2	1	32
4	MV2-S2	4	1	64
5	MV2-S1	4	2	64
6	MV2-S2	8	1	96
7	MViTBlock-L2	8	1	96
8	MV2-S2	16	1	128
9	MViTBlock-L4	16	1	128
10	MV2-S2	32	1	160
11	MViTBlock-L3	32	1	160

其中: MV2 表示 MobileNetV2 区块, S1、S2 表示 Stride 为 1 和 2, MViTBlock 表示 Mobile-ViT 区块, LX 表示 Transformer 的层数为 x 层。

模型输入图像的尺寸为 256×256 , 分别于模型 $\frac{1}{8}x$, $\frac{1}{16}x$, $\frac{1}{32}x$ 的 MobileNetV2 区块后依次增加 Mobile-ViT 区块。

MobileNetV2 区块用于提取带有偏置归纳的卷积局部特征信息; Mobile-ViT 区块通过计算不同区域之间信息的关联度, 可以有效获取具备全局感知的注意力信息特征。

此外, 由于 Transformer 结构计算复杂度高, 基于低分辨率下的特征处理有利于提升模型的运算速度。下面将详细介绍 MobileNetV2 区块和 Mobile-ViT 区块结构。

1) MobileNetV2 区块;

MobileNetV2 区块采用深度可分离卷积有效地降低了卷积的运算复杂度和可训练参数数量, 极大地提高了模型的推理速度。

通过 1×1 的卷积实现特征维度的升维, 缓解狭窄维度

会降低信息流特征传递的问题,采用 3×3 的深度可分离卷积提取表达特征,再采用 1×1 的卷积还原其输入维度。在本网络中包含有两种类型的 MobileNetV2 区块结构,如图 4 所示。步长 S 取值为 1 的区块结构如图 4 (a) 所示,采用残差连接,防止反向传播时梯度消失,其输入输出特征尺度保持不变;步长 S 取值为 2 的区块结构如图 4 (b) 所示,用于降低特征尺度,实现特征下采样。

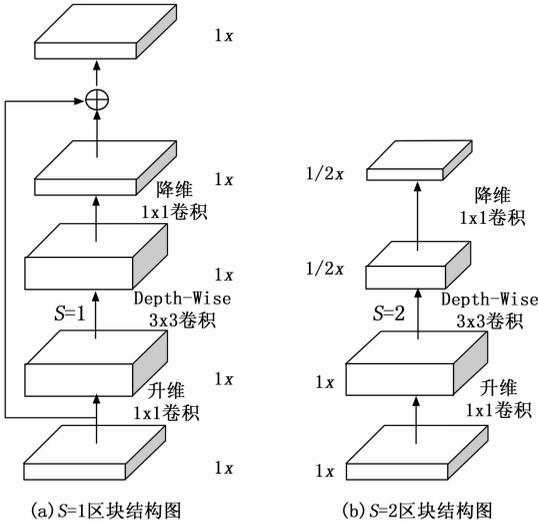


图 4 MobileNetV2 区块结构

2) Mobile-ViT 区块:

受限于 Transformer 处理需要二维线性数据流,因此,需将输入的三维特征转换为二维数据进行 Transformer 处理,将处理后输出进行维度还原回复空间维度。因此,Mobile-ViT 区块包含了展开、Transformer 层和折叠等三部分组成。其具体结构如图 5 所示。

展开部分,输入特征先后经过 3×3 和 1×1 的卷积,得到尺度为特征 F 将 F 划分成 N 块 patches,每块为 F_s ,其具体表达如公式 (1) 所示:

$$F \in \mathbb{R}^{(H \times W \times d)} \xrightarrow{split} NF_s \in \mathbb{R}^{(h \times w \times d)} \quad (1)$$

式中, H, h 表示特征高; W, w 表示特征宽; d 表示通道数; N

$$= \frac{H}{h} \times \frac{W}{w}; \text{其中 } H \text{ 要整除 } h, W \text{ 要整除 } w.$$

