

基于遥感图像的 YOLOv3 模型算法优化

刘德健¹ 郭荣佐^{1*} 宫禄齐²

¹(四川师范大学 四川 成都 610101)

²(中国科学院计算技术研究所 北京 100190)

摘要 为了使 YOLOv3 算法能在遥感设备上实时地进行目标检测,模型压缩是常用的解决方案。根据模型压缩后精度会有损失的特点,先为模型添加注意力机制算法再对模型进行训练,并以此为基础提出基于融合卷积层与 BatchNormal 层后对模型进行通道剪枝方法,对 YOLOv3 进行通道剪枝,得到压缩后的 YOLOv3 目标检测模型,并对剪枝后的模型进行微调后,恢复模型的精确度。通过实验得到的结果,融合后剪枝的方法在 mAP 损失 0.6% 的基础上,对 YOLOv3 的模型大小压缩 94.93%,检测速度提升了 150.6%;通过实验证明,该模型能够运用到对于实时性较高和检测精度较高的遥感图像目标检测,并且适用于存储空间较小的遥感设备。

关键词 遥感图像 深度学习 注意力机制 卷积融合 模型剪枝

中图分类号 TP3 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2025.01.034

YOLOV3 FUSION PRUNING BASED ON REMOTE SENSING IMAGE

Liu Dejian¹ Guo Rongzuo^{1*} Gong Luqi²

¹(Sichuan Normal University, Chengdu 610101, Sichuan, China)

²(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract In order to enable the YOLOv3 algorithm to perform real-time target detection on remote sensing equipment, model compression is a common solution. According to the characteristics of the loss of accuracy after the model is compressed, this paper added the attention mechanism algorithm to the model and trained the model. On this basis, we proposed a channel pruning method based on the fusion of the convolutional layer and the BatchNormal layer. Channel pruning was performed on YOLOv3, and we obtained the compressed YOLOv3 target detection model. After fine-tuning the pruned model, the accuracy of the model was restored. The experimental results show that the method of pruning after fusion in this paper reduces the size of the YOLOv3 model by 94.93% and increases the detection speed by 150.6% with only 0.6% mAP loss. The experiment proves that the model can be applied to real-time performance remote sensing image target detection with higher and higher detection accuracy, and suitable for remote sensing equipment with small storage space.

Keywords Remote sensing image Deep learning Attention mechanism Convolution fusion Model pruning

0 引言

近几十年来,由于我国航空航天技术的不断发展,对遥感图像的研究也成为热门课题。随着遥感卫星的不断部署与发展,遥感图像也呈高分辨率趋势,所以遥感图像在林业、农业、工业等有了广泛的应用。但

是要取得遥感图像精确的目标检测结果,仍有诸多困难,所以,构建一个合适的目标检测模型来迎接这些困难是重要的一步。对于遥感图像的研究,以往的研究主要是利用计算机对遥感图像中各类地物的光谱信息和空间信息进行分析,选择特征,并用一定的手段将特征空间划分为互不重叠的子空间,然后将图像中的各个像元划归到各个子空间对应的地物类别去,这属于

收稿日期:2021-10-26。国家自然科学基金项目(11905153,61701331)。刘德健,硕士生,主研领域:模型部署,物联网技术。
郭荣佐,教授。宫禄齐,硕士生。

遥感图像的分类^[1],在这方面已经取得了重大的进展。然而,随着科技的进步,分类的要求远远不能满足任务的需求,目前对于遥感图像的任务要求更进一步,即需要识别图像内的具体内容,遥感图像的目标检测也不断进步,成为重要一环。

与此同时,在目标检测领域,神经网络也在与时俱进,本次实验中用到的基于卷积神经网络的目标检测,也是神经网络中计算机视觉领域的重要技术。目前,基于卷积神经网络的目标检测算法主要分为两大类:双阶段目标检测(two-stage detection)和单阶段目标检测(one-stage detection)。双阶段目标检测以 R-CNN^[2]为基础不断演进的 Fast R-CNN^[3]、Faster R-CNN^[4]、Mask R-CNN^[5]等;单阶段目标检测算法以 YOLO(You Only Look Once)^[6]为基础不断演进的 YOLOv3^[7]、YOLOv4^[8]、YOLOv5、SSD^[9]等。由于双阶段目标检测算法有预选框的存在,使该类模型的精确度较高,但是模型的体积较大,导致实时性较差,对于实时性较高的遥感图像并不适用。目前大多数遥感图像嵌入式设备已适用于 YOLOv3 模型,但对于实时性较高的遥感图像不适用,然而在大量的目标检测算法中,YOLOv3 目标检测模型既有较高的检测精确度,又有相对较快的检测速度,最近效果更加出色的算法存在不支持的情况,并且如 SSD 这样的单阶段目标检测模型各方面效果并不如 YOLOv3 神经网络模型效果好,所以本文采用 YOLOv3 模型,来支持遥感设备的可用性。

遥感图像检测的多数任务中有实时性的需求,而神经网络模型的实时性与模型结构以及模型参数量(体积)密不可分,神经网络不断进步的同时,模型的优化、压缩、加速也在不断改进,模型压缩的方法主要分为剪枝^[10-11]、量化^[12-13]、低秩分解^[14-15]、知识蒸馏^[16-17]。其中,模型剪枝在各种各样的模型压缩加速中脱颖而出,在保证模型精度不丢失的情况下,更加有效地加快了模型推理速度。模型剪枝主要思想是通过剪掉模型中不重要的参数进行裁剪,以此来减少模型的参数量与模型的计算量。模型剪枝按照移除的粒度大小可以分为两种类型:非结构化剪枝^[18]和结构化剪枝^[19]。剪枝是保证模型精确度与模型推理速度的合理方法,以此为基础,对已有的神经网络模型进行调优,获得适合遥感图像的体积小、速度快的神经网络模型。不难看出,在模型压缩加速中剪枝是相对更有效方法,所以,采用模型剪枝的方法,对遥感图像进行检测的神经网络模型进行剪枝,以此来获得神经网络模型压缩加速的效果。

