

DOI:10.19651/j.cnki.emt.2212249

## 基于 Transformer 改进 YOLOv5 的山火检测方法研究\*

钱承山<sup>1</sup> 沈有为<sup>1</sup> 孙宁<sup>2</sup> 戴仁天<sup>1</sup>

(1.南京信息工程大学自动化学院 南京 211800; 2.无锡学院 无锡 214000)

**摘要:** 构建智慧生态林业中核心环节为对森林火灾的监测及防范,为了第一时间扑灭火源防止火势蔓延并将可能发生的山火隐患于第一时间消除,提出了两种适用于无人机高空巡检的森林火灾检测模型 YOLO\_MC 与 YOLO\_MCLite。其中 YOLO\_MC 可对标准图像中的明火及烟雾进行检测,并基于 YOLO\_MC 模型进行轻量化设计,提出适用于热图像中高温区域的检测模型 YOLO\_MCLite。在网络结构的设计中,首先在常规的卷积神经网络中融合加入了 Transformer 模型,提升了主干网络对于全局特征信息的感知能力;同时对 Transformer 模型进行轻量化设计,首先在网络结构上通过分组计算的形式减少 tokens 数量以降低计算量,其次通过通道注意力机制对特征块的通道数进行去冗余并提权来减少 tokens 的维度参数以降低计算复杂度,并且采用蒸馏算法从所设计的网络中提取出超轻量化网络应用于无人机红外影像的森林高温点检测,以预防森林火灾的发生。经过实验得出以下数据:所设计的两个检测模型中,其中适用于标准图像中对于明火及烟雾的检测准确率可达 94.8%,适用于热图像对高温点的检测准确率可达 97.2%,并且在英伟达 JETSON TX2 嵌入式设备上测试两个模型的帧率分别达到了 22.5 以及 32.4。实验结果表明本文所设计网络能够对森林火灾进行有效检测并可以通过检测高温点及时预防火灾发生。

**关键词:** 目标检测;森林火灾;神经网络;Transformer 模型;蒸馏学习

**中图分类号:** TP391.4 **文献标识码:** A **国家标准学科分类代码:** 510.1050

## Research on improved YOLOv5 forest fire detection method based on Transformer

Qian Chengshan<sup>1</sup> Shen Youwei<sup>1</sup> Sun Ning<sup>2</sup> Dai Rentian<sup>1</sup>

(1. College of Automation, Nanjing University of Information Science &amp; Technology, Nanjing 211800, China;

2. Wuxi University, Wuxi 214000, China)

**Abstract:** The core link in the construction of smart ecological forestry is the monitoring and prevention of forest fires. In order to put out the fire source to prevent the spread of fire and eliminate the hidden danger of forest fires in the first time, two forest fire detection models, YOLO\_MC and YOLO\_MCLite, are proposed for UAV aerial inspection. Among them, YOLO\_MC can detect open flame and smoke in standard images. Based on the lightweight of YOLO\_MC model, a detection model YOLO\_MCLite suitable for high temperature region in thermal images is proposed. In the design of the network structure, the Transformer model is firstly integrated into the conventional convolutional neural network, which improves the perception ability of the backbone network for global feature information; at the same time, lightweight design of the Transformer model is performed. First, reduce the number of tokens in the form of group computation on the network structure to reduce the amount of tokens. Second, de-redundant and weighted the number of channels in feature blocks through channel attention mechanism to reduce the dimension parameters of tokens to reduce the computational complexity. And the distillation algorithm is also used to extract the ultra-lightweight network from the designed network and apply it to the detection of forest high temperature points in the infrared image of the UAV to prevent the occurrence of forest fires. After experiments, the following data are obtained: the detection accuracy of the designed detection model for open flame and smoke can reach 94.8%, and the detection accuracy for high temperature points can reach 97.2%. And the test frame rate on the NVIDIA JETSON TX2 embedded device reached 22.5 and 32.4 respectively. The experimental results show that the network designed in this paper can effectively detect forest fires and prevent fires in time by detecting high temperature points.

**Keywords:** target detection; forest fire; neural network; Transformer model; distillation learning

## 0 引言

全球气候变暖除了导致冰川融化海平面上升等危害

外,也使得全球进入了森林火灾的高发期。据统计在 2021 年我国共发生 616 起森林火灾,虽然发生次数处于历史最低位,但也对生态环境在一定程度上造成了不可逆的影响。

收稿日期:2022-11-30

\* 基金项目:无锡市现代产业发展资金项目(20201012)、江苏省研究生科研实践创新计划项目(SJCX22\_0349)、江苏省科技副总项目(FZ20200099)资助

近年来退耕还林、封山育林等一系列政策的实施使得我国林场资源愈加丰富,而及时有效地进行森林火灾的监测和预警也显得更加重要。国家林业局所发部的《全国森林防火规划(2016-2025年)》,提出了新时期应积极更新森林防火治理体系,持续推进森林林场防火系统的现代化治理能力。而在传统森林防火措施中,仍主要依靠人工监测,在工作效率及监测覆盖面上存在有较大的短板,无法实现森林防火资源信息的共享及全方位的监控。基于上述问题本文拟设计一种针对于森林火灾遥感检测网络,以配合相关硬件设备实现智能化林场管理。

人工智能的发展促进了各领域智能化的构建,越来越多的计算机视觉技术应用于安防监测、遥感监测等工作中,在本文所述的应用场景同样也有类似的解决方案。如栗俊杰等<sup>[1]</sup>提出了一种基于YOLOv2-Tiny的无人机火灾检测与云台跟踪方法,首先对YOLOv2-Tiny的网络参数进行优化以得到网络最优模型,并将该模型部署与K210开发板上,当检测到火灾情况时可通过画面信息实时控制云台实现对火灾现场的画面跟踪;李宁等<sup>[2]</sup>设计了一种由无人机搭载和云台相机组成的图像采集平台,通过采集影像回传地面进行视觉处理实时识别监测火灾的发生,在火灾识别方面该团队提出了一种基于YOLOv3算法优化的improved-YOLOv3,相比于原网络加强了网络对特征的融合度,从而增加了网络的检测能力。在提升火灾烟雾的检测精度方面,汪睿卿等<sup>[3]</sup>提出一种融合细节特征与混合注意力机制的YOLOv4改进算法,该团队设计了细节特征融合模块,将主干网络中的底层细节特征引入高层特征图,得到具有丰富多尺度信息的融合特征,并且在通道和空间维度上采用混合注意力机制对融合特征的图权重进行重新赋值,在增强烟雾目标特征的同时抑制无关区域特征,使烟雾特征表达具有更好的鲁棒性。以上3个研究案例证明了将计算机视觉应用于火灾烟雾检测任务中的可实施性,并且均采用了YOLO算法作为检测网络,YOLO算法单阶段的检测流程使得其在检测速度上具有较大优势更适用于该类型的检测任务。但是,在整体设计中仍存在一定欠缺,轻量化的网络结构设计在嵌入性上较好,但检测精度不够理想;高精度的检测模型则仅适用于本地PC端检测运行,嵌入式设备算力无法满足实时运算。例如,王斌等<sup>[4]</sup>以YOLOv4-tiny作为主干网络提出了一种面向火焰快速检测的轻量化深度网络,将网络中的卷积层使用深度可分离卷积进行替换以减少参数,但为了弥补精度的损失,其团队在FPN模块中加入了通道注意力机制;李欣健等<sup>[5]</sup>同样采用了深度可分离卷积减少YOLO网络的参数,其团队则采用了数据增强及位置损失函数优化以提高网络检测精度,基于此提出了一种适用于复杂场景下的轻量化火焰方法。以上案例说明目前在嵌入式设备的轻量化网络部署中,主要采用了减少网络参数的方式,但是网络参数的锐减同样会对网络检测精度造成影响,因此需要采用其他方式提升