将 $F_s \in \mathbb{R}^{(h \times w \times d)}$ 按照 h, w 维度进行压平,得到 $F_p \in \mathbb{R}^{(p \times d)}$,其具体表达式如式 (2) 所示:

$$F_s \in \mathbb{R}^{(N \times h \times w \times d)} \Rightarrow NF_s \xrightarrow{flatten} NF_p \Rightarrow F_p \in \mathbb{R}^{(N \times p \times d)} \quad (2)$$

式中, $p = w \times h; F_p$ 表示展开后的特征; F_s 表示 split 后的特征。

然后,经过 Transformer 层部分的处理, $F_s(N, h, w, d)$ 转变为 $F_p(N, p, d)$,此部分展开处理与原有的不同,会将 h, w, d 三个特征维度进行压缩,得到二维的特征。然而本区块展开后的特征 F_p 为三维特征。在本区块中,将得到的特征输入 F_p 进行按 n 次拆分输入 Transformer 的处理结构中,其结构如图 6 所示。

输入特征拆分为 $p_1 \dots p_n, F_p \in \mathbb{R}^{(N \times 1 \times d)}$,其中 $1 \leq i \leq n$; F_p 经过线性层得到 $F_Q, F_K, F_V \in \mathbb{R}^{(nd)}$ 共 3 个特征,将 d 分成 j 组,每组为 d_k 个进行多个注意力处理。其表达式如公式 (3) 所示:

$$F_t = \sigma\left(\frac{F_Q \times F_K^T}{\sqrt{Z}}\right) \quad (3)$$

式中, Z 表示元素个数, \times 表示矩阵乘法, σ 表示 Softmax 函数; $F_t \in \mathbb{R}^{(N \times N)}$ 。

将得到的 F_t 与 F_v 矩阵相乘得到 F_a' ,再将特征联合其它多头得到 F_a 。其表达如公式 (4)、(5) 所示:

$$F_a' = F_t \times F_v \quad (4)$$

$$F_a = \text{Concat}(F_a') \quad (5)$$

式中, $F_t \in \mathbb{R}^{(N \times N)}$ 表示 Q 特征, $F_v \in \mathbb{R}^{(N \times d_k)}$ 为 V 特征, $d_k = \frac{d}{j}$,表示分组特征分为 j 组,每组数量为 $d_k, F_a \in \mathbb{R}^{(N \times d)}$ 表示注意力特征, Concat 表示融合层;

将 F_a 通过激活层和两层线性层得到 F_{out} ,其具体表达式如式 (6) 和式 (7) 所示,通过聚合 n 层输入折叠层。

$$F_{p_i} = \text{FFN}(\text{GeLU}(F_a)) \quad (6)$$

$$F_p = \text{Concat}(F_{p_i}) \quad (7)$$

式中, $F_{p_i} \in \mathbb{R}^{(N \times d)}$ 表示; $\text{FFN}(\cdot)$ 表示两层线性层, GeLU 表示高斯误差线性激活层。