模型压缩可以使模型加速,从而提升模型的性能,但是由于目前遥感图像大多数为高空或者高速状态进

行成像,导致遥感图像成像的像素较低,在对遥感图像目标检测模型进行训练后,往往精度差强人意,为了追求神经网络模型更高的精确度,引入了注意力机制算法。用于机器翻译的注意力模型(AM)^[20]是最开始提出的注意力机制,现已被广泛应用于神经网络模型。注意力机制是模仿人的视觉机制,人类的视觉系统总是更加关注图像中辅助判断的部分信息,并且忽略掉不重要的信息。注意力机制就是想让系统学会注意力,能够忽略无关信息而关注重点信息。许多优异的注意力机制算法,很大程度上提升了模型的精确度,目前比较主流的注意力机制分为两大类,单路注意力机制:SE-Net(Squeeze and Excitation)^[21]、ECA-Net(Efficient Channel Attention)^[22]等;双路注意力机制:SK-Net(Selective Kernel Networks)^[23]、ResNeSt^[24]、CBAM(Convolutional Block Attention Module)^[25]等。随着注意力机制不断发展,在人工智能领域,注意力机制已成为神经网络结构的重要组成部分,并且广泛应用于自然语言处理、目标检测等神经网络模型中,为模型添加注意力机制,能够更好地提升模型的精度。

遥感图像的发展,预示着科技的进步,本文拟优化已有的剪枝方法对深度神经网络模型进行压缩加速,使模型能够适应当前不断发展的遥感技术。对模型的优化方法,多种多样,但是适用于遥感图像的方法需要在不断的实验中发现,本文在修改 YOLOv3 主干网络后,经过多次的实验选择了适合的注意力机制结构,以此来增加模型的精度,并且经过训练完后的模型对其融合 Conv 层与 BatchNormal 层达到进一步压缩模型的目的,融合后对其进行剪枝以达到适应遥感图像的需求的神经网络模型。

1 方法描述

随着遥感图像技术的不断发展,对遥感图像进行检测的神经网络模型也需要不断进步。遥感图像目标检测任务相对于普通图像目标检测而言,遥感图像目标检测任务需要具有轻量化、实时性、高精度等特点,然而大多数目标检测模型难以满足此需求。本次实验针对目标检测模型在遥感图像上检测中的不足,提出相应的改进方法,以达到遥感图像目标检测任务的要求。首先本次实验选取大多数遥感图像嵌入式设备已适用的 YOLOv3 模型对其进行改进。首先,根据遥感图像目标小、类别少、图片大的特点,改进 YOLOv3 的主干网络为 Resnet18 网络;其次,根据遥感图像检测任务高精度特点,并且模型压缩后精度会有损失的特性,提前在训练时添加注意力机制来提高模型的精度;然后,根据遥感图像检测任务实时性、轻量化的特点,提

出融合剪枝的方法对模型进行压缩;最后,经过对剪枝后模型的微调,恢复剪枝前的精度。

1.1 改进 YOLOv3 模型 Backbone

本次实验中,将 YOLOv3 的主干网络 Darknet-53 改进为 Resnet18^[26],如图 1 所示。首先,根据遥感图像检测任务的实时性高、类别少、图片大等特点,选择参数量少,模型层数少、计算量小、体积小的网络模型。与此同时,根据神经网络中不同尺度特征图重要性的分析,对 Resnet18 网络进行改进,来增强对小目标特征的提取能力。其次,由于遥感图像的目标检测大多数情况下是对小目标进行检测,如本次实验使用的数据集,是对遥感图像中飞机、汽车小目标进行检测,YOLOv3 中的主干网络 DarkNet53 对小目标的特征提取效果表现一般。本次将 Resnet18 残差网络作为 YOLOv3 的主干网络,一方面,因为 Resnet 在第一个卷积层采用 7×7 的卷积核,大大提高了模型的感受野,所以 Resnet 网络对小物体特征提取较为友好,另一方面,Resnet18 网络层数相比于 DarkNet53 网络层数更加少,体积更加小,参数量也更少,这样将大幅缩小 YOLOv3 的模型体积,所以选择 Resnet18 作为本文优化 YOLOv3 的主干网络。最后,结合遥感图像分辨率较低的特点,优化 YOLO 层 anchor boxes,相比于 YOLOv3-DarkNet53,该模型在遥感图像上的精确度不相上下,但是模型体积更小,参数量更少,网络层数也更少。