精度避免参数减少过多造成精度断崖式下降的结果。

综上,为更好地构建智慧林业,本文通过设计轻量化的网络结构使其能够搭载与嵌入式设备中实现离线检测,在这一环节中主要通过对支路模块的参数控制及采用蒸馏学习的方式提取轻量化模型结构;并通过优化网络模块的方式在原主干网络中加入Transformer模块增强网络对于特征的提取能力,在保证轻量化的前提条件下提升模型检测精度,提升网络整体鲁棒性。

## 1 相关理论

为提升网络的检测性能,本文在主干网络中加入了Transformer模型,以提升网络对于全局特征信息的提取。Transformer模型最初应用于自然语言领域,其采用“编码器-解码器”及自注意力机制构建了非循环的网络结构,解决了循环网络无法并行计算,无法计算长特征序列等问题。不同于LSTM、GRU等模型对于特征序列的计算需要按序列顺序输入,Transformer模型建立序列特征中各个特征的联系则通过并行计算的形式,这也使得Transformer能够应用于计算机视觉领域中<sup>[6-7]</sup>。对于在计算机视觉领域中特征图输入到Transformer模型中的计算流程如下所述:首先需要将二维的图像数据进行序列化操作,以 $H \times W \times C$ 大小的特征图是 $x \in R^{H \times W \times C}$ 为例,通过分割的方式将特征图划分为 $N(N = HW/P^2)$ 个大小为 $P \times P \times C$ 的特征块,并将每个特征块延展为一维向量 $token$ ,变换后输入的特征图可表示为 $x_p \in R^{N \times (P^2 \times C)}$ 。其次将序列化后的特征即tokens输入到Transformer模块的编码器当中,对于编码器的组成结构如图1所示。

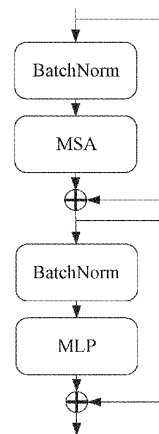


图1 Transformer模型编码器组成结构

在编码器内主要有两个层结构:多头自注意力机制层(multi-head self-attention, MSA)与多层感知机层(multi-layer perceptron, MLP),MSA层负责构建全局特征信息,而MLP层负责对信息进行分类,同时在MSA与MLP层前均配有层归一化层,归一化层与每个关键层组成残差连接。而对于Transformer模型的核心多头自注意力机制的

计算表达如下:多头自注意力机制由多个自注意力机制合并组成,每个自注意力机制通过计算  $Q$  (Query)、 $K$  (Key)、 $V$  (Value) 3 个向量实现全局信息的动态聚合。假设输入的特征序列为  $x_p$ , 通过线性映射矩阵  $U_{QKV}$  将  $x_p$  进行投影得到  $Q$ 、 $K$ 、 $V$  这 3 个向量,并计算  $Q$  与  $K$  之间的相似度  $S$ , 并将相似度  $S$  与  $V$  进行加权计算得到单个 token 序列的自注意力值:

$$SA = softmax\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (1)$$

式中:  $d$  为每个特征序列的维度。最后将每个特征序列的自注意力机制进行合并,即可得到多头自注意力机制:

$$MSA = [SA_1(z); SA_2(z); \dots; SA_n(z)]U_{msa} \quad (2)$$

式中:  $U_{msa}$  为映射矩阵,负责将每个自注意力头进行拼接聚合得到最终的多头自注意力值。

目前对于 Transformer 模型应用于计算机视觉领域中有如下 2 个关键问题需要解决,1) 如何降低 Transformer 的计算参数,由于 Transformer 的计算量与特征图大小、维度、划分特征块个数呈正相关,特征图越大、维度越多、划分的特征块越密集均会增加 Transformer 模块的计算量<sup>[8]</sup>; 2) 如何提高特征块之间的交互性,在标准 Transformer 模型中对特征图采用等分的形式划分特征块,特征块与特征块之间无关联,全具特征的关联性较弱<sup>[9]</sup>。因此在本文的 Transformer 模块设计中将着重解决以上两个问题以提升 Transformer 模块的应用性能<sup>[10-11]</sup>。

## 2 网络设计

### 2.1 主体网络设计

本文所述检测任务为无人机高空巡航对林场、森林等树木植被茂密地区的烟雾、火灾进行检测。在该检测任务中主要存在有如下难点,1) 由于植被遮挡的原因,背景干扰较强,使得漏检情况较多; 2) 当巡航面积较大时,巡航高度较高,相应目标在成像中的像素区域也较小,不易检测; 3) 由于检测模型运行环境为嵌入式设备,因此为保证嵌入式设备的算力能够应对,对于网络参数量同样具有一定的要求。综上在网络的设计环节中,由于 Transformer 模型的高参数量使得搭建全 Transformer 网络并不适用于本文任务中,因此本文选用了卷积神经网络中具有高检测速度的 YOLO 网络作为基础网络<sup>[12-15]</sup>, 依靠卷积网络对于局部特征的敏感性增强对于火焰、烟雾特征细节的提取能力。并设计新的轻量化 Transformer 模块,利用 Transformer 对全局特征的敏感性提升网络对于灾情区域的定位监测,将 Transformer 模块与其主干卷积网络进行合并,设计构建新特征提取网络,二者相结合后网络结构如图 2 所示。

如图 2 所示,在 YOLO 主干网络中共包含有 4 个跨级局部网络 (cross stage partial network, CSP) 模块,在每个 CSP 模块所对应的特征图维度中增加了 Transformer 模块支路,由原先的单支路重构为并行支路。CSP 模块其整体