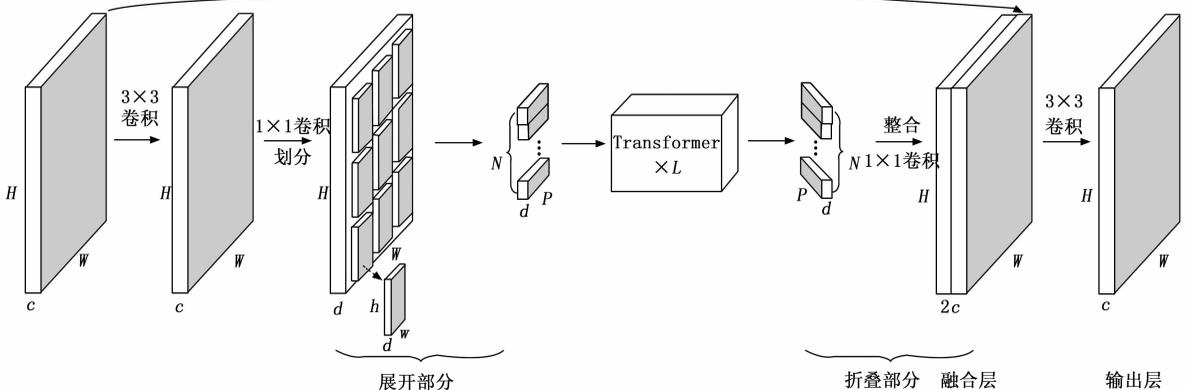


图 5 Mobile-ViT 区块结构

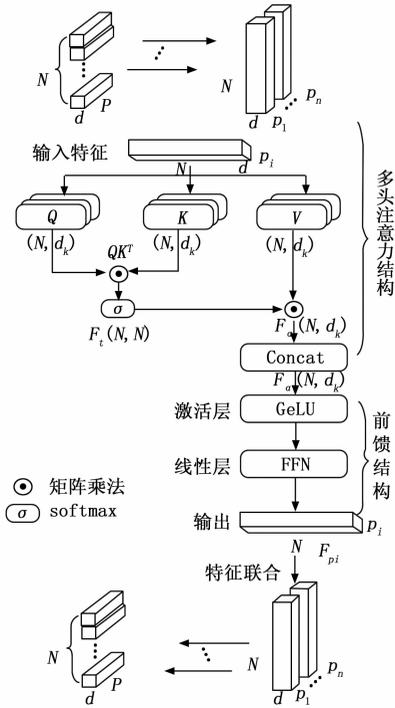


图 6 Transformer 结构

在原有的 Transformer 自注意力结构需要与 0 特征图里面每个区域进行相似关联度计算, 如图 7 (a) 所示。但在本文中, Transformer 中每个 patch 采用了 $p = w \times h$ 分块操作, 并分别进行注意力特征计算, 由于 patch 在输入之前进行了 3×3 的卷积核计算, 而每个 patch 的 w 和 h 的选取都为 2, 小于 3。如图 7 (c) 所示。patch 中的元素可以选择只与相同位置的元素进行相似度计算。而不需要像 Transformer 一样对所有的位置进行计算, 一定程度上提升了 Transformer 推理的速率, 如图 7 (b) 所示。

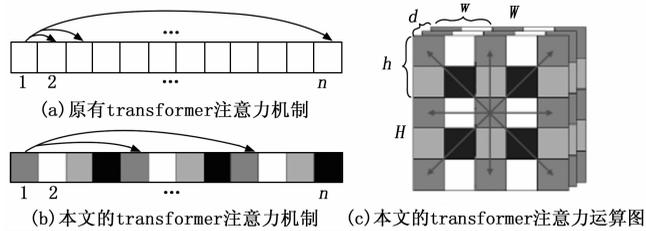


图 7 Transformer 自注意力机制对比

折叠层部分与展开层刚好相反, 将展开的部分还原成三维特征, 如图 5 所示。其具体的表达式如式 (8) 和式 (9) 所示:

$$F_p \in \mathbb{R}^{(N \times p \times d)} \Rightarrow NF_p \xrightarrow{\text{reverse}} NF_s \Rightarrow F_s \in \mathbb{R}^{(N \times h \times w \times d)} \quad (8)$$

$$F_s \in \mathbb{R}^{(N \times h \times w \times d)} \xrightarrow{\text{reshape}} F \in \mathbb{R}^{(H \times W \times d)} \quad (9)$$

式中, 符号表达与式 (1) 和式 (2) 中的一致, 并先后经过 1×1 卷积和 3×3 卷积得到输出特征, 其表达如式 (10) 所示:

$$F = \text{Conv}3 \times 3(\text{Conv}1 \times 1(F)) \quad (10)$$

式中, F 表示特征, $\text{Conv}3 \times 3$ 表示 3×3 卷积, $\text{Conv}1 \times 1$ 表示 1×1 卷积操作。

