文章编号：2095-2163(2023)12-0023-09 中图分类号：TP392 文献标志码：A

基于 YOLOv7 的通用目标检测模型

钟 玲，陆国芳
（沈阳工业大学 软件学院，沈阳 110870）

摘 要：针对通用目标检测领域在自动提取特征的过程中会提取错误的目标检测区域信息，本文以 YOLOv7 模型作为基线模型进行改进，有效地提高检测精度。首先，在 YOLOv7 模型的主干网络中引入改进的注意力机制，在上采样模块中采用双三次插值，以增强浅层和深层的特征融合效果，减少区域信息丢失；其次，通过设计动态 IOU 阈值实现动态非极大值抑制，解决固定阈值导致检测边界框冗余的问题，提升准确性；最后，采用剪枝算法对网络模型进行轻量化处理，并使用深度可分离卷积替换原始卷积。实验结果显示，本文模型在数据集上的准确率、F1 值和召回率均高于其他模型，说明本文建立的基于 YOLOv7 模型改进的通用目标检测算法的有效性。

关键词：YOLOv7 模型；通用目标检测；注意力机制；双三次插值；剪枝算法；深度可分离卷积

A general object detection algorithm based on YOLOv7 model

ZHONG Ling, LU Guofang
（School of Software, Shenyang University of Technology, Shenyang 110870, China）

Abstract: In response to the generic target detection domain that extracts wrong target detection region information in the process of automatic feature extraction, this paper uses YOLOv7 model as the baseline model. An improved attention module is introduced in the backbone network of YOLOv7, and the upsampling algorithm is changed by using dual triple interpolation in the upsampling module; meanwhile, an optimized non-maximum suppression (NMS) method is implemented in detection, and dynamic IOU thresholds are designed to achieve dynamic NMS, which solves the problem of redundancy of detection bounding boxes due to fixed thresholds and reduces the false alarm rate; finally, a pruning algorithm is used to network Finally, the pruning algorithm is used to lighten the network model and replace the original convolutional model with depth-separable convolution. The experimental results show that the Acc, F1 values and recall rates of the model in this paper are higher than those of other models on the data set, which can illustrate the effectiveness of the improved YOLOv7-based general-purpose target detection model established in this paper.

Key words: YOLOv7 model; generic target detection; attention mechanism; bicubic interpolation; pruning algorithm; depth-separable convolution

0 引 言

机器视觉旨在模仿人类视觉系统，使计算机系统具备理解和解释图像或视频数据的能力。为了从图像中提取目标物体的特征参数，判断其类别，首先需要预先设置好边界分类条件；然后通过图像处理技术，对输入的图像进行分析，提取目标物体的关键特征参数，如像素分布、边缘、分辨率、纹理等关键信息；最后将目标物体的关键信息与已知类别的关键信息进行匹配，从而确定物体的类别[1]。在传统的图像处理中，经典的图像区域检测有基于 Canny 的检测、基于 Sobelmask 的检测和基于 LoG 算子的检测等检测方法[2]。然而，这些方法很容易受到背景噪音的影响，背景越复杂检测就越困难[3]。近年来，以神经网络为主体的机器视觉研究取得了巨大突破，其中卷积神经网络在目标特征提取方面展现出了强大优势[4]。

1 相关内容

YOLOv7 是 YOLO 系列最新的模型，具有更快的检测速度和更高的检测精度[5]。YOLOv7 模型在主干网络中引入了高效线性注意力模块（E-

基金项目：国家自然科学基金(61540069)；辽宁省教育厅科研基金(LJGD2020017)。

作者简介：陆国芳(1999-)，男，硕士研究生，主要研究方向：计算机视觉。
通讯作者：钟 玲(1970-)，女，硕士，副教授，主要研究方向：图像处理、数据挖掘。Email：xuzongzh@ sina.com

收稿日期：2022-12-17

ELAN)、CBS(Conv-BatchNorm-Swish)模块、CBM(Conv-BatchNorm-Mish)等模块,这些模块有助于提高模型的性能和效率[6]。YOLOv7 模型的网络结构如图 1 所示。