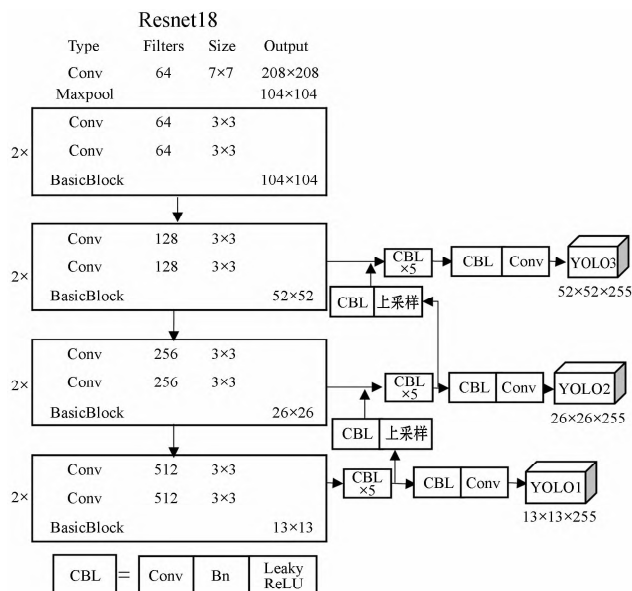


图 1 YOLOv3-RESNET18 结构

1.2 添加注意力机制

注意力机制的提出是提升神经网络模型精确度的重要一环,根据 Conv 与 BatchNormal 融合后的最小单元 RR(ResnetConv2-D + ReLU) 提取特征的特点,在训练时选取 CBAM 注意力机制网络(整体结构如图 2 所示),该方法用到注意力机制中的两个模块,分别是通

道域和空间域,通过通道注意力机制模块与空间注意力机制模块这两个模块得到前面卷积层细化后的特征信息。通过 CBAM 的两个模块,能够让神经网络更加有效地知道网络传播过程中哪些信息需要更加突出,并且让网络知道哪些信息需要抑制,从而根据该思想来提高神经网络模型的精确性。

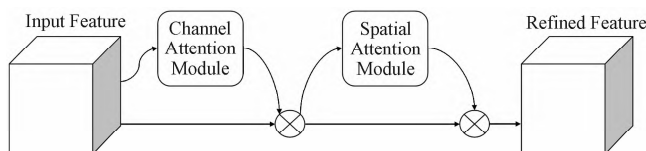


图 2 CBAM 注意力机制结构

根据 YOLOv3 模型的特点,同样改进 Resnet18 为对图像 3 次进行特征提取,分别是在 32 倍降采样、16 倍降采样、8 倍降采样,这 3 次采样形成了对特征图的多尺度提取。第一次添加注意力机制的位置在 32 倍降采样后,32 倍降采样的感受野最大,是对大目标的检测,所以在该处添加 CBAM 注意力机制模块;第二次添加注意力机制是在 16 倍上采样融合 16 倍降采样后,在融合后添加注意力机制的原因是网络越深的特征表达效果越好,比如在进行 16 倍降采样检测,如果直接使用 16 倍下采样的特征来检测,这样就使用了浅层特征,这样效果一般并不好。如果想使用 32 倍降采样后的特征,但深层特征的大小太小,因此使用步长为 2 的上采样,把 32 倍降采样得到的特征图的大小提升一倍,也就成了 16 倍降采样后的维度,这样就可以进行深层特征的检测;融合特征层的操作既可以提高非线性程度、增加泛化性能,提高网络精度,又能减少参数、提高实时性。同理,第三次注意力机制添加位置在 8 倍上采样融合 8 倍降采样后。添加 CBAM 注意力机制后的结构如图 3 所示。

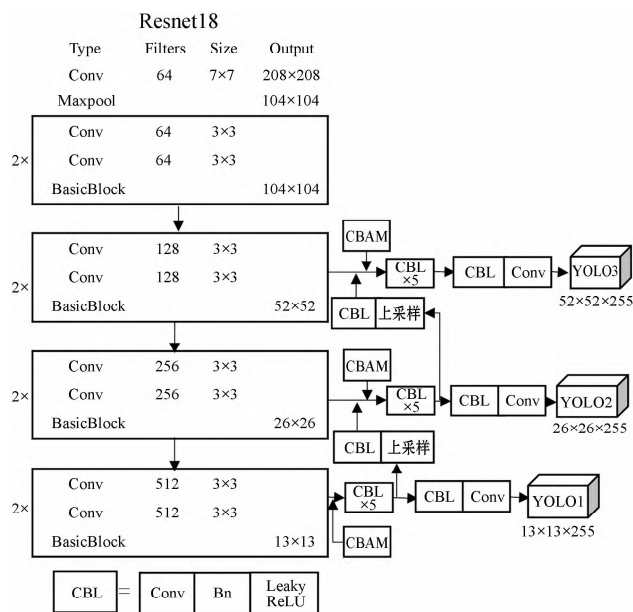


图 3 YOLOv3-RESNET18-CBAM 结构

在添加 CBAM 注意力机制时采取相应的消融实验,探究注意力机制模块位置对模型精度的影响,实验根据卷积神经网络提取特征的特点,选出注意力机制积极影响最好的位置。实验结果如图 4 所示,可以看出,CBAM 注意力机制对实验精度产生积极的影响,添加的位置不同实验精度也有所不同。