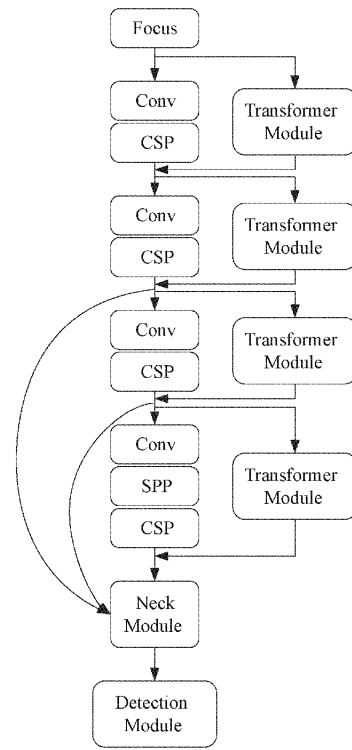


图 2 特征提取网络结构

为残差结构,本文所采用的 CSP 模块结构如图 3 所示。

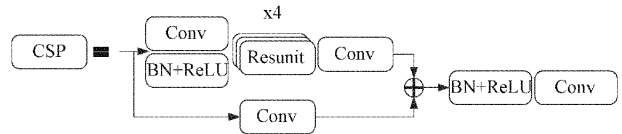


图 3 CSP 模块结构

在每个 CSP 模块内同样具有两个并行的支路,在上方支路中特征图首先由  $3 \times 3$  卷积层及 ReLU 激活函数进行激活,之后由 4 个 Resnet 残差模块进行特征提取,对提取后的特征再由  $3 \times 3$  卷积层进行特征整合。而在下方支路中则仅由一个  $3 \times 3$  卷积层对输入特征进行映射与上方支路的输出特征图进行合并,对合并后的特征同样采用 ReLU 激活函数进行激活并由  $1 \times 1$  卷积层调整特征图的通道数。在主干网络中,主要通过 CSP 模块对目标的局部特征进行提取,并通过多次下采样的方式缩小特征图使得在卷积核感受野不变的情况下提取不同粒度的特征信息。其次,在本文设计的主干网络中,在 CSP 模块进行局部特征信息提取的同时,通过加入 Transformer 模块强化网络对于全局特征信息的提取,即每个 CSP 模块均对应一个 Transformer 模块,采用 concat 层对两个模块的输出进行合并以提升特征图的丰富度。在整体结构中卷积模块与 Transformer 模块构成并行的结构,在每个维度中两个模块的输出会先进行合并再传递至下一维度中。并且考虑到网络的深度因素,不同深度的网络层具有不同级别的特征

信息,在本文所设计的网络中同样将浅层、中层模块输出的特征传递至网络深层进行3个粒度特征信息的合并,以进一步提升特征提取网络所输出特征图的特征信息完整性。

在融合操作中,考虑到将标准Transformer模块移植进网络后,由于Transformer本身的计算量较大,加入后对网络模型的实时性具有较大的影响,同时其对于特征图的特征块划分机制具有一定的局限性(各个特征块的关联性不强)。因此本文在对主干网络进行优化的同时,对于Transformer模块的结构及计算机制进行重构,以提升其计算速度及整体性能。首先在Transformer模块的结构设计上,本文将并行多头自注意力计算拆解为不同的小组进行并行计算,每组设置了不同大小的特征图分块参数,对分块得出不同大小的特征块组进行长距离特征构建,以提升特征丰富度。基于上述优化方案,本文所设计的Transformer\_MC(MultiChannel)结构如图4所示。

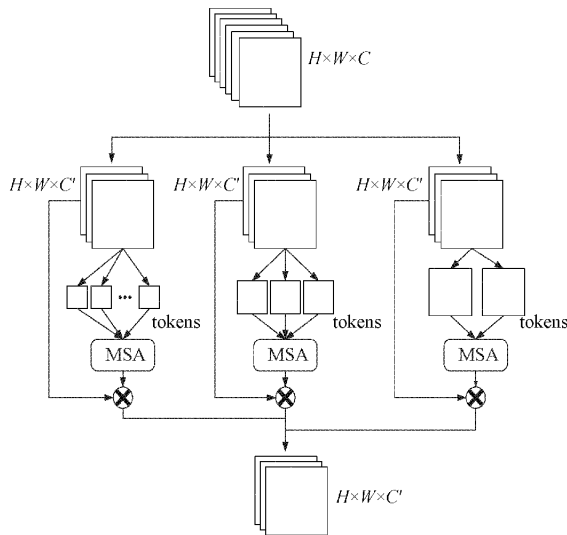


图4 Transformer\_MC模块结构

如图4所示,优化后的Transformer\_MC模块工作流程为:输入的特征图首先根据通道数采用交叉分组的形式将特征图均分为3个子特征图,交叉分组的形式确保了分组后3个子特征图仍具有特征关联,避免了特征断层的问题;其次在对特征图进行块划分时,对于特征块的大小参数 $p$ ,每组子特征图均设定了不同大小,使得在后续的多头自注意力计算及特征融合中,能够涵盖不同大小的特征块,即在同一个特征图中划分得到的特征块具有;并且 $p$ 参数设置得越大,所划分得到的特征块数越少,在一定程度上可减少多头自注意力的计算参数,之后对3个子特征图分别进行多头自注意力计算,并将计算得到的具有长距离关联特征信息的特征图进行合并,以得到信息关联度更强的全局特征信息。除了通过控制特征图划分特征块的大小来决定token的数量以减少计算量外,还可通过减小每个特征块的通道数来控制token的参数量以减少多头自注意力的计算量。因此,本文在对输入特征图采用交叉分组后得到

的3个子特征图进行二次处理。由于每个子特征图所要分割的特征块大小不同,即对感知区域大小不同,因此在分组后本文分别对3个子特征图采用空间注意力机制提取出对应于每个分块大小 $p$ 其所对应的重点区域。同时,通过采用多个通道注意力机制对特征图的通道进行加权重组,以其中一个子特征图为例,该过程如图5所示。

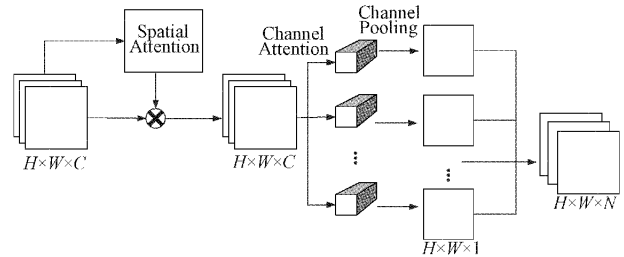


图5 支路轻量化

假设输入的特征图为 $x \in R^{H \times W \times C}$ ,首先采用空间注意力机制对该特征图中的重点区域进行提取:

$$y' = SA(x) \quad (3)$$

式中: $SA$ 表示为Spatial Attention空间注意力机制。其次设定了 $N$ 个通道注意力机制对输出 $y'$ 进行通道注意力计算:

$$y_i = CA_i(x) \quad (4)$$

式中: $y_i$ 表示每个通道注意力机制的输出, $CA_i$ 对应不同的通道注意力权重。最后将 $y_i$ 的维度变换为 $H \times W \times 1$ ,并将变换后的 $y_i$ 进行拼接即可得到参数轻量化后的子特征图,再对各子特征图划分不同大小的特征块进行多头自注意力计算。综上,将基于Transformer\_MC模块优化的YOLO网络命名为YOLO\_MC以进行区分。YOLO\_MC网络中对于每个维度的特征提取流程如图6所示。