1.3.2 颈部结构

颈部结构可分为特征融合层和特征金字塔层两个模块。特征融合层分别从主干网络中多个特征尺度上获取卷积特征和 Transformer 自注意力特征; 特征金字塔层通过采用并行的双层金字塔结构实现特征多尺度运算, 如图 2 所示。

特征融合层分别从主干网络的 3 个特征尺度上引出一组卷积特征和 Transformer 自注意力特征, 将提取的特征分别经过 1×1 的卷积层进行特征联合得到多尺度特征。

特征金字塔层可分为两层: 第一层是上采样过渡层, 第二层为下采样过渡层。在上采样过渡层中, 将输入的多尺度特征并行通过上采样与浅层特征进行相加融合, 其融合的过程为: 深层特征采用双线性插值的方式实现 $2x$ 上采样后, 结合 1×1 的卷积与浅层特征实现通道维度对齐; 浅层特征通过 1×1 的卷积层与上采样后的深层特征通过对应元素相加的方式融合。

在下采样过渡层中, 将浅层特征通过下采样与深层特征进行融合, 下采样过渡层中的浅层特征采用 2×2 的不重叠平均池化层实现特征下采样。其余操作与上采样过渡层基本相同。

与原有算法相比, 特征融合层整合了卷积特征和 Transformer 自注意力特征, 有利于整合局部偏置归纳信息和全局注意力特征信息。金字塔层通过采用并行的双层金字塔结构实现多尺度信息融合和运算解耦, 有利于提升模型精度和提高模型推理速度。

1.3.3 预测层

预测层包含有 3 个检测分支, 分别为置信度、分类和坐标回归 3 种。将输入的 3 个尺度的预测特征分别采用 3×3 的卷积层进行目标预测。本置信度表示为目标预测类别的概率值, 置信度预测范围为 $[0, 1]$, 由 $P_r(\text{Object}) * \text{IOU}(\text{Pred}, \text{Groundtruth})$ 组成, 其中 $P_r(\text{Object})$ 包含物体则表示 $P_r(\text{Object})$ 取 1, 否则取值为 0。判断是否包含该类物体, $\text{IOU}(\text{Intersection over Union})$ 表示预测框与目标框的重叠区域占总覆盖区域占比, 这有利于筛选出背景和前景目标区域。类别预测是对包含有目标的框进行概率预测, 其值范围为 $[0, 1]$, 判断该目标框的概率匹配某类别的概率, 总共包含 C 个类。回归框预测包含有 4 个值 (x, y, w, h) 其中 x 表示横坐标, y 表示纵坐标, w 表示宽度, h 表示高度。Anchor 的个数和尺寸由数据集由聚类统计得到。因此, 每次需要预测 $K * (4 + 1 + C)$ 个值, K 为 Anchor 个数。

1.4 优化目标

对应预测层的 3 个检测分支, 对应的优化目标损失包含有置信、分类、坐标回归 3 类损失。其中置信损失采用 MSE 损失计算, 其如式 (11) 所示:

$$L_{conf} = \sum_{i=0}^N \sum_{j=0}^K P_{ij}^{obj} * (\hat{C}_i - C_j)^2 \quad (11)$$

式中, N 表示预测框的个数, K 表示 anchor 的个数。 P_{ij}^{obj} 表示第 i 个网格中第 j 个 box 是否包含目标, 包含则为 1, 不包含则为 0。 \hat{C}_i 表示预测框与 anchor 的 IOU 置信度标签, C_j 表示第 j 个真实的 IOU 值。

分类损失采用交叉熵损失计算, 其具体表达式如公式 (12) 所示:

$$L_{class} = - \sum_{i=0}^N \sum_{j=0}^K P_{ij}^{obj} * \sum_{c \in class} \{ \hat{p}_i(c) \log [p_i(c)] + [1 - \hat{p}_i(c)] \log [1 - p_i(c)] \} \quad (12)$$