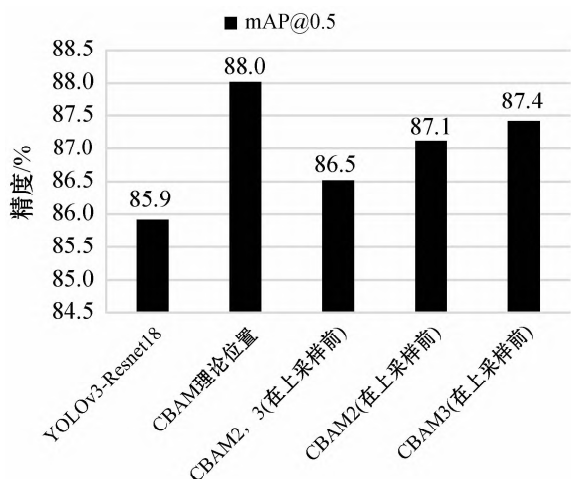


图 4 CBAM 位置对精度影响

1.3 卷积层与 BatchNormal 层的融合

一般放在卷积层之后的 BatchNormal^[27]层在卷积神经网络中,能够加速网络的收敛,还能够防止网络的过拟合问题。BatchNormal 层在为模型训练时起到了积极的作用,但是,模型训练完后进行前向推理时,由于增加了 BatchNormal 层的运算,不仅降低了模型的性能,还增加了模型的存储大小。当前,有很多先进的深度神经网络模型,如 ResNet、MobileNet^[28]、ShuffleNet^[29]等都添加了 BatchNormal 层来优化训练时的网络模型,因此,将 BatchNormal 层的参数通过数学推导来合并到卷积层,从而提升模型的前向推理速度,是优化模型一种有效的方式。

卷积层计算式如下:

$$Y = W \times X + B \quad (1)$$

BatchNormal 层计算式如下:

$$\mu = \frac{1}{m} \sum_{i=1}^m x_i \quad (2)$$

$$\delta^2 = \frac{1}{m} \sum_{i=1}^m (Y - \mu)^2 \quad (3)$$

$$x_i = \frac{Y - \mu}{\sqrt{\delta^2 + \epsilon}} \quad (4)$$

由式(2)、式(3)、式(4)可得式(5)。

$$y_i = \gamma \times x_i + \beta \quad (5)$$

通过式(1) - 式(5)可以进行推导:

$$Y = \gamma \times \left(\frac{W \times X + B}{\sqrt{\delta^2 + \epsilon}} \right) + \beta \quad (6)$$

进行变换可得融合后卷积计算式(7)。

$$Y = \frac{\gamma \times W}{\sqrt{\delta^2 + \epsilon}} \times X + \frac{\gamma \times (B - \mu)}{\sqrt{\delta^2 + \epsilon}} + \beta \quad (7)$$

即

$$Y = W_{\text{merged}} \times X + B_{\text{merged}} \quad (8)$$

通过式(8)即可以得到卷积层与 BatchNormal 层融合后的式子,利用代码根据该式将训练好的 YOLOv3-Resnet18 模型中的卷积层与 BatchNormal 层进行融合,融合阶段可在进行遥感图像推理阶段,也可以将融合后的模型进行保存,本文为压缩加速神经网络模型,将融合后的模型保存下来,融合后 YOLOv3-Resnet18 模型的层数大量减少。

1.4 融合后通道剪枝

传统的通道剪枝方法原理如图 5 所示,图 5 中假设第 i 层的 C_{i1} 和 C_{i3} 通道对应的权重值为 0.008 和 0.002,可以近乎等于 0,因此该层中的这两个通道在网络传播过程中的贡献可以近似看作不存在,可以丢掉。利用该思想,逐层地将模型中权重低于某个百分比的所有通道去除掉,得到一个更加紧凑的神经网络,为模型剪枝的主要方法。

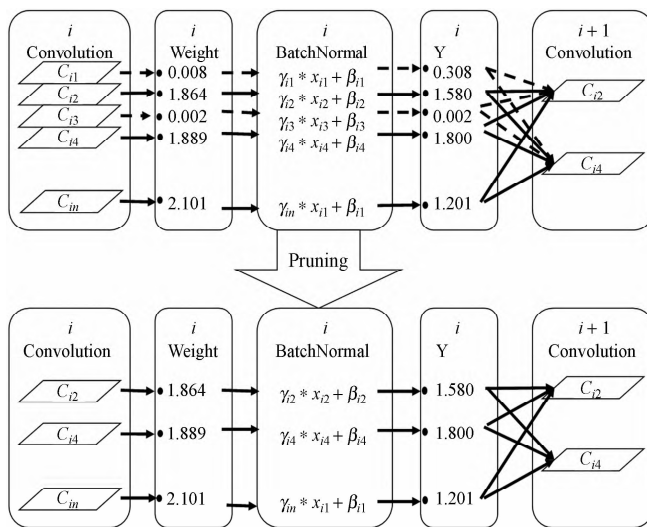


图 5 通道剪枝示意图

但是从图 5 中不难看出,在 C_{i1} 卷积核经过 BatchNormal 计算后,得到的 Y 值可能并不接近于 0,如果此时将该 C_{i1} 卷积核剪去,模型结果会受到一定程度的影响。如果模型层数多时,而且不经过 BatchNormal 层后进行剪枝,模型精度也许会在一定程度上损失,所以,以此为基础,根据式(7)得:

$$Y = \frac{\gamma \times W}{\sqrt{\delta^2 + \epsilon}} \times X + \frac{\gamma \times (B - \mu)}{\sqrt{\delta^2 + \epsilon}} + \beta$$

式中: γ 为 BatchNormal 的缩放因子; β 为平移参数; μ 和 δ^2 为均值和方差; ϵ 为作用参数,防止分母为 0; Y 是 BatchNormal 计算后的结果,由于 Y 的值需要进行 ReLU

激活函数,如果 Y 值很小时,经过激活函数后对网络的贡献度也很小,以此为基础,本文提出融合 Conv 与 BatchNormal 层后剪枝的方法,该方法根据融合后的式

$$(7) \text{ 参数 } W_{a_{\text{merged}}} = \frac{\gamma}{\sqrt{\delta^2 + \epsilon}} \text{ 与 } B_{b_{\text{merged}}} = \frac{\gamma \times (B - \mu)}{\sqrt{\delta^2 + \epsilon}} + \beta$$