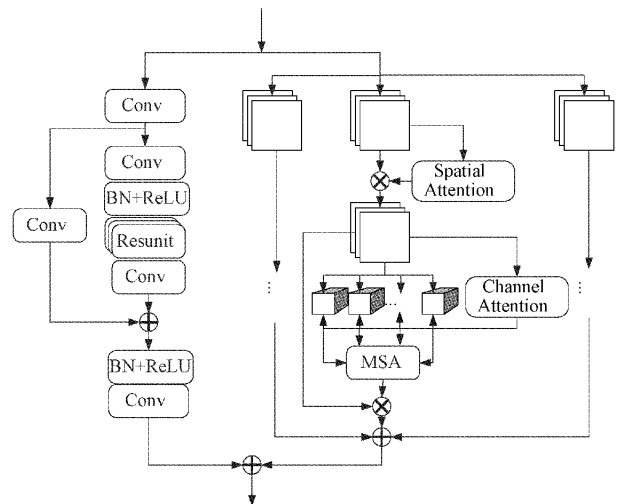


图6 特征提取详细流程

## 2.2 热图像检测网络设计

结合在实际应用场景中,当火势比较小时且受树木的遮挡,火光及烟雾存在有较大的漏检可能。针对于这种情

况,本文在摄像头模组中增加热成像仪,通过采集热图像检测图片中的高温区域,筛选出可疑区域以做出警报。由于热成像仪是通过探测热量,将其转换为对应的电信号进而转换为热图像,因此高温区域与低温区域在图像中有较大的色差与对比度,可通过对高亮区域进行检测以得出,整体检测任务较简单。同时在热图像中的目标与常规图像中目标形态相似,即二者具有在低级形态特征中具有一定的共性。因此对于热图像的检测网络,本节将 2.1 节中常规图像的检测模型结合热图像数据集迁移至新的 YOLO\_MC 网络;其次,由于热图像中的目标具有明显的显著性特征,检测难度较低,可以使用更精简的网络进行检测,故在获得 YOLO\_MC 热图像检测模型后,本文采用蒸馏学习的方式进一步提取轻量化网络模型<sup>[16 18]</sup>。

本文所采用的蒸馏方式为特征蒸馏,针对于教师网络中主干网络的层输出特征进行学习,在本文中教师网络采用的为 YOLO\_MC,而学生网络则将教师网络中 CSPDarkNet 网络部分的层数及参数进行 1/2 缩减,使得整体网络更浅更窄。蒸馏过程如图 7 所示。

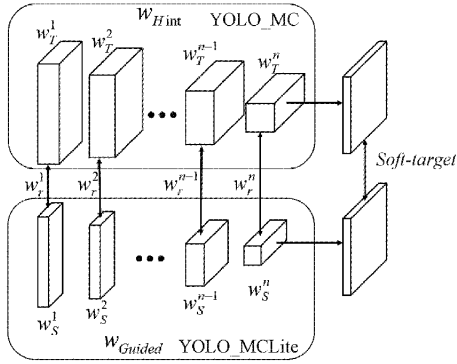


图 7 蒸馏学习示意图

在蒸馏过程中主要有两个阶段,第 1 阶段为学习教师网络的隐藏层特征,即对特征信息进行蒸馏。在这一阶段中,对与教师网络的层结构输出可表示为  $W_{Hint}$ , 学生网络的层结构输出表示为  $W_{Guided}$ , 由于在训练的初始阶段,学生网络的网络参数处于随机状态,因此需要通过一个映射函数将  $W_{Guided}$  与  $W_{Hint}$  进行匹配,该函数用  $W_r$  进行表示,并且学生网络的输出与教师网络输出的差异越小越好,即代表学习到相应特征。在该阶段中采用 MSE 差异作为损失函数:

$$L_{HT}(W_{Guided}, W_r) = \frac{1}{2} \| u_h(x; W_{Hint}) - r(v_g) \|^2 \quad (5)$$

式中:  $u_h$  表示教师网络第 1 层~第  $N$  层所对应的函数,  $v_g$  表示学生网络从第 1 层~第  $n$  层所对应的函数,而  $r$  表示为卷积层调整器。其次在第 2 阶段中,与标准蒸馏网络相同,采用 Soft-target 对教师网络进行学习,采用的损失函数为:

$$L_{KD}(W_S) = \mathcal{L}(y_{true}, P_s) + \lambda \mathcal{L}(P_T^r, P_S^r) \quad (6)$$

式中:  $\mathcal{L}$  为交叉熵函数,  $\lambda$  为可调节参数,  $W_S$  为学生网络整体的参数,而其他符号所代表的含义分别为:

$$P_s = \text{softmax}(a_s) \quad (7)$$

$$P_T = \text{softmax}(a_T) \quad (8)$$

$$P_T^r = \text{softmax}\left(\frac{a_T}{\tau}\right) \quad (9)$$

$$P_S^r = \text{softmax}\left(\frac{a_S}{\tau}\right) \quad (10)$$

式中:  $a_T, a_S$  分别代表教师网络和学生网络的 softmax 输出,  $\tau$  为 softmax 函数的缩放因子。综上,采用迁移学习与蒸馏学习的方式获取了对于热图像的检测网络,本文命名为 YOLO\_MCLite 以进行区分。

### 3 实验分析

在本章中采用以下评估指标对 YOLO\_MC 及 YOLO\_MCLite 网络的检测性能进行评估,其中精确率 (Precision) 评估模型对目标类别分类的准确性:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

召回率 (Recall) 评估模型对测试集中目标的查全能力:

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

其中,  $TP$  为将正类预测为正类的个数;  $FP$  为将负类预测为正类的个数;  $FN$  为将正类预测为负类的个数。

F1-Score 为精确率与召回率的调和平均数,反应模型综合能力:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

平均检测精度 (AP) 评估模型对每一个类别的检测准确率:

$$AP = \int_0^1 P_{smooth}(r) dr \quad (14)$$

平均检测精度均值 mAP 评估模型的综合检测准确率:

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (15)$$

其次,在模型轻量化方面,本文采用以下指标进行评估,网络参数 (Parameter) 评估模型空间复杂度:

$$Parameter = (C_{in} \times K^2 + 1)C_{out} \quad (16)$$

式(16)为单层卷积层参数量计算公式,  $C_{in}$ 、 $C_{out}$  分别为特征图输入前与输入后的通道数;  $K$  为卷积核的大小。帧率 FPS: 计算模型每秒检测帧数(主观指标)对于本次实验所采用的数据集,出于安全角度考虑本文采用了开源数据集,在标准图像的检测中采用了北亚利桑那大学的 FLAME 数据集、比尔肯特大学的 VISFire 数据集及 CVPR 数据库中的 KMU Fire and Smoke 数据集;在热图像的检测测试中采用的为 FLAME 数据集。对以上数据集进行数

据清洗及重新标注得到本文所使用的数据集参数如表1所示。

在标准图像中同时包含有火焰及烟雾的图像有7 526张,仅包含有火焰的图像有5 999张,仅包含烟雾的图像有1 728张。在实验中对于数据集的划分,将所有类型的图片均按照7:2:1的比例划分为训练集、测试集与验证集。

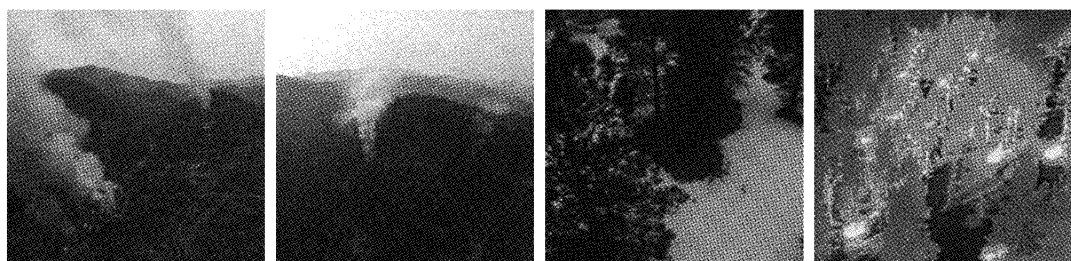


图8 数据集样例

数据集中的样本数据涵盖了不同飞行高度所拍摄到的森林火灾场景,使得训练得出的模型能够适用于不同检测环境。其次在本次实验中所采用的硬件环境与软件环境分别如表2所示。

表2 硬件 & 软件环境配置

硬件	软件环境
CPU	Inter Core i7_13700k
GPU	NVIDIA RTX 3070
嵌入式设备	JETSON TX2
操作系统	Win10
编程语言	Python 3.8
深度学习框架	Pytorch 1.8.0
CUDA	11.3
cuDNN	8.2.0

对于网络训练时的超参数设置,本文设置了不同的超参数组,对每组设置均经过多次训练,对测试结果取平均值,以最优组的结果为本次实验中的超参数设置组。在超参数的设置中,本文设置迭代次数为300个epoch,迭代每次输入的数据样本张数batchsize为16,初始学习率设置为0.001,在第200个epoch后学习衰减为0.0001直至训练完成。

### 3.1 消融实验

首先采用消融实验法对YOLO\_MC的检测性能进行对比测试,在消融实验中本文以YOLOv5s作为基础算法,分别在YOLOv5s中加入了标准Transformer模块,以及在标准Transformer模块的输入端采用本文所设计的输入支路轻量化操作与采用多路多头计算方式的Transformer模块进行Transformer模块的优化性能对比,最后一个为加入本文所设计的Transformer\_MC模块,以验证在整体结构上的差异与Transformer模块的使用差异,对比结果

对于所采用的数据集样例如图8所示。

表1 数据集详情

数据集	火焰	烟雾	综合视频时长/min
标准图像	13 525	9 254	37
热图像	7 451(高温区域)		19

如表3所示。

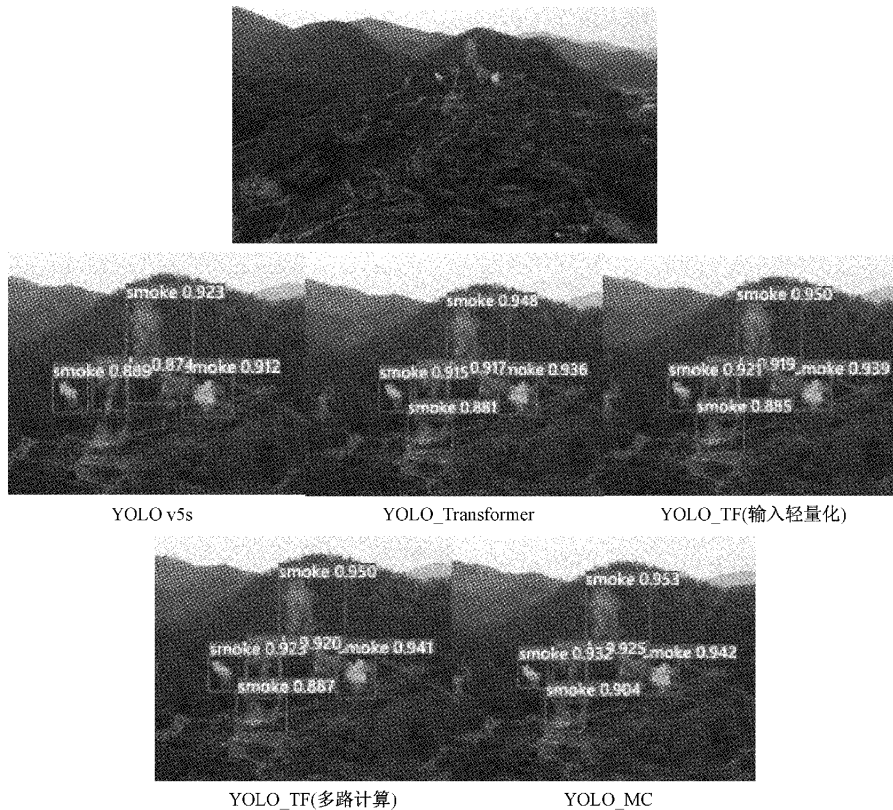
表3 模型检测性能消融对比结果 %

模型	Precision	Recall	F1-Score	mAP
YOLO v5s	90.1	94.5	92.2	91.4
YOLO_Transformer	92.8	96.7	94.7	93.5
YOLO_TF(输入轻量化)	93.1	96.8	94.9	93.8
YOLO_TF(多路计算)	93.3	97.0	95.1	94.1
YOLO_MC	93.8	97.2	95.4	94.8