式中, $class$ 表示类别数, $\hat{p}_i(c)$ 表示第 i 个网格类别 c 是否匹配, 若预测类别和真实类别匹配则为 1, 否则为 0, $p_i(c)$ 为预测为 c 类的概率分数, 范围为 $[0, 1]$ 。

坐标回归损失主要衡量预测真实值与 anchor 之间的偏移值的区间, 这样可以有效地降低坐标预测的不稳定性。采用 MSE 作为其损失如公式 (13) 所示:

$$L_{loc} = \sum_{i=0}^N \sum_{j=0}^K P_{ij}^{obj} [(t_{xi} - \hat{t}_{xi})^2 + (t_{yi} - \hat{t}_{yi})^2 + (t_{wi} - \hat{t}_{wi})^2 + (t_{hi} - \hat{t}_{hi})^2] \quad (13)$$

式中, $\hat{t}_{xi}, \hat{t}_{yi}, \hat{t}_{wi}, \hat{t}_{hi}$ 分别为预测第 i 个网格目标的中心点横坐标、纵坐标, 宽度偏移值和高度偏移值。 t_{xi}, t_{yi}, t_{wi} 及 t_{hi} 分别表示第 i 个网格的中心点横坐标、纵坐标、宽度偏移值和高度偏移值。其真实值的计算相关的表达式如公式 (14) 和 (15) 所示:

$$t_{xi} = G_{xi} - C_{xi}; t_{yi} = G_{yi} - C_{yi} \quad (14)$$

$$t_{wi} = \log(G_{wi}/C_{wi}); t_{hi} = \log(G_{hi}/C_{hi}) \quad (15)$$

式中, G 表示真实框, C 表示 anchor 的坐标, x, y, w, h 表示中心点的坐标、宽度和高度。

总体的损失为置信、分类和坐标回归 3 类损失之和。其表达如公式 (16) 所示:

$$L = \lambda L_{loc} + \beta L_{class} + \eta L_{conf} \quad (16)$$

式中, L 表示总体损失, λ, β 和 η 为系数常数, 分别取值为 1、1 和 0.5。

1.5 电网变压器油液渗漏检测算法模型实现流程

电网变压器油液渗漏检测算法模型具体实现的流程如图 8 所示。首先将输入图像进行尺寸缩放, 将图像固定在 256×256 的输入大小, 再将图像均值化, 采用训练样本的均值和方差作为图像均值化的估计参数。

然后, 将图像输入到 Mobile ViT 主干网络中进行多尺度特征提取, 将提取的特征实现多尺度特征融合, 进而构建出多尺度的特征金字塔层特征; 在多尺度特征上进行目标预测, 得到大量的不同置信度和目标区域重叠的目标候选框, 需要进一步进行筛选, 主要由置信度阈值和重叠区域非极大抑制两方面构成。一方面, 通过设置置信度阈值排除一部分低置信度的目标候选区域; 另一方面将筛选出高置信度区域进行目标框非极大抑制操作 (Non-Max Suppression), 将挑

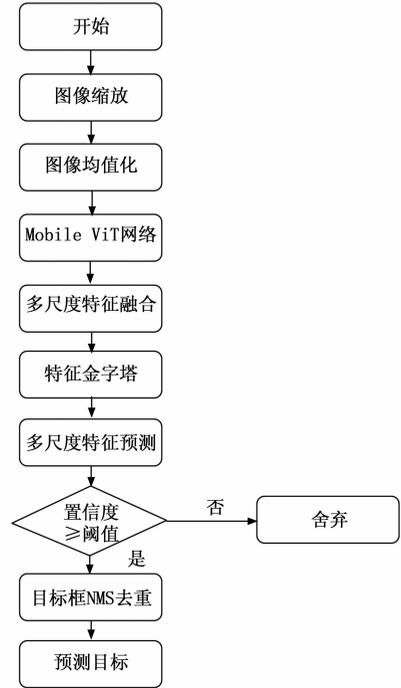


图 8 电网变压器油液渗漏检测算法模型流程

选出置信度最高的目标框作为正样本和标准, 依次计算与其它目标框的交并比。最后, 将交并比高的目标框进行去除, 再继续从剩下的低交并比目标框中重复此操作, 直到筛选完所有的目标框, 此操作进一步去除了高置信度且重叠的目标框, 得到最后的目标预测框。