在式子中的贡献度来对模型进行剪枝,可以得出:根据 $W_{a_{\text{merged}}}$ 来计算该卷积核贡献度,比 γ 参数更加精确,同时当参数 $W_{a_{\text{merged}}}$ 很小, $B_{b_{\text{merged}}}$ 很大时,此时的贡献度也会很大,所以根据这两个参数的贡献度来进行剪枝。

该方法不仅能在一定程度上提升模型的精确度,而且能够减少模型的网络层数,即减少模型的参数量与模型的体积,效果如图 6 所示。

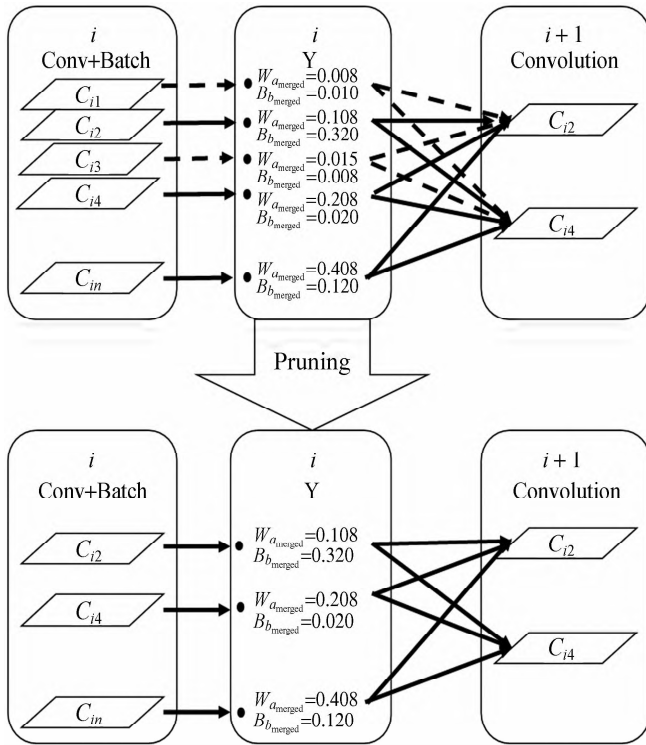


图 6 融合后通道剪枝示意图

特别注意, YOLOv3-Resnet18 的主干网络为残差网络,属于多分支网络, Resnet18 包含 4 个基本的 BasicBlock,每个 BasicBlock 分别对应于不同的尺度的特征,并且每个 BasicBlock 结构都是由 3×3 的尺度的卷积核组成。由于 Resnet18 属于残差网络,存在 shortcut,所以 BasicBlock 的输入通道数与其内部的卷积核个数必须保持相等,才能保证残差网络的加法运算,因此在进行通道剪枝时需要注意 3×3 的尺度的卷积核剪枝后的数量需要与 BasicBlock 的输入通道保持一致。

1.5 剪枝后模型微调

设 w_i 为未剪枝模型第 i 层卷积层的权重矩阵,其

卷积核个数为 n_i ; w'_i 为剪枝后模型第 i 层卷积层的权重矩阵,其卷积核个数为 n'_i 。第 i 层卷积层的剪枝率

$$p_i = \frac{n_i - n'_i}{n_i}, \text{ 当 } p_i \text{ 大于某阈值 } \epsilon \text{ 时,即判定为剪枝后模型}$$

保留原模型信息过少,对精度影响较大,需对这类卷积层进行信息优化。剪枝后模型精度相比于原模型精度存在一定程度上的损失,微调是模型剪枝后重要的一步,利用剪枝后的模型对数据集进行重新训练,使剪枝后模型中的卷积核提升对数据集的敏感程度。微调实现对剪枝后模型某些卷积层的信息补偿,进而在微调后获得更高精度。

1.6 方法总结

本文提出对 YOLOv3 模型的压缩加速优化主要步骤大体分为以下的四步:首先,由于 YOLOv3 中自带的 DarkNet53 主干网络,参数量大,模型层数多,导致模型体积大,进行推理时计算量大,并且 DarkNet53 没有较大的卷积核在小目标的特征提取方面效果在一定程度上不如 Resnet18,所以对 YOLOv3 的主干网络进行更换;其次,根据模型压缩后精度上会有损失的特点,再对 YOLOv3-Resnet18 模型进行训练时添加 CBAM 注意力机制模块,根据神经网络特征提取的特点,并经过多次的消融实验,总共为 YOLOv3-Resnet18 模型添加 3 个 CBAM 注意力机制模块;然后,在得到添加注意力机制后的模型后,对模型进行卷积层与 BatchNormal 层的融合,根据融合式子的推导,构建相关的算法后,对模型的卷积层与 BatchNormal 层进行融合;最后,对融合后的 YOLOv3-Resnet18 模型进行通道剪枝,通过融合式子中的贡献度对模型进行剪枝,移除融合卷积层中几乎无贡献度的卷积核,并对其进行微调,恢复到模型原有的精度。

2 实验

2.1 数据集

为验证本文方法的有效性,采用的数据集 UCAS-AOD, UCAS-AOD 数据集是一个遥感影像数据集,该数据集由中科大发布,初次发布于 2014 年,后于 2015 年补充,包含汽车(car)、飞机(plane)两类目标,以及背景负样本。该数据集中,图片大小为 1280×659 。样本具体信息如表 1 所示。本文对 YOLOv3、YOLOv3-SPP、YOLOv3-tiny、YOLOv4、YOLOv4-tiny 等代表性模型进行训练,与本文提出的算法模型进行对比以验证本文算法的效果。