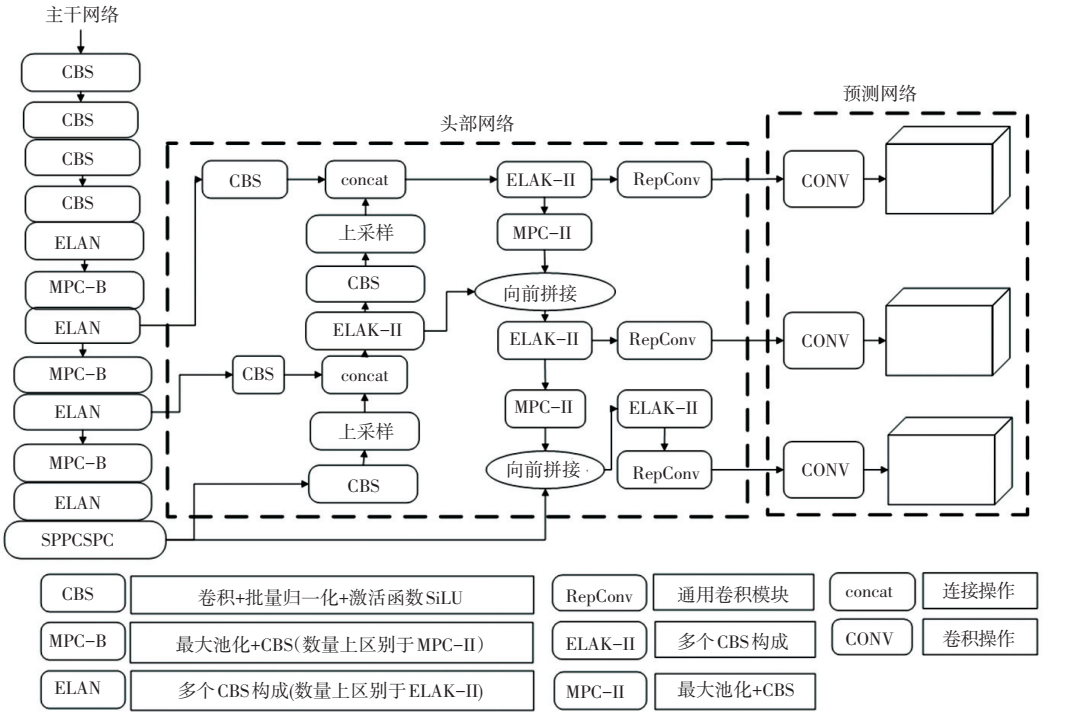


图 1 YOLOv7 网络结构

Fig. 1 YOLOv7 network structure

YOLOv7 网络由输入模块(Input)、主干网络(Backbone)、头部(Head)和预测(Prediction)4 个模块构成。Input 模块将输入的图像缩放至统一像素大小,以便满足主干网络的输入要求;主干网络由若干 BConv 层、E-ELAN 层以及 MPConv 层组成,其中 BConv 由卷积层、批量归一化层(Batch Normalization,BN)、LeakyReLU 激活函数组成,用于提取不同尺度的图像特征[7];E-ELAN 层通过引导不同特征组的计算块学习更多样化的特征,在不破坏原有梯度路径的情况下提高网络的学习能力;MPConv 卷积层在 BConv 层的基础上增加了 Maxpool 层,构成上下两个分支,上分支通过一个最大池化层对图像进行下采样,使特征图的长宽减半,下分支则先通过一个 1×1 的卷积运算,再通过一个步长为 2,大小为 3×3 的卷积运算,使图像的通道数减半,最后将上下分支提取到的特征进行融合,提高了网络的特征提取能力。Head 模块由路径聚合特征金字塔网络(Path Aggregation Feature Pyramid Network,PAFPN)组成,通过引入自底向上的路径使得底层信息更容易传递到高层,从而实现了不同层次特征的高效融合[8]。预测模块对 head 模块输出的不同尺度特征图进行图像通道数调整,利用 1×1 卷积来进行置信度、类别和锚框的预测[9]。虽然 YOLOv7 模型在常见任务场景(如行人、车辆检测)中表现出色,但检测仍然存在许多问题:

(1)与常见的场景相比,通用目标大多任意摆放、形状特征不定,采用水平框对图像中的目标进行标注,丢失了目标的方向性特征,增加了目标检测的难度;

(2)实际应用中,由于目标之间重叠或遮挡,造成在复杂背景下的目标检测不准确;

(3)在特征提取和特征融合模块中的 E-ELAN 模块虽然能增强网络的学习能力,但没有充分利用各节点的特征图输出,没有考虑到各模块对特征的提取能力不同。

针对上述问题,本文对 YOLOv7 模型进行改进:

(1)在主干网络中引入了改进的注意力模块,以增强模型对区域的关注能力;

(2)提出了优化的非极大值抑制方法,以解决区域边界框的冗余问题;

(3)设计了一个辅助的残差网络模块,以从目标中提取更多的特征信息,从而减少检测误差;

(4)设计并实现了一个结构化模型压缩算法,使得模型能够更好的部署在移动设备或嵌入式设备上。

2　通用目标检测模型

2.1　模型自适应

2.1.1　自适应网络系统

在目标检测任务中，YOLOv7 模型会生成许多边界框，为了提高准确性，需要在这些边界框中选择最有可能包含目标的框。非极大值抑制方法能消除多个重叠的边界框，并保留最可能包含目标的一个边界框。非极大值抑制方法的操作过程：首先，对所有的边界框按照预测的目标概率得分从高到低进行排序；其次，选择得分最高的边界框，并将其添加到最终的结果中；对于剩余的边界框，计算其与已选中的边界框的交并比，如果某个边界框的交并比（IOU）超过了设定的阈值，则将这个边界框删除；最后，重复上述过程，直到所有的框都被考虑过。通常，IOU 阈值是一个固定的概率值，由于可以在同一区域得到多个检测结果，因此固定的 IOU 阈值不适用于目标识别。