2 实验和结果分析

2.1 数据采集及增广

目前, 针对电网中电网变压器油液渗漏领域内, 没有通用的开源数据集。通过整理配电日常工作中变压器出现油液渗漏异常时的视频数据, 对其进行了筛选, 截取电网变压器油液渗漏图像, 同时结合网络上开源的电网变压器油液渗漏图像, 总共获得了 4 000 张电网变压器油液渗漏图像, 形成了电网变压器油液渗漏数据集。图 9 为本文构建数据集的部分数据示例。

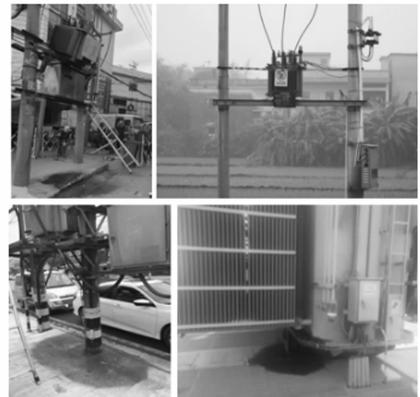


图 9 电网变压器油液渗漏样本示例

再根据 4:1 的原则做训练集测试集划分。训练集的图片数据为 3 200 张, 测试集的图片数据为 800 张。

同时本文在训练算法模型前, 采用了旋转、翻转、明暗度调整、旋转、随机擦除和 Mosaic 等方式对训练集进行了数据增广。其示例如图 10 所示。

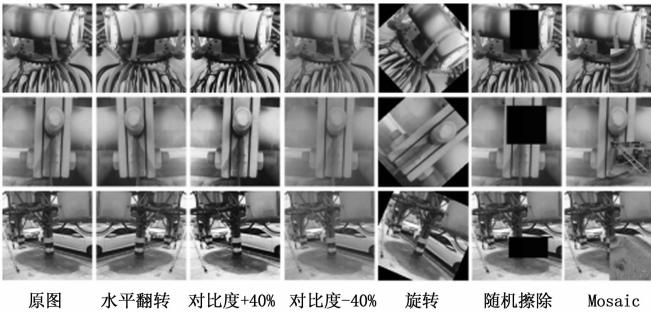


图 10 数据集增广图像

2.2 实验环境及参数

本文实验固定输入图片尺寸为 256×256 , 采用在 Image Net 上预训练的模型作为主干网络。在训练中将固定主干模型的参数, 同时, 新增层参数采用 Kaiming 初始化方式进行初始化。在检测训练中, 本文设定学习率为 0.001, 终止学习率为 1×10^{-7} , 采用步进学习率 (Step_lr) 的方式训练 35 个周期, 分别在 10, 20, 25, 30 上逐次降低 0.1, 批尺度为 64。采用自适应随机梯度下降的方式进行权重优化, 其中衰减权重系数设置为 0.95。锚定框尺寸采用 YoloV4 的聚类策略进行设定, 其长宽尺寸比为 1:1, 1:2, 2:1, 1:5。本文实验设备为 NVIDIA RTX2080ti GPU 计算平台, 在 UBUNTU18.04 操作系统下, 使用 PyTorch2.0 深度学习框架训练算法模型。

2.3 模型评估指标

为了评估检测模型的检测效果, 本文统一采用目标检测常用的精度评价指标, 值越大表示检测模型的总体识别准确率越高。同时, 为了准确测试出目标检测模型检测的速度, 本文统一采用速度评价指标 FPS (frame per second), 即模型每秒处理图像的张数, 其值越大, 证明模型检测速度越快。