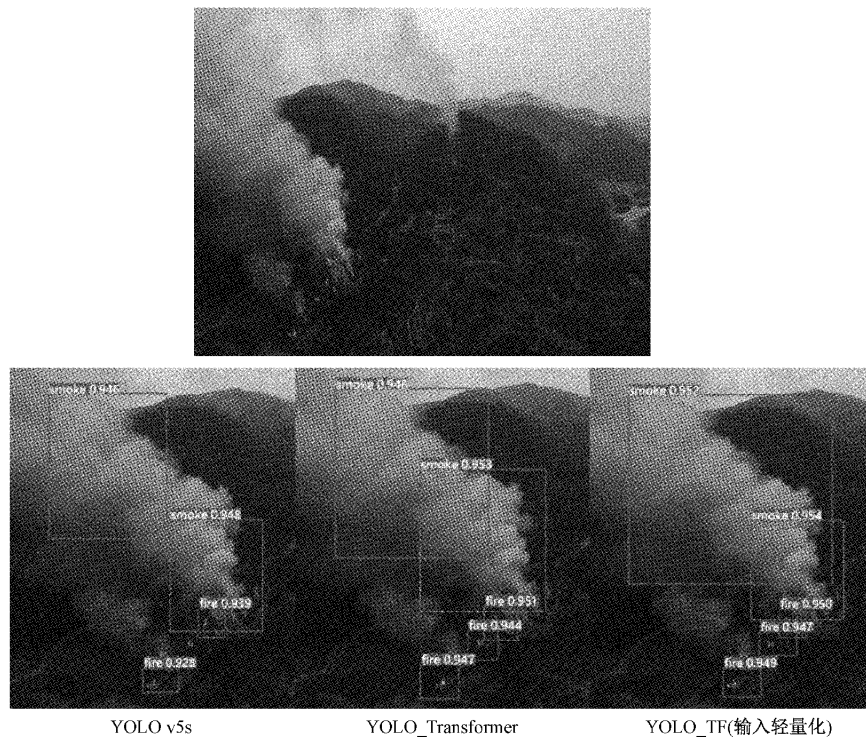
从表3的对比数据中,在Precision和Recall指标中YOLO\_Transformer与YOLO\_MC相比于YOLOv5s均有所提高,说明在纯卷积模型的网络结构中加入Transformer模型能够提升网络的分类能力与查全能力,特别是对于Recall指标,加入Transformer模型后相比于之前有了较大幅度的提升,说明Transformer模型的多头自注意力构建全局特征信息,提升了网络对目标的搜索能力。其次,通过将Transformer\_MC模块分解为两部分移植入标注Transformer模块中以验证各模块对于网络改进的有效性。首先将Transformer模块的输入端做支路轻量化处理,在该轻量化的过程中分别采用了空间注意力机制与通道注意力机制对关键区域进行提取,在轻量化的同时也完成了对于冗余特征的处理,仅对加权重点特征信息进行多头自注意力计算,从表3的测试结果也体现出了对有效特征进行提权后对于各项指标性能有一定的增加。再者将Transformer模块中的多头自注意力计算通过对卷积通道分组拆解为多路进行,而每一路中对于特征块的划分均采用了不同的大小,使得在特征块与特征块之间存在有重叠交集,最终输出的特征具有更强的关联性,在一定程度上也提升了网络整体的检测性能。最后,从YOLO\_Transformer网络与YOLO\_MC网络的对比中,采用Transformer\_MC模块相比于标准Transformer模块在

Precision 指标与 Recall 指标上要更优,证明了 Transformer\_MC 模块采用分组计算的形式提升了 tokens 的交互性,并且在分组前采用空间注意力机制对关键区域进行标记,同样提升了模型性能,使得网络的总体性能要优于采用标准 Transformer 模块。从 F1-Score 与 mAP 指

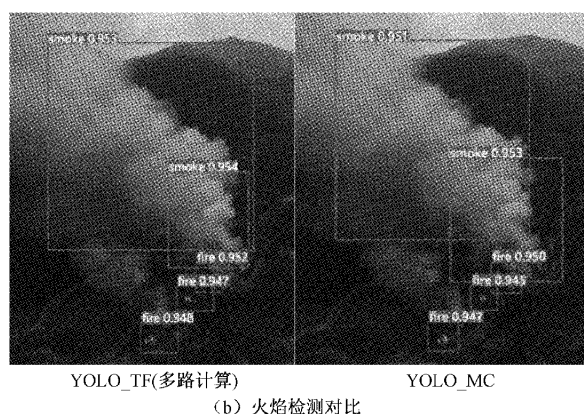
标的对比中,同样验证了在检测性能方面,采用卷积模型与 Transformer 模型相结合的方式要优于纯卷积模型、本文所设计的 Transformer\_MC 模块在该任务中要优于标准 Transformer 模块。对于各模型的实际检测效果对比如图 9 所示。



(a) 烟雾检测对比



(b) 火灾检测对比



(b) 火焰检测对比

图9 实际检测效果对比

从图9的对比情况中,可以看出在特征提取网络中加入Transformer模型后能够提升网络对于一些小目标的检测能力,并且在检测精度上也有明显提升。将标准Transformer分别采用支路轻量化设计与多路多头自注意力计算后,对于目标的检测精度与标准Transformer相比有小幅提升。而对于YOLO\_MC与YOLO\_Transformer的比较中,二者的差异主要体现在检测精度,YOLO\_MC的精度更高于YOLO\_Transformer。而对于网络的轻量化消融实验,测试结果如表4所示。

表4 模型轻量化消融对比结果

模型	Parameter	FPS
YOLO v5s	7 022 326	28.1
YOLO_Transformer	18 523 778	12.1
YOLO_TF(输入轻量化)	14 258 425	18.6
YOLO_TF(多路计算)	15 692 177	17.3
YOLO_MC	9 147 654	22.5

注:FPS测试硬件环境为英伟达JETSON TX2

在轻量化消融实验中,YOLO\_Transformer网络计算量相比于YOLOv5s锐增,主要原因在于标准Transformer模块自身参数量较大,加入后使得网络整体的参数量增加,并且在检测帧率上也有明显的下降。而对于YOLO\_MC网络,其所采用的Transformer\_MC模块采用空间注意力机制生成通道数更少的子特征图,并且通过增大特征图的划分大小以减少tokens的生成数量,这两项优化的具体效果如上表所示。采用Transformer\_MC模块的网络整体的参数量相比较与YOLOv5s仅有小幅增加,且检测帧率仍能保持20 fps以上。

综合表3与4的测试结果,可以证明本文所设计网络YOLO\_MC在保证轻量化的同时在本文所述任务中仍具有出色的检测精度,可满足实际应用场景。

### 3.2 蒸馏模型对比分析

通过3.1节实验验证了本文所设计YOLO\_MC网络

在标准可见光成像中的火焰及烟雾检测有优秀的表现。而对于热图像输入的检测中,本文基于YOLO\_MC网络模型蒸馏提取了YOLO\_MCLite网络,对于该网络模型在热图像中的检测性能如表5所示。

表5 蒸馏模型检测性能测试结果 %

模型	Precision	Recall	F1-Score	mAP
YOLO v5s	93.1	95.2	94.1	94.6
YOLO_MC	97.8	98.6	98.2	97.3
YOLO_MCLite	97.4	98.4	97.8	97.2