由于固定的 IOU 阈值没有过滤掉多余的检测结果，本文提出了自适应非极大值抑制方法提高检测结果的可靠性。自适应非极大值抑制方法提供了一个动态的 IOU 阈值，当多个检测结果相似时，提供较大的 IOU 阈值以消除重复的检测结果；当多个检测结果不同时，提供较小的 IOU 阈值以保留更多的检测结果。假设检测结果集为 D，T 为检测得分最高的边界框，最高检测框与检测结果 D 中每个 d_i 之间的 IOU 平均值 tb_m，公式（1）：

$$tb_m = \frac{1}{n}\sum_{i=1}^{n} IOU(T, d_i) \tag{1}$$

其中，n 为检测结果个数。

对比 tb_m 与默认 IOU 阈值（默认值通常为 0.5），取其中的最大者作为最终 IOU 阈值。

在自适应非极大值抑制方法中，当检测结果的大 IOU 值占比较高时，阈值也较高，重复检测结果会被剔除。当较小的值所占比例较高时，阈值也较小，保留较多的检测结果。因此，本文提出的自适应非极大值抑制方法更适合目标检测。

2.1.2　双三次插值

目标检测图像通常有很多边缘信息，对图像进行上采样时，会增加图像的像素，通常在图像需要上采样时使用插值方法。目标检测图像通常有很多边缘信息，不同的插值方法会产生不同的上采样效果，从而影响检测精度。在空间上采用适当的插值方法会使上采样的结果更加平滑和连续，也会减少不必要的噪音。

在目标检测中，需要更精确的感知区域来更好地表达物体特征[10]。在 YOLOv7 模型中，主干网络中的浅层特征包含丰富的细节信息、位置信息、颜色和纹理等低级语义信息，而深层特征包含丰富的高级语义信息。Head 模块将来自主干网络中不同层的特征图融合在一起，从而获得更全面、丰富的特征表示[11]。

在目标检测中，对于具有相对较小或复杂纹理的目标训练时希望模型能够提取更多的目标特征，因此本文优化了 YOLOv7 模型的插值方法。在 YOLOv7 模型中，双线性插值方法用于上采样处理，核心是在两个方向上进行线性插值。在上采样过程中，双线性插值只考虑插值点周围 4 个直接相邻像素点的影响，而不考虑邻近像素点的影响，从而忽略了物体的细节信息，比如边缘、纹理等信息。而在上采样模块中使用双三次插值，不仅考虑了 4 个相邻的像素点，还考虑了相邻像素点的影响，因此可以保留更详细的信息。双三次插值以其在原始图像中的位置计算其周围的 16 个像素点的加权平均值，根据这些加权平均值来估计目标像素的值，可以从浅层特征图获取更多的低级语义信息和细粒度信息，这对于目标检测非常重要。

2.2　注意力机制

注意力机制的核心就是只关注重点的信息，就像人类在观察周边事物时，往往是着眼于特征明显的区域仔细观察，这些区域就是人们感兴趣的区域，因此这些区域会被分配到更多的注意力。这个特点应用在目标检测中，网络模型可以把注意力集中在图像的局部特征上，形成注意力热图。通过权重属性来衡量图像中不同区域的注意力，从而进一步提取感兴趣区域的特征，同时舍弃掉不感兴趣的区域[12]。双重注意力机制通过引入通道注意力和空间注意力来提高模型对输入特征的关注度，从而改善模型在目标检测中的性能。本文根据双重注意力机制设计注意力模块，并将其添加到主干网络中，使检测模型更加集中于检测的区域。通道注意力结构和空间注意力结构如图 2 和图 3 所示。

通道注意力结构通过在通道维度上计算特征的平均值和标准差来生成一组特定的权重，以指导网络分配不同通道特征之间的相互作用，从而使网络能够更好地表达数据特征，通道注意力的计算公式如式（2）所示：