2.4 消融实验

为了验证本文方法各部分的有效性, 本文对提出的改进的方法 (MViT-YoloV4) 各组成部分进行消融实验, 消融实验的结果如表 2 所示。消融结果表明: 直接结合 MobileNet ViT 和 YoloV4 的目标检测方式相对原有 YoloV4 目标检测方法有着一定差距, 单纯引入注意力网络结构并不能带来性能上的提升。数据增广对模型性能有一定的提升; 增加特征融合层提升了 2.7 个百分点, 有效地说明了全局信息特征和局部信息特征的有效融合有利于提升模型的检测效果; 增加特征金字塔层可以提升了 3.2 个百分点; 特征融合层与特征金字塔层的增加对模型的表现效果最佳, 二者总共提升了 5.2 个百分点。此外, 本文尝试引入原有

YoloV4 模型的 SPP (Spatial pyramid pooling) 结构用于特征维度的对齐, 发现其并不能较好提升模型的表现能力。本文推测可能是 SPP 层可能会破坏原有的全局感知特征, 导致造成信息的丢失和特征不对齐等问题, 这与纯卷积结构的作用效果并不完全一致。

表 2 消融实验结果

改进策略	1	2	3	4	5	6
Mobile Vit	✓	✓	✓	✓	✓	✓
数据增广	—	✓	✓	✓	✓	✓
特征融合层	—	—	✓	—	✓	✓
特征金字塔	—	—	—	✓	✓	✓
SPP	—	—	—	—	—	✓
精度/%	88.7	90.1	92.8	93.3	95.3	94.7

2.5 对比实验结果与分析

为了验证本文方法整体的有效性, 在相同实验环境下, 训练和测试了 Faster R-CNN、SSD、YoloV3、YoloV4、MV2-YoloV4 等模型, 与本文设计的模型进行了精度、速度和参数量进行全方位对比。实验结果如表 3 所示。

表 3 模型精度、速度对比

方法	尺寸	精度/%	速度/fps	参数量/M
Faster R-CNN	600	91.5	5	65.3
SSD	300	89.5	43	25
YoloV3	320	90.3	45	59
YoloV4	320	92.7	48	61.5
MV2-YoloV4	320	91.1	79	10.4
MViT-YoloV4	256	95.3	50.6	11.8

如表 3 所示, 本文方法为 MViT-YoloV4, 在精度上远高于 Faster R-CNN、SSD、YoloV3 等方法, 同时比原有的 YoloV4 模型提升了 2.6 个百分点, 相比 MV2-YoloV4 (MobileNetV2-YoloV4) 模型提升 4.2 个百分点。同时, 在模型速度对比中, 本文方法仅低于 MV2-YoloV4, 高于对比的其他方法, 说明轻量化模型比基线模型上存在着一定的速度优势。此外, 本实验发现推理速度与参数量之间的关系并不存在着对应的近似线性的关系。YoloV4 模型参数量比 YoloV3 要多, 但是其推理速度反而要略快些, MV2-YoloV4 参数量是 YoloV4 的 1/6, 但速度却快了不到 1 倍。

从实验结果表明本文的方法在电网变压器油液渗漏检测任务中检测精度和速度均取得了较好的效果。相对于 YoloV4 模型, 本文方法在电网变压器油液渗漏检测精度由 92.7% 提高到 95.3%。此外, 在边缘设备上进行测试对比发现, MViT-Yolo V4 在 Atlas NPU 的基础上其推理速度可以达到 43FPS。图 11 为电网变压器油液渗漏实际检测效果图。

3 结束语

如何快速并准确发现设备异常是保障电网安全与稳定



图 11 测试效果示例图

的前提之一。针对电网变压器油液渗漏检测问题,本文提出一种改进 YoloV4 的电网变压器油液渗漏检测方法。该方法使用 Mobile-ViT 作为模型的骨干结构,利用卷积和 Transformer 有效地提取目标的局部和全局信息特征,降低了计算量。同时本文在改进 YoloV4 模型的基础上,针对性地提出了一种多尺度特征融合方法,实现了局部和全局信息的多尺度特征融合,增强了上下文语义表达。本文方法有效地增强了检测模型的泛化能力和适用性,提高了电网变压器油液渗漏准确率和速度。但是 Transformer 结构的计算复杂度依旧相对较高,下一步将继续在提高模型检测准确率和速度的基础上,研究进一步降低模型的计算量和误检率的可能性方案。