表 1 UCAS-AOD 数据集的具体信息

类名	图像	目标样本
飞机	1 000	7 482
汽车	510	7 114

本文中所有模型训练的优化方法均采用随机梯度下降(SGD)算法,批训练数量(batch size)为64,每个模型都进行300个epoch的训练。训练中,初始学习率为0.001,学习率采用LambdaLR(学习率自定义调整)方法,学习率最后会衰减到 5×10^{-4} ,权重衰减值为 5×10^{-4} ,动量系数设置为0.937;在进行推理时,所有训练和优化后的模型均使用单卡 GeForce RTX 2080 Ti,图片进入网络前预处理,图片尺寸均调整为 416×416 ,每16幅图为一个Batch,进行推理。

2.2 模型融合剪枝自身对比

本文在UCAS-AOD数据集上对本文提出的添加CBAM注意力机制训练后,进行融合通道剪枝实验,对剪枝后的模型命名为YOLOv3-Resnet18-CBAMP并对其进行微调。经过多次实验后,通过分析本文方法生成模型的mAP、模型体积,以及模型的速度FPS(Frames Per Second)后,选取最优的剪枝率80%,对未进行添加CBAM机制的模型、融合后通道剪枝模型等结果进行对比如表2所示。未进行剪枝的YOLOv3-Resnet18模型,在mAP得到保证的情况下,模型体积大、参数量大、FPS低,基本满足遥感图像实时性的要求,本文提出的剪枝方法将优化YOLOv3-Resnet18模型,使其完全能适应遥感图像目标检测的要求;注意力机制能使模型在精确度上有一定的提升,在添加CBAM注意力机制后,模型的体积与参数量及FPS基本保持不变,模型的mAP有了0.021小幅度的提升;对添加注意力机制的模型进行融合剪枝后,相比于YOLOv3-Resnet18-CBAM,模型的体积大小下降81.81%,参数量下降了79.76%,计算量GFLOPS下降了59.68%,模型mAP损失了0.013,FPS增加了89.09%,显然已经达到了遥感图像实时性的要求;最后,通过对通道剪枝后的YOLOv3-R18-CBAMP模型进行微调,mAP基本能够恢复到添加CBAM后模型的mAP。

表 2 融合后通道剪枝对比

模型	Size/MB	Params	mAP@0.5	FPS
YOLOv3-R18	62.86	1.64×10^7	0.859	113
YOLOv3-R18-CBAM	62.87	1.64×10^7	0.880	110
YOLOv3-R18-CBAMP (Fused-pured)	11.98	3.32×10^6	0.867	208
YOLOv3-R18-CBAMP (fine-tuned)	11.98	3.32×10^6	0.876	208

并且实验过程中,对YOLOv3-Resnet18-CBAMP模型进行卷积与BatchNormal融合时,模型的层数相比于未融合模型层数减少40层,如图7所示。

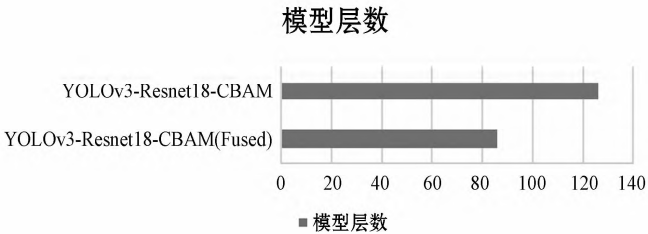


图 7 融合卷积与BatchNormal层YOLOv3-Resnet18-CBAM模型层数对比

2.3 融合剪枝对比

实验中比较了本文提出的剪枝方法、文献[30]和文献[31]的剪枝方法,图8为通道剪枝率的变化,剪枝模型mAP的变化曲线。可以看出添加注意力机制模块后,模型mAP的起点便已优于未添加注意力机制的模型;在剪枝率为60%之前,本文通道剪枝方法,与文献[30]和文献[31]的剪枝方法得到的mAP基本没有损失;然而当剪枝率大于60%时,文献[30]和文献[31]方法的mAP损失均大于本文提出的剪枝方法;当剪枝率在60%和90%之间时,本文的剪枝方法得到的模型mAP均优于文献[30]和文献[31]得到的模型mAP;因此本文的融合后通道剪枝方法优于文献[30]和文献[31]的通道剪枝方法。

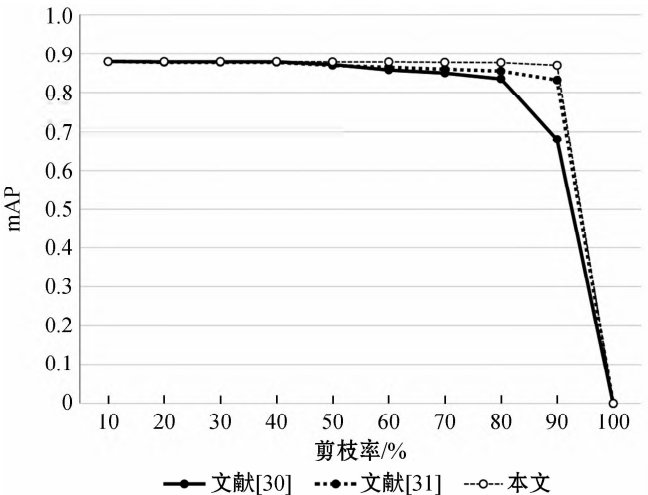


图 8 通道剪枝后模型mAP的变化曲线图

如表3所示,在实验过程中通过与文献[30]和文献[31]的剪枝方法进行对比,比较、分析文献[30]和文献[31]在UCAS-AOD遥感图像数据集上训练后不同剪枝率效果,在mAP损失可以接受的前提下,为文献[30]和文献[31]中设计的剪枝方法选取最优剪枝率;并对比本文设计的YOLOv3-Resnet18-CBAMP模型,由结果可以看出,本文提出添加注意力机制的融合剪枝后YOLOv3-Resnet18-CBAMP模型相比于未剪枝