$$A_c(x) = \alpha(FC(\mathrm{AvgPool}(x)) + FC(\mathrm{MaxPool}(x))) \tag{2}$$

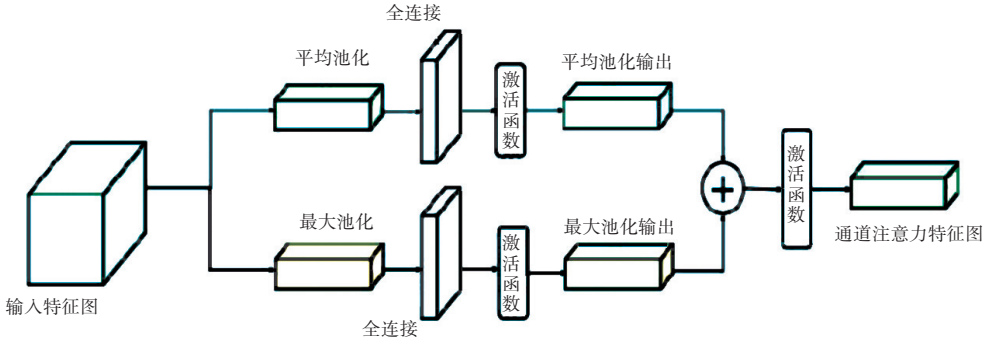


图 2 通道注意力结构图

Fig. 2 Channel attention structure diagram

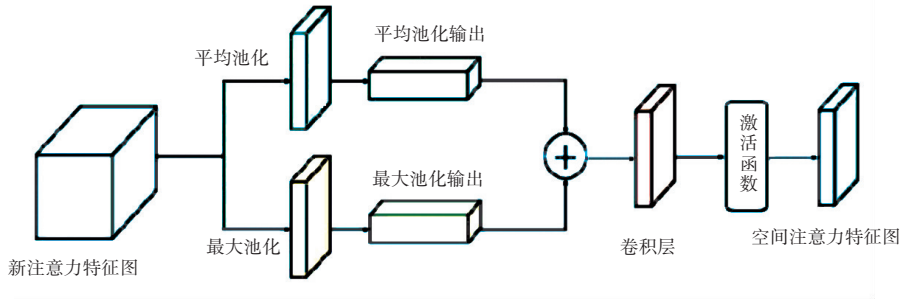


图 3 空间注意力结构图

Fig. 3 Spatial attention structure diagram

其中，α 代表 sigmoid 激活函数；AvgPool(x) 代表平均池化层；MaxPool(x) 代表最大池化层。

空间注意力结构通过学习注意力权重来确定哪些区域的特征更加重要，并将这些区域的特征进行增强。注意力权重可以基于不同信息源，如图像的空间分布、颜色分布、纹理等。学习了注意力权重后，空间注意力会将权重应用到特征表示上，对每个特征的元素进行加权操作，从而增强特征表示中重要的特征信息，过滤掉无关的噪声信息，空间注意力的计算公式如式(3)所示：

$$A_s(X) = \alpha(f^{3\times3}([\text{Avgpool}(X); \text{Maxpool}(X)])) \quad (3)$$

其中，α 代表 sigmoid 激活函数，$f^{3\times3}$ 代表 3×3 的卷积操作。

双重注意力融合模块如图 4 所示，将主干网络提取到不同尺度的特征图作为中间值输入到双注意力融合模块中，先用通道注意力模块进行处理，获得通道维度的注意力图，再通过空间注意力模块获取空间维度注意力图；把得到的注意力图与输入图像进行特征融合运算，生成一张新的特征图，具体计算方式如式(4)、式(5)所示：

$$X = A_c(x) \otimes x \quad (4)$$

其中，x 为通道注意力网络的输入图像，$Ac(x)$ 为输入图像在通道维度上的运算，通过在通道维度上的运算来实现输入图像在不同通道间的差异化。

$$Y = A_S(X) \otimes X \quad (5)$$

其中，X 为经过差异化处理的权重值与输入图像在通道上经过线性加权操作后得到的特征图，$As(X)$ 为输入图像在空间维度上的运算，实现特征图不同元素位置重要性的差异化处理。

最后，将 X 作为 $As(X)$ 的输入，实现对特征图的连续处理，将 $As(X)$ 运算的输出值与原特征图进行线性加权获得最终输出 Y。

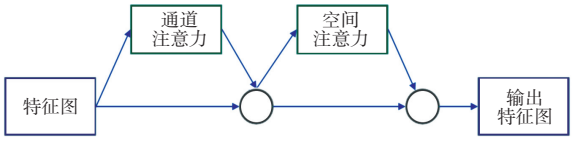


图 4 双重注意力融合模块结构图

Fig. 4 Attention fusion module structure diagram

本文在 YOLOv7 模型的 Head 模块中使用双三次插值方法对图像进行上采样，在预测模块中引入注意力模块进行不同尺度的特征融合，优化非极大值抑制方法以消除多个重叠的边界框，得到目标检测模型 AB-YOLOv7(Attention-Bilbic-Yolov7) 如图 5 所示。