参考文献:

- [1] 孔祥玉, 赵帅, 贾宏杰, 等. 智能电网中电力设备及其技术发展分析 [J]. 电力系统及其自动化学报, 2012, 24 (2): 21-26.
- [2] 蔡定国, 唐金权. 干式变压器用绝缘材料, 绝缘结构与系统综述 [J]. 绝缘材料, 2019, 52 (11): 1-8.
- [3] 饶帆. 油浸式变压器的绝缘材料与应用 [J]. 新型工业化, 2020, 10 (1): 34-37.
- [4] 王一凡, 张水龙, 赖泽彬, 等. 核电站变压器漏油原因分析与缺陷处理 [J]. 设备管理与维修, 2021, 499 (10): 31-32.
- [5] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C] //2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 1: 886-893.
- [6] 高纪东, 王正群, 夏进. 基于 HOG 特征的加权稀疏表示人脸识别算法 [J]. 计算机技术与发展, 2022, 32 (7): 64-69.
- [7] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 2197-2206.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C] //Computer Vision-ECCV 2016; 14th European Conference, Springer International Publishing, 2016:

21-37.

- [9] 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, 2016: 779-788.
- [10] 白钰杰, 裴以建, 朱秀军. 基于改进 YoloV4 模型的乳腺钼靶图像肿块检测 [J]. 计算机工程与科学, 2023, 45 (4): 654-664.
- [11] FARHADI A, REDMON J. Yolov3: An incremental improvement [C] //Computer Vision and Pattern Recognition, 2018, 1804: 1-6.
- [12] 彭道刚, 黄孟俊, 威尔江, 等. 基于 YOLOv4 的火电厂指针仪表识别方法研究 [J]. 计算机应用与软件, 2023, 40 (6): 166-172.
- [13] 杜少聪, 张红钢, 王小敏. 基于改进 YOLOv5 的钢轨表面缺陷检测 [J]. 北京交通大学学报, 2023, 47 (2): 129-136.
- [14] 邢伟寅, 李礁, 钟乐海, 等. 改进 Retina Net 的农作物病虫害检测模型研究 [J]. 四川农业大学学报, 2023, 41 (1): 153-157.
- [15] EN S, HE K, GIRSHICK R, et al. Faster r-CNN: Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems, 2015: 1440-1448.
- [16] 高玮玮, 杨亦乐, 方宇, 等. 多特征尺度融合改进 Faster-RCNN 视网膜微动脉瘤自动检测算法 [J]. 光子学报, 2023, 52 (4): 220-231.
- [17] HE K, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C] //2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2980-2988.
- [18] 倪汉杰, 初秀民, 张彬鹏, 等. 基于改进 Cascade-RCNN 的内河航标检测算法 [J]. 中国航海, 2022, 45 (3): 99-105.
- [19] 王程远, 赵俊莉, 计晓斐, 等. 基于改进的 VGG 网络的烟柜状态识别模型 [J]. 计算机工程与设计, 2023, 44 (6): 1796-1803.
- [20] 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.
- [21] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.
- [22] 祁宣豪, 智敏. 图像处理中注意力机制综述 [J/OL]. 计算机科学与探索: 1-20 [2023-07-24]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20230629.1447.002.html>.
- [23] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale [C] //International Conference on Learning Representations (ICLR), 2021: 57-69.
- [24] 付忠广, 王诗云, 高玉才, 等. 基于 Mobile-VIT 的旋转机械故障诊断方法 [J]. 汽轮机技术, 2023, 65 (2): 119-121.