与表3中YOLOv5s、YOLO\_MC的检测结果相比,在热图像检测中YOLOv5s与YOLO\_MC的各项检测性能指标均要高于表3,说明热图像中目标更易检测。分析YOLO\_MCLite的测试结果,其各项指标均高于YOLOv5s且与YOLO\_MC相接近,说明采用蒸馏学习后YOLO\_MCLite的网络性能仍与教师网络YOLO\_MC相同。各模型对于高温区域的检测情况如图10所示。

由图10可见,在检测性能方面的差异与标准图像相同,采用Transformer模型后降低了模型的漏检率;并且经过蒸馏学习后的模型YOLO\_MCLite在检测精度上仍与YOLO\_MC相接近无较大差距。其次对轻量化性能进行测试分析,测试结果如表6所示。

在轻量化测试中,由于YOLO\_MCLite将原CSPDarkNet网络中的部分层结构及网络层参数进行删减,使得网络计算量及检测帧率都要小于YOLO v5s。结合表5测试结果,可证明在热图像的检测中,经YOLO\_MC蒸馏得到的YOLO\_MCLite在该场景下的检测任务中,在保证有高精度的同时具有更少的计算量,网络性能相比较于其他网络更优。

### 3.3 综合对比

本文所设计网络结构主要侧重于网络的轻量化,以保证能够在嵌入式设备中运行,为此在进一步实验中,本文参照相关山火检测方法,以轻量化为主要参考依据,选用



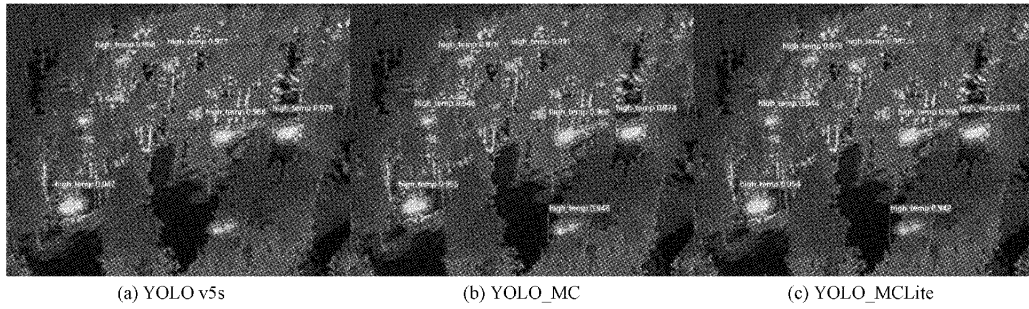


图 10 实际检测效果对比

表 6 蒸馏模型轻量化对比测试

模型	Parameter	FPS
YOLO v5s	7 022 326	28.1
YOLO_MC	9 147 654	22.5
YOLO_MCLite	4 528 784	32.4

在山火检测任务中常用的轻量化主干网络搭配 YOLO 检测器和其他常用网络<sup>[19-25]</sup>,从综合检测性能及轻量化两个维度与本文所设计网络 YOLO\_MC 及 YOLO\_MCLite 进行对比。首先在标准图像中火焰及烟雾对比检测结果如表 7 所示。

其次在热图像中高温区域对比检测结果如表 8 所示。

表 7 标准成像下多模型检测结果对比

模型	F1-Score/%	mAP/%	Parameter	FPS
YOLO_MC	95.4	94.8	9 147 654	22.5
YOLO_ShuffleNetv2	89.6	87.2	3 558 714	35.7
YOLO_MobileNetV3	90.3	89.5	5 217 475	30.1
YOLOP	93.7	92.6	16 557 524	15.1
YOLOX	94.1	92.7	18 757 524	14.8
YOLOv7	94.7	93.5	19 957 524	13.5
Faster RCNN	94.2	93.4	45 581 474	3.8
SSD	93.8	91.2	24 987 524	6.4

表 8 热图像下多模型检测结果对比

模型	F1-Score/%	mAP/%	Parameter	FPS
YOLO_MCLite	97.8	97.2	4 528 784	32.4
YOLO_ShuffleNetv2	92.3	89.0	3 558 714	35.7
YOLO_MobileNetV3	94.1	91.8	5 217 475	30.1
YOLOP	95.2	93.9	16 557 524	15.1
YOLOX	95.6	94.5	18 757 524	14.8
YOLOv7	96.7	95.2	19 957 524	13.5
Faster RCNN	96.4	95.8	45 581 474	3.8
SSD	95.6	93.2	24 987 524	6.4

由表 7 与 8 的测试结果当中, YOLO\_MC 虽然在轻量化程度上不及 YOLO\_ShuffleNetv2 与 YOLO\_MobileNetV3,但其检测帧率在 JETSON TX2 上仍能满足实际实时检测,并且其检测性能评估指标 F1-Score 与 mAP 要远高于这二者。同时与 YOLOP、YOLOX、YOLOv7、Faster RCNN、SSD 网络相比, YOLO\_MC 则是在检测精度与轻量化上都要优于以上网络,在本文所述任务中的应用性更强;而 YOLO\_MCLite 则在网络参数上与

YOLO\_ShuffleNetv2、YOLO\_MobileNetV3 处于同一层次,并且检测精度要优于二者。且与 YOLOP 等其他网络相比,在检测速度与检测精度上均有明显优势。

其次将本文所设计网络 YOLO\_MC 与现有的森林火灾检测算法进行比较,验证本文所提出网络模型在森林火灾检测任务中所具有的优越性与先进性。该对比实验中各模型源码通过开源下载,在本文所诉数据集上进行测试对比。首先针对于标准图像的火灾、烟雾检测对比,对比

结果如表9所示。

表9 标准成像下现有方法对比

模型	F1-Score/ %	mAP/ %	Parameterer	FPS
YOLO_MC	95.4	94.8	9 147 654	22.5
文献[26]	91.7	89.4	5 475 214	28.6
文献[27]	93.9	93.5	69 862 265	0.5
文献[28]	92.2	92.8	13 462 713	17.2

文献[26]基于YOLOv5s进行优化将YOLOv5s的主干网络替换为轻量化网络Shufflenetv2,其次在主干网络与Neck模块的连接处加入为轻量化网络设计的位置注意力模块(coordinate attention,CA),它能够将图片不同的位置信息聚合到通道中,使所被检测的对象关注度得以提高。这一改进方法得利于主干网络的轻量化使得整体检测速度相对由于本文YOLO\_MC模型,但在检测精度上与本文所设计模型差距较大;文献[27]采用纯Transformer模块搭建检测网络,该网在Swin Transformer网络结构中对窗口自注意力机制进行改进,采用了KNN自注意力提高对小块噪声的识别,使用Augmentation数据增强方法增加模型的泛化能力。与本文相比同样采用了Transformer模型,但对于模型特征则缺少了由卷积模型所提取的局部特征,因此由于特征完整性与本文相比较弱使得在检测性能方面略低于本文网络,其次是由于Transformer模型的参数量较大,使得整体模型不具备有轻量化的特性。文献[28]提出基于改进GhostNet-FCOS的火灾检测算法,该算法以目标检测网络FCOS为基础降低通道维数并选用GhostNet作为特征提取网络,以实现轻量化火灾检测算法,同时引入动态卷积,在不增加网络宽度和深度的情况下优化主干网络的基础模块,提高对形态多变的火焰图像的特征提取能力。该网络与本文网络相比其主要差距仍然为特征数据的完整性,该网络主要依靠动态卷积对火灾区域的特征信息进行重点提取,缺乏全局特征,因此在遥感图像上对于火焰及烟雾的感知能力并不及本文网络。其次针对于红外成像的检测模型对比结果如表10所示。