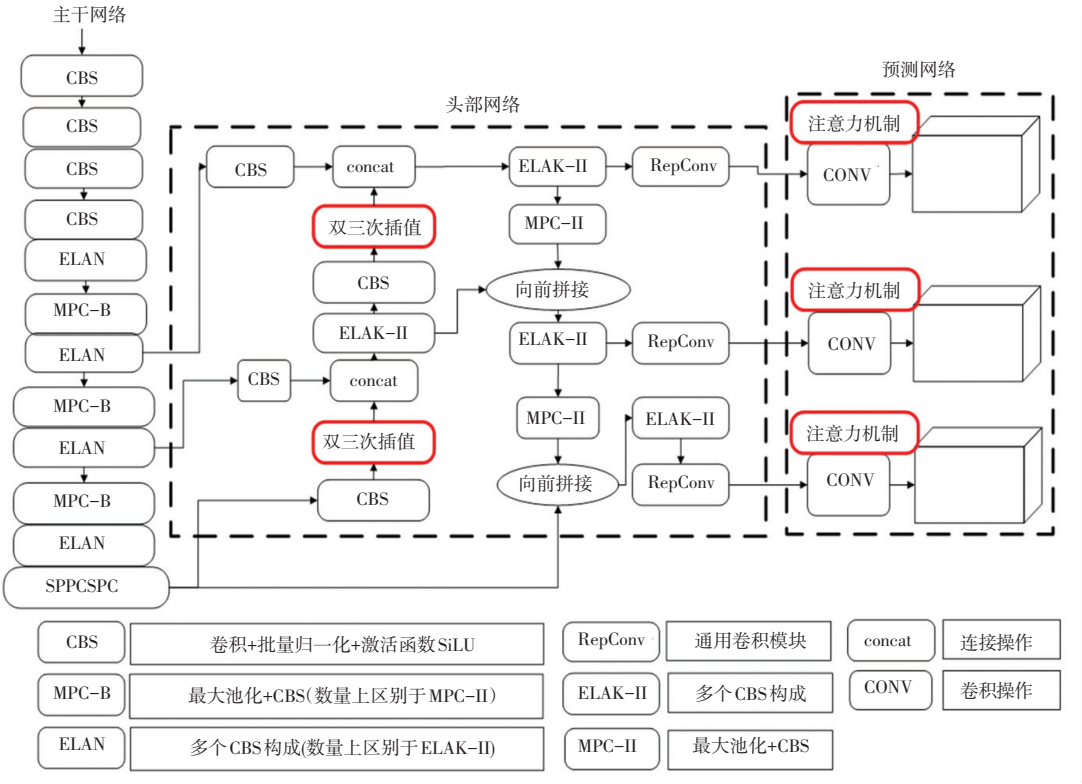


图 5　AB-YOLOv7(Attention-Bilbic-Yolov7)模型

Fig. 5　AB-YOLOv7(Attention-Bilbic-Yolov7) module

2.3　模型轻量化

2.3.1　基于 MobileviT 的轻量化改进

MobileViT 是一种可用于移动设备的轻量级视觉模型，结合了卷积神经网络和 Vision Transformer 的优点，能够加快网络的推理和收敛速度，使得网络更加稳定高效[13]。本文使用 MobileVit 作为 AB-YOLOv7 模型的特征提取网络，MobileVit 网络结构见表 1。

表 1　MobileVit 的网络结构

Table 1　MobileVit network structure

输入的图像尺寸	每个特征层所经历模块	输出的特征图通道数	Transformer 模块数	每一次操作的步长
$256^2\times3$	conv2d	16	-	2
$128^2\times16$	MV2	32	-	1
$128^2\times32$	MV2	64	-	2
$64^2\times64$	MV2	64	-	1
$64^2\times64$	MV2	64	-	1
$64^2\times64$	MV2	96	-	/2
$32^2\times96$	MVIT	96	2	1
$32^2\times96$	MV2	128	-	2
$16^2\times128$	MVIT	128	4	1
$16^2\times128$	MV2	160	-	2
$8^2\times160$	MVIT	160	3	1
$8^2\times160$	Conv2d	640	-	1
$8^2\times640$	Avgpool 8×8	-	-	-
$1^2\times640$	FC	-	-	-
$1^2\times k$	Conv2d	-	-	-

在表 1 中,MVIT 是 MobileVit 的核心模块,由局部信息编码模块、全局信息编码模块和特征融合模块组成,能够将局部特征信息和全局特征信息融合,从而充分的提取图像信息。MV2 表示 MobileNetV2 模块,MV2 先使用 1×1 卷积实现升维,再通过 3×3 的逐通道卷积提取特征,最后使用 1×1 卷积实现降维,以提高网络的表征能力。由于 MobileVit 网络第 10 层的特征被下采样了 32 倍,如果继续下采样会导致位置信息大量丢失。因此,放弃 MobileVit 网络的第十层之后,将剩余的网络作为模型的核心特征提取网络。

2.3.2 基于模型剪枝的轻量化改进

为了进一步对网络模型进行精简,本文设计并实现了一个结构化模型压缩算法,使模型能更好的部署在轻量级设备上。首先,通过训练一个大规模的过参数化模型来获得最佳的基础网络性能;其次,对大型模型进行剪枝和训练,即重新调整网络结构中的通道或层数,从而得到一个简化的网络结构;最后,使用剪枝后大型网络中保留的参数来初始化剪枝后的网络,继续在训练集上进行几轮微调以进一步调整模型参数。在微调过程中,对卷积层内核的权重进行排序,由于权重较低的卷积核对网络模型效果影响较低。因此,移除权重值排序最低的卷积核以及其对应的特征映射图,并清除相关的连接。在原来剪枝掉的卷积层中,重新生成新的卷积核并进行初始化权重赋值,用剩下的权重值对新的卷积核进行赋值并训练,得到最终的模型权重。

将结构化模型压缩算法引入到 AB-YOLOv7 模型中,在保持模型性能的前提下进一步减小模型的复杂度,使其能应用在资源受限的设备上,得到本文最终目标检测模型 ABL-YOLOv7 (Attention-Bilbic-Light-Yolov7),模型结构如图 6 所示。