前的模型,模型的体积大小下降 81.81%,参数量下降了 79.76%,mAP 增加了 0.017,FPS 增加了 84.07%。YOLOv3-R18-CBAMP 在模型的大小、模型的精度以及模型的检测速度这三项因素上均优于文献[30]和文献[31]的模型。

表 3 本文剪枝方法、文献[30]、文献[31]对 YOLOv3-Resnet18-CBAM 的剪枝方法比较

模型	Size/MB	Params	mAP@0.5	FPS
YOLOv3-R18	62.86	1.64×10^7	0.859	113
YOLOv3-R18 (75% pruning) ^[30]	19.24	4.59×10^6	0.849	151
YOLOv3-R18 (85% pruning) ^[31]	35.46	8.66×10^6	0.853	172
YOLOv3-R18-CBAMP (本文)	11.98	3.32×10^6	0.876	208

2.4 轻量化模型对比

本文在 UCAS-AOD 数据集上对各个 YOLOv3、YOLOv4 原模型,以及一些衍生的轻量级模型进行训练与对比。实验结果如表 4 所示。

表 4 原模型、轻量化模型、本文模型对比

模型	Size/MB	Params	mAP@0.5	FPS
YOLOv3	236.60	6.15×10^7	0.882	83
YOLOv3-SPP	35.30	8.85×10^6	0.875	145
YOLOv3-tiny	34.60	8.24×10^6	0.833	248
YOLOv4	245.8	6.97×10^7	0.898	71
YOLOv4-tiny	29.60	7.21×10^6	0.850	233
YOLOv3-R18-CBAMP (本文)	11.98	3.32×10^6	0.876	208

通过分析表 4 结果可知,通过训练不同的模型,在保证遥感图像所需准确率与实时性达到基本要求的前提下:YOLOv3、YOLOv4 原模型体积大、参数量大、计算量大,虽然 mAP 较高,但 FPS 太低,基本不能满足遥感图像实时性的特点,本文提出的 YOLOv3-Resnet18-CBAMP 模型相比于原模型在各方面都具有较大的优势;与 YOLOv3-SPP、YOLOv3-tiny、YOLOv4-tiny 等 YOLO 系列的轻量级模型相比,YOLOv3-SPP 在检测速度 FPS 上大大低于 YOLOv3-Resnet18-CBAMP,虽然 YOLOv3-Resnet18-CBAMP 检测速度 FPS 比 YOLOv3-tiny、YOLOv3-SPP 低,但是 YOLOv3-Resnet18-CBAMP 在模型体积上均比实验中 YOLO 系列的轻量级模型小,以及精度 mAP 比 YOLOv3-tiny 高 0.043,比 YOLOv4-tiny 高 0.026。本文提出的对 YOLOv3 的优化方案,在模型精度上保

证了有效性,在速度上满足遥感图像实时性的特点,并且平衡了模型的检测速度与精度,所以满足遥感图像精确性、高实时性的特点,并且能够适用于对精度、实时性要求高的遥感图像检测中。

由于遥感图像大多数为嵌入式,嵌入式设备计算力相对较小,而且存储空间也较小,根据该特点设计的 YOLOv3-Resnet18-CBAM 神经网络模型正是满足该需求,如图 9 所示,本文设计模型的轻巧性与精确性,以此来部署到遥感嵌入式设备上。

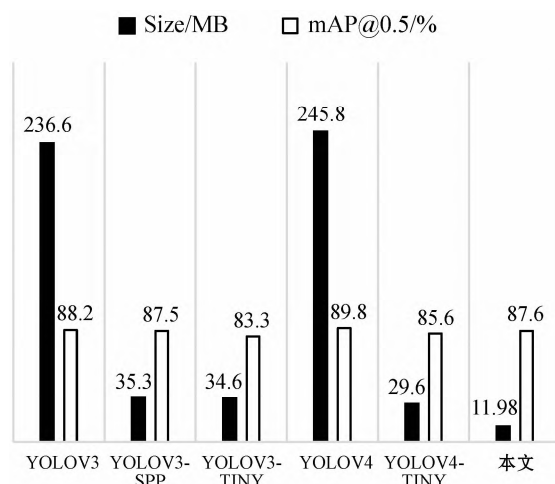


图 9 原模型、轻量化模型、本文模型对比

3 结 语

目前,硬件的不断进步,使深度学习能够更多地从理论应用于实践,遥感图像的发展也带动深度学习的不断进步。通过优化 YOLOv3 模型,对其修改 Backbone,本文在知晓压缩模型后有精度损失的前提下,先为模型添加注意力机制,使精度在本身的基础上提高,并进行融合 Conv 和 BatchNormal 层后进行剪枝,在中科大发布的遥感数据集 UCAS-AOD 上进行实验证明,本文提出的 YOLOv3-Resnet18-CBAMP 目标检测模型,虽然在大多数性能上优于 YOLOv3 原模型,但是在检测精度上与原模型有些许的差距;尽管在检测精确度上高于 YOLOv3-tiny 与 YOLOv4-tiny,但是 YOLOv3-tiny 与 YOLOv4-tiny 是通用模型,导致结构上与本文为 UCAS-AOD 数据集剪枝后得到的模型在结构上差异大,YOLOv3-Resnet18-CBAMP 在浮点计算上高于 YOLOv3-tiny,结构设计不如 YOLOv4-tiny,导致在推理速度上不及这两个模型。科技就是在不断的试错中发展的,本文选择注意力机制是在模型剪枝后达不到遥感图像精度高的需求时,经过多次的试错与探索,才找到适合 YOLOv3-Resnet18 模型的注意力机制模型。在以后的研究中,可以利用最新的 YOLOv5 方法来进一步优化