表10 热图像下现有方法对比

模型	F1-Score/ %	mAP/ %	Parameterer	FPS
YOLO_MCLite	97.8	97.2	4 528 784	32.4
文献[29]	94.3	95.6	6 875 147	27.6

文献[29]采用CSPNet34作为主干网络进行测试,相比于YOLO结构其减少了多粒度检测,以单个检测器进行检测以提升速度。与本文YOLO\_MCLite模型相比其在

检测精度与检测速度上均较差,其主要原因本文网络模型通过蒸馏学习获得,在保证网络精简的同时网络参数仍具有较高有效性。综合表9与10的对比实验,可说明本文所设计网络模型在森林火灾检测中的优越性。

## 4 结 论

本文所提出的YOLO\_MC网络基于CSPDarkNet网络加入了Transformer模型,并设计了轻量化的Transformer模块以应用在主干网络中,以实现兼顾轻量化与高检测精度的网络模型,应用于森林、林场等大范围的高空无人机火灾巡检。并且采用蒸馏学习的方法,从YOLO\_MC网络中提取出网络参数更少,轻量化程度更高的YOLO\_MCLite网络,应用于热图像中高温区域的检测,以实现森林火灾的预警。通过实验结果表明,本文所设计的Transformer\_MC模块有效的控制了参数量,融合至CSPDarkNet网络中,在不大幅增加参数量的情况下,有效提升了检测精度,且能满足JETSON TX2嵌入式设备的搭载使用。本文所设计网络为构建智慧林业提供了可行方案。

## 参考文献

- [1] 栗俊杰,毛鹏军,淡文慧,等. 基于YOLOv2-Tiny的无人机火灾检测与云台跟踪研究[J]. 消防科学与技术,2022,41(1):108-112.
- [2] 李宁,刘青,熊俊,等. 基于无人机的火灾检测系统设计与实现[J]. 火灾科学,2022,31(1):46-51.
- [3] 汪睿卿,王慧琴,王可. 融合细节特征与混合注意力机制的火灾烟雾检测[J]. 液晶与显示,2022,37(7):900-912.
- [4] 王斌,李靖,赵康,等. 面向火焰快速检测的轻量化深度网络研究[J]. 计算机工程与应用,2022,58(17):256-262.
- [5] 李欣健,张大胜,孙利雷,等. 复杂场景下基于CNN的轻量火焰检测方法[J]. 模式识别与人工智能,2021,34(5):415-422.
- [6] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30.
- [7] ZHU X, SU W, LU L, et al. Deformable DETR: Deformable transformers for end-to-end object detection [J]. ArXiv Preprint, 2020, ArXiv:2010.04159.
- [8] MEHTA S, GHAZVININEJAD M, IYER S, et al. Delight: Deep and light-weight Transformer [J]. ArXiv Preprint,2020, ArXiv:2008.00623.
- [9] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International

- Conference on Computer Vision, 2021: 10012-10022.
- [10] 丁志江,李丹,马志程,等. 基于 Transformer 的车道线分割算法研究[J]. 电子测量与仪器学报,2022, 36(10):227-234.
- [11] 朱相华,智敏,殷雁君. 基于 2D CNN 和 Transformer 的人体动作识别[J]. 电子测量技术,2022, 45(15): 123-129.
- [12] 史梦圆,高俊钊. 改进 YOLOv5 的建筑物破损检测算法研究[J]. 国外电子测量技术,2022,41(8):84-90.
- [13] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [ J ]. ArXiv Preprint, 2020, ArXiv: 2004.10934.
- [14] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [ J ]. ArXiv Preprint, 2021, ArXiv:2107.08430.
- [15] 彭继慎,孙礼鑫,王凯,等. 基于模型压缩的 ED-YOLO 电力巡检无人机避障目标检测算法[J]. 仪器仪表学报,2021,42(10):161-170.
- [16] TANG R, LU Y, LIU L, et al. Distilling task-specific knowledge from bert into simple neural networks [ J ]. ArXiv Preprint, 2019, ArXiv: 1903.12136.
- [17] AMIK F R, TASIN A I, AHMED S, et al. Dynamic rectification knowledge distillation [ J ]. ArXiv Preprint,2022, ArXiv:2201.11319.
- [18] KUSHAWAHA R K, KUMAR S, BANERJEE B, et al. Distilling spikes: Knowledge distillation in spiking neural networks[C]. 2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 2021: 4536-4543.
- [19] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient cnn architecture design[C]. Proceedings of the European Conference on Computer Vision(ECCV), 2018: 116-131.
- [20] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3 [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.
- [21] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [ J ]. ArXiv Preprint, 2021, ArXiv:2107.08430.
- [22] WU D, LIAO M W, ZHANG W T, et al. YOLOp: You only look once for panoptic driving perception[J]. Machine Intelligence Research, 2022, 19(6): 550-562.
- [23] 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 [ J ]. ArXiv Preprint,2022, ArXiv:2207.02696.
- [24] REN 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, 28.
- [25] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector [ C ]. European Conference on Computer Vision. Springer, Cham, 2016: 21-37.
- [26] 皮骏,刘宇恒,李久昊. 基于 YOLOv5s 的轻量化森林火灾检测算法研究[J]. 图学学报,2023,44(1):26-32.
- [27] 叶铭亮,周慧英,李建军. 基于改进 Swin Transformer 的森林火灾检测算法[J]. 中南林业科技大学学报, 2022,42(8):101-110.
- [28] 张融,张为. 基于改进 GhostNet-FCOS 的火灾检测算法[J]. 浙江大学学报(工学版), 2022, 56(10): 1891-1899.
- [29] 戴颖成,陈知明,刘峰,等. 基于无人机红外影像的森林火灾燃烧点检测方法[J]. 中南林业科技大学学报, 2022,42(9):102-114.

### 作者简介

钱承山,教授,硕士生导师,主要研究方向为智能终端与物联网应用、非线性系统控制、自动检测技术。

E-mail: qianchengshan@163.com

沈有为(通信作者),硕士研究生,主要研究方向为图像处理、深度学习、嵌入式开发。

E-mail:1219545797@qq.com