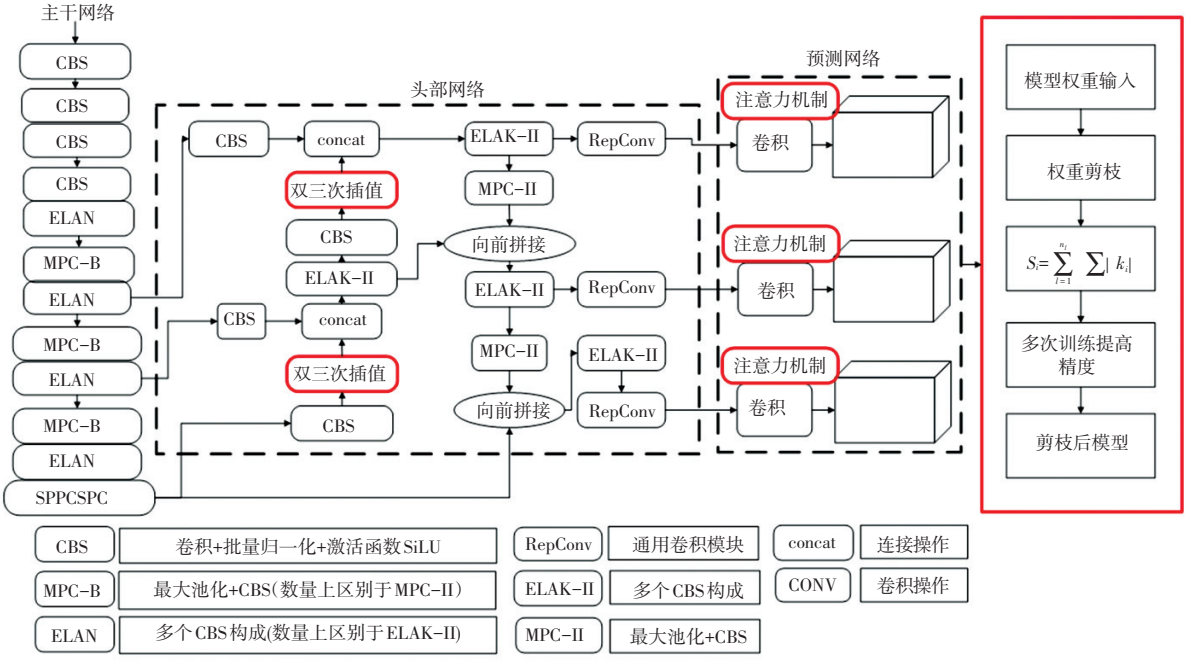


图 6 ABL-yolov7 模型结构

Fig. 6 ABL-yolov7 module structure

3 实 验

3.1 实验设置

本文采用 VisDrone2019-DET 数据集进行改进的通用目标检测模型实验。该数据集的数据在不同的场景下拍摄,涵盖了城市、乡村和工业区等多种场景,具有多样性;数据集包含多个目标类别,包括行人、车辆、自行车等常见的目标;数据集中的图像背景复杂程度不同,有些较为稀疏,有些则较为拥挤;在各种天气和光照条件下的拍摄,包含白天和夜晚两种光照条件,有些图像则在阴天和雾霾条件下拍摄。VisDrone2019-DET 数据集每个图像都具有精确的目标边界框标注,提供了目标在图像中的位置、大小信息和目标类别的标签,方便进行目标分类的

研究和评估。

本文实验的基础实验环境见表 2。

表 2　实验环境配置

Table 2　Experimental environment configuration

环境选项	配置信息
操作系统	Windows
开发语言	Python
深度学习框架	tensorflow
CPU 型号与主频	Intel Xeon Gold 6248R 3.00 GHz
GPU 型号和显存	Nvidia RTX 3090 24 GB
内存容量	256 GB
硬盘	4 T

本文的训练模型由 SGD 优化器在 Nvidia RTX 3090 上进行训练，训练图片分辨率为 640×640×3，批次大小为 12，迭代运行次数为 300，学习率为 0.002，学习率动量为 0.9。

通过目标检测领域通用的性能指标，包括正确率（*Acc*）、召回率（*R*）以及 *F*1 值（*F*1 - score）来评价改进模型的有效性。

准确率是指在预测为正例的样本中实际为正例所占的比例，数值越大表示预测的结果中实际为正例的占比越高，公式（6）：

$$Acc = \frac{TP}{TP + FP} \tag{6}$$

其中，*TP* 表示正确的分类样本数量，*FP* 表示将负样本预测为正例的数量。

召回率是指在所有实际为正例的样本中被正确预测为正例的样本所占的比例，公式（7）：

$$R = \frac{TP}{TP + FN} \tag{7}$$

其中，*FN* 表示将正样本预测为负例的数量。

*F*1 值通过计算准确率和召回率的调和平均值来评估模型的性能，综合考虑了准确率和召回率两个指标，公式（8）：

$$F1 = \frac{2 * R * Acc}{R + Acc} \tag{8}$$

3.2　对比实验

为了验证改进模型 ABL-YOLOv7 的有效性，与主流的旋转目标检测模型 ReDet、SCRDet、R3Det 和水平目标检测模型 RetinaNet、FCOS、YOLOv7 进行了对比实验，结果见表 3。