模型,还可以利用知识蒸馏的方法对剪枝后模型精度进一步提升。

参 考 文 献

- [1] Dou P, Shen H F, Li Z W, et al. Remote sensing image classification using deep-shallow learning[J]. Applied Earth Observations and Remote Sensing, 2021, 14: 3070 – 3083.
- [2] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580 – 587.
- [3] Li J N, Liang X D, Shen S M, et al. Scale-aware fast R-CNN for pedestrian detection[J]. IEEE Transactions on Multimedia, 2015, 20(4): 985 – 996.
- [4] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137 – 1149.
- [5] 杨智宏, 贺石中, 冯伟, 等. 基于 Mask R-CNN 网络的磨损颗粒智能识别与应用[J]. 摩擦学报, 2021, 41(1): 105 – 114.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2015: 779 – 788.
- [7] 张奔, 徐锋, 李晓婷, 等. 针对军事集群目标的 YOLOv3 改进算法研究[J]. 火力与指挥控制, 2021, 46(5): 81 – 85.
- [8] 张欣, 张永强, 何斌, 等. 基于 YOLOv4-tiny 的遥感图像飞机目标检测技术研究[J]. 光学技术, 2021, 47(3): 344 – 351.
- [9] 王良玮, 陈梅, 李晖, 等. DF-SSD: 一种基于反卷积和特征融合的单阶段小目标检测算法[J]. 计算机与现代化, 2021(6): 18 – 23.
- [10] LeCun Y, Denker J, Solla S. Optimal brain damage[J]. Advances in Neural Information Processing Systems, 1989(2): 598 – 605.
- [11] Li H, Kadav A, Durdanovic I, et al. Pruning filters for efficient convnets[EB]. arXiv:1608.08710v1, 2016.
- [12] Gong Y C, Liu L, Yang M, et al. Compressing deep convolutional networks using vector quantization[EB]. arXiv: 1412.6115, 2014.
- [13] Dettmers T. 8-Bit approximations for parallelism in deep learning[EB]. arXiv:1511.04561, 2015.
- [14] Lebedev V, Ganin Y, Rakhuba M, et al. Speeding-up convolutional neural networks using fine-tuned CP-Decomposition[EB]. arXiv:1412.6553, 2014.
- [15] Oseledets I V. Tensor-train decomposition[J]. SIAM Journal on Scientific Computing, 2011, 33(5): 2295 – 2317.
- [16] Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network[EB]. arXiv:1503.02531, 2015.
- [17] Zagoruyko S, Komodakis N. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer[EB]. arXiv:1612.03928, 2016.
- [18] 徐嘉荟. 基于模型剪枝的神经网络压缩技术研究[J]. 信息通信, 2019(12): 165 – 167.
- [19] Li H, Samet H, Kadav A, et al. Pruning filters for efficient conv-nets[EB]. arXiv:1608.08710v1, 2016.
- [20] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[C]//International Conference on Learning Representations, 2015: 1409 – 1423.
- [21] 张红斌, 熊其鹏, 蒋子良, 等. 联合 SENet 异构层特征融合与集成学习的材质图像识别[J]. 控制与决策, 2022, 37(6): 1632 – 1642.
- [22] Avallone S. An energy efficient channel assignment and routing algorithm for multi-radio wireless mesh networks[J]. Ad Hoc Networks, 2012, 10(6): 1043 – 1057.
- [23] Yuan S B, Wu B, Li P. Intra-pulse modulation classification of radar emitter signals based on a 1-D selective kernel convolutional neural network[J]. Remote Sensing, 2021, 13(14): 2799.
- [24] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.
- [25] Chen B Y, Zhang Z H, Liu N, et al. Spatiotemporal convolutional neural network with convolutional block attention module for micro-expression recognition[J]. Information, 2020, 11(8): 48 – 52.
- [26] 管贤淑, 葛万成. 基于 ResNet18 的减速带识别及其环境影响研究[J]. 通信技术, 2021, 54(3): 597 – 603.
- [27] 陈强普, 桑军, 项志立, 等. BN 对 VGG 神经网络的影响研究[J]. 合肥工业大学学报(自然科学版), 2018, 41(1): 35 – 39.
- [28] 李昂, 王晟全, 张晨. 基于 Rsnet 框架和 Mobilenet 迁移网络的遥感机场目标识别[J]. 重庆理工大学学报(自然科学版), 2021, 35(6): 131 – 136.
- [29] 李荔, 瞿洪桂, 孙家乐. 基于 ShuffleNet 的人脸图像质量评估方法[J]. 电子技术与软件工程, 2021(7): 132 – 135.
- [30] Zhang P Y, Zhong Y X, Li X Q. SlimYOLOv3: Narrower, faster and better for real-time UAV applications[C]//IEEE International Conference on Computer Vision Worksho, 2019: 37 – 45.
- [31] He Y, Liu P, Wang Z W, et al. Filter pruning via geometric median for deep convolutional neural networks acceleration[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2019: 4340 – 4349.