表 3　不同模型情感分析对比结果

Table 3　Comparison results of sentiment analysis of different models

%

模型	准确率	F1 值
ReDet	82.03	71.37
SCRDet	82.56	70.92
R3Det	85.51	74.54
RetinaNet	73.16	63.99
FCOS	69.13	64.71
YOLOv7	83.87	74.32
ABL-YOLOv7	91.42	88.09

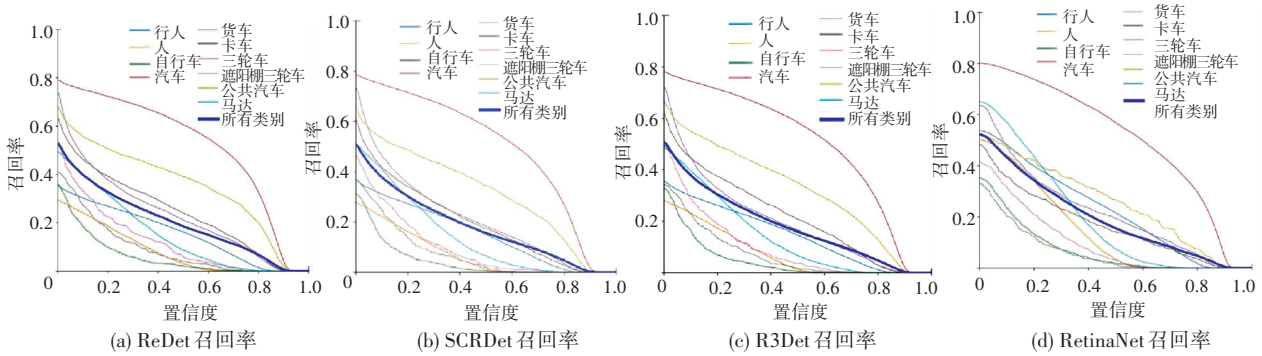
由表 3 可知，ABL-YOLOv7 模型与主流的旋转目标检测模型 ReDet、SCRDet、R3Det 相比准确率分别提高了 9.39%、8.86%、5.91%，*F*1 值分别提高了 16.72%、17.17%、13.55%、，24.1%、23.38%、13.77%；与水平目标检测模型 RetinaNet、FCOS、YOLOv7 相比，准确率分别提高了 18.26%、22.29%、7.45%，*F*1 值分别提高了 24.1%、23.38%、13.77%。

各模型的召回率如图 7 所示，可见 ABL-YOLOv7 模型与旋转目标检测模型 ReDet、SCRDet、R3Det 和水平目标检测模型 RetinaNet、FCOS、YOLOv7 相比召回率有所提升，表明 ABL-YOLOv7 模型能够提高复杂场景下的通用目标检测率。

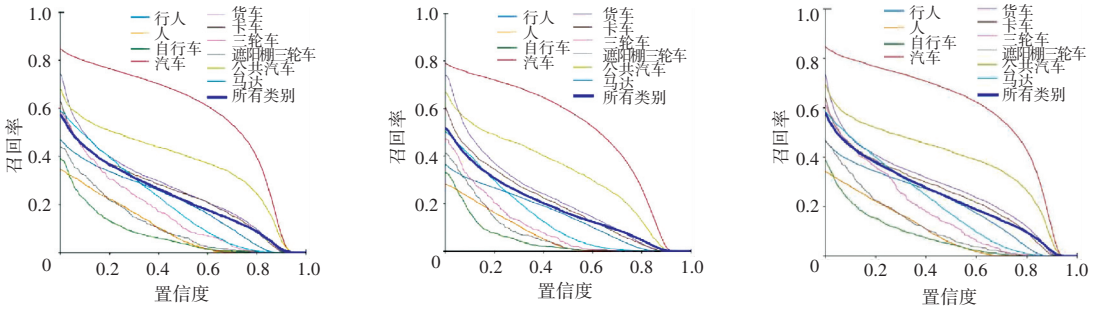
3.3　消融实验

为了进一步验证 ABL-YOLOv7 模型各个模块的有效性，本文进行了消融实验，实验结果见表 4。由表 4 可知，当分别去掉自适应网络、双三次插值、轻量化及剪枝时，准确率分别下降了 3.32%、3.11%、4.51%，*F*1 值分别下降了 2.88%、0.99%、2.18%；当只有自适应网络、双三次插值、轻量化及剪枝、注意力机制其中的一种时，准确率和 *F*1 值急剧下降，表明自适应网络、双三次插值、轻量化及剪枝、注意力机制能提升对通用目标的检测效果。

YOLOv7 模型与 ABL-YOLOv7 模型在 VisDrone2019 -DET 数据集上的部分检测效果如图 8 所示。可见对于遮挡目标，YOLOv7 模型存在漏检，检测效果不理想；ABL-YOLOv7 模型能够检测出遮挡目标，并且检测精度有所提高。



(a) ReDet 召回率 (b) SCRDet 召回率 (c) R3Det 召回率 (d) RetinaNet 召回率



(e) FCOS 召回率 (f) YOLOv7 召回率 (g) ABL-YOLOv7 算法召回率

图 7 各种算法召回率对比图

Fig. 7 Comparison chart of recall rates of various algorithms

表 4 ABL-YOLOv7 各模块消融实验结果

Table 4 Ablation experimental results of each ABL-YOLOv7 module %

序号	自适应网络	双三次插值	注意力机制	轻量化及剪枝	*Acc*	*F*1
1	√	√	√	√	91.42	88.09
2		√	√	√	88.10	85.21
3	√		√	√	88.31	87.10
4	√	√	√		86.91	85.91
5			√	√	81.37	80.25
6	√			√	84.63	82.59
7	√	√			80.88	80.69
8	√				76.63	75.40
9		√			74.53	70.29
10			√		72.48	70.10
11				√	76.28	79.59

(a) 原始 YOLOv7 (b) ABL-YOLOv7 模型

图 8 实验效果对比

Fig. 8 Comparison of experimental effect

4　结束语

对于不同场景下的通用目标检测，在自动提取特征的过程中面临着目标检测区域信息提取错误的问题。本文以 YOLOv7 为基础网络模型，针对通用目标检测模型进行改进，主要完成的工作如下：

(1)针对区域边界框的冗余问题，提出了优化的非极大值抑制方法，在操作过程中通过引入动态 IOU 阈值来替换固定的 IOU 阈值，从而解决固定 IOU 阈值导致检测边界框冗余的问题，提高检测模型的准确性和鲁棒性；

(2)通过引入双三次插值方法，计算目标像素周围的 16 个像素点的加权平均值进行插值，增强浅层特征与深层特征的语义融合，减少区域信息丢失，使上采样的结果更加平滑和连续；

(3)为了提取目标更多的特征，在预测模块中引入双重注意力机制，即空间注意力分支通过在每个空间位置上计算权重，以捕捉不同位置之间的依赖关系；通道注意力分支通过在每个通道上计算权重，得到与通道相关的注意力向量，从而捕捉全局通道相关性。将两个注意力分支的输出结合在一起，使模型在学习过程中可以同时关注空间位置和通道间的信息，从而更全面地理解输入数据的特征。

(4)为了加快网络的推理和收敛速度，使用 MobileVit 作为特征提取网络，设计并实现了一个结构化模型压缩算法，对模型进一步精简。

本文在 VisDrone2019-DET 数据集上进行实验，实验结果表明本文改进的模型在通用目标检测性能上表现更佳。

参考文献

[1] NOH Y, KOO D, KANG Y M, et al. Automatic crack detection on concrete images using segmentation via fuzzy cmeans clustering [C]//Proceedings of the International Conference on Applied System Innovation(ICASI). IEEE, 2017:877-880.

[2] YOUM M, YUN H, JUNG T, et al. High-speed crack detection of structure by computer vision [C]//Proceedings of the KSCE 2015 Convention Civil Expo and Conference. Gunsan, Korea, 2015: 28-30.

[3] LIU Y, YUN H, JUNG T, et al. Richer convolutional features for edge detection[J]. IEEE Trans. Pattern Anal. Mach. Intell, 2019, 41(8):1939-1946.

[4] MANINIS K K, PONT T J, ARBELÁEZ P, et al. Convolutional oriented boundaries: From image segmentation to high-level tasks [J]. IEEE Transactions on Pattern Analysis and Mechine Intelligence, 2018, 40(4): 819-833.

[5] 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[EB/OL]. [2022-07-06]. https://arxiv.org/abs/2207.02696.

[6] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4: scaling cross stage partial network[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021: 13024-13033.

[7] JIANG T T, CHENG J Y. Target recognition based on CNN with leakyReLU and PReLU activation functions[C]//Proceedings of the 2019 IEEE Conference on Sensing, Diagnostics, Prognostics, and Control(SDPC). New York: IEEE Press, 2019: 718-722.

[8] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. [2021-07-18]. https://arxiv.org/abs/2107.08430.

[9] HAN J M, DING J, XUE N, et al. ReDet: A rotation-equivariant detector for aerial object detection[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2021: 2786-2795.

[10] YANG X, YANG J R, YAN J C, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2020: 8231-8240.

[11] YANG X, YAN J, FENG Z, et al. R3Det: Refined single-stage detector with feature refinement for rotating object [C]//Proceedings of the IEEE Conference on Association for the Advancement of Artificial Intelligence(AAAI). New York: IEEE Press, 2021:3163-3171.

[12] 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. 2017:2980-2988.

[13] YANG Q, ZHANG S, CHEN W, et al. A lightweight network for high dynamic range imaging[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA: IEEE, 2022: 824-832.