智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

双向特征融合与注意力机制结合的目标检测

赵文清， 杨盼盼

引用本文:
赵文清, 杨盼盼. 双向特征融合与注意力机制结合的目标检测[J]. 智能系统学报, 2021, 16(6): 1098–1105.
ZHAO Wenqing, YANG Panpan. Target detection based on bidirectional feature fusion and an attention mechanism[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(6): 1098–1105.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202012029

DOI: 10.11992/tis.202012029
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20210903.1028.002.html

双向特征融合与注意力机制结合的目标检测

赵文清[1,2], 杨盼盼[1]
（1. 华北电力大学 控制与计算机工程学院, 河北 保定 071003; 2. 复杂能源系统智能计算教育部工程研究中心, 河北 保定 071003）

摘　要: 目标检测使用特征金字塔检测不同尺度的物体时, 忽略了高层信息和低层信息之间的关系, 导致检测效果差; 此外, 针对某些尺度的目标, 检测中容易出现漏检。本文提出双向特征融合与注意力机制结合的方法进行目标检测。首先, 对 SSD(single shot multibox detector) 模型深层特征层与浅层特征层进行特征融合, 然后将得到的特征与深层特征层进行融合。其次, 在双向融合中加入了通道注意力机制, 增强了语义信息。最后, 提出了一种改进的正负样本判定策略, 降低目标的漏检率。将本文提出的算法与当前主流算法在 VOC 数据集上进行了比较, 结果表明, 本文提出的算法在对目标进行检测时, 目标平均准确率有较大提高。

关键词: 特征金字塔; 双向融合; 特征提取; SeNet 注意力机制; 样本; 语义信息; 目标检测; 深度学习
中图分类号: TP391　　**文献标志码:** A　　**文章编号:** 1673−4785(2021)06−1098−08

中文引用格式: 赵文清, 杨盼盼. 双向特征融合与注意力机制结合的目标检测 [J]. 智能系统学报, 2021, 16(6): 1098–1105.
英文引用格式: ZHAO Wenqing, YANG Panpan. Target detection based on bidirectional feature fusion and an attention mechanism[J]. CAAI transactions on intelligent systems, 2021, 16(6): 1098–1105.

Target detection based on bidirectional feature fusion and an attention mechanism

ZHAO Wenqing[1,2], YANG Panpan[1]
(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China; 2. Engineering Research Center of the Ministry of education for Intelligent Computing of Complex Energy System, Baoding 071003, China)

Abstract: When using a feature pyramid to detect objects of different dimensions, the relationship between high- and low-level information is ignored, resulting in a poor detection effect; in addition, for targets of a certain scale, detection is easily missed. In this paper, a method combining bidirectional feature fusion and an attention mechanism is proposed for target detection. First, the deep and shallow feature layers of the single-shot multibox detector (SSD) model are fused, then the obtained features are fused with the deep feature layer. Second, the channel attention mechanism is added to the two-way fusion to enhance semantic information. Finally, an improved positive and negative sample decision strategy is proposed to reduce the target misdetection rate. The algorithm proposed in this paper is compared with the current mainstream algorithms in the VOC dataset. The results show that the average accuracy of the proposed algorithm is greatly improved when detecting targets.

Keywords: feature pyramid; bidirectional fusion; feature extraction; SeNet attention mechanism; sample; semantic information; target detection; deep learning

目标检测是计算机视觉领域的重要研究方向。现阶段的目标检测方法主要有 2 种: 一种是基于分类的两阶段法, 另一种是基于回归的单阶段法。2014 年 Girshick 等[1] 首次提出 R-CNN(region convolutional neural networ, R-CNN) 使用选择性搜索[2] 生成的区域建议。该算法相比传统算法

收稿日期: 2020−12−17.　**网络出版日期:** 2021−09−03.
基金项目: 河北省自然科学基金项目 (F2021502013); 中央高校基本科研业务费面上项目 (2020MS153, 2021PT018).
通信作者: 赵文清. E-mail: zhaowenqing@ncepu.edu.cn.

得到了很大的提升, 但是不能满足实时性要求。在 2015 年, Girshick 等[3-4] 在 R-CNN 和 SPPNet (spatial pyramid pooling net) 网络的基础上提出了 Fast-RCNN 网络。Fast R-CNN比 R-CNN 快, 但是这个网络还是不能实现端到端的训练, 因此 He 等[5] 就端到端的问题提出了解决方法, 提出了 Faster-RCNN 网络。Faster-RCNN 将区域建议阶段和分类阶段结合到一个模型中, 允许端到端学习。从 R-CNN 到 Faster-RCNN 的发展过程, 都是先得到候选框, 对候选框做处理之后再进入分类器做处理, 这种类型的方法统称为两阶段目标检测法。两阶段方法不能满足实时性要求, 基于回归的单阶段检测法应运而生。在单阶段检测方法中最具代表性的是 YOLO[6-7] 系列和 SSD[8] 系列。YOLO 算法检测速度快, 但是容易出漏检。为了改善 YOLO 算法中的缺陷, Liu 等[8] 提出 SSD 算法, SSD 使用多尺度特征进行分类和回归任务, 但是其效果仍有待提高。为了提高目标检测的准确率, DSSD[9] 将反卷积技术应用于 SSD的所有特征图, 以获得放大的特征图。但是, 由于将反卷积模块应用于所有特征图, 因此存在模型复杂度增加和速度降低的局限性。此外, 针对尺度较小的目标在检测中存在的问题, 相关学者提出了一系列改进算法。Singh 等[10] 提出 SNIP(scale normalization for image pyramids) 算法, 该算法主要思想是对特征图进行不同倍数采样, 通过实验给出相对最优的检测小目标的特征图尺寸, 最后通过 Soft-NMS 融合不同分辨率下的检测结果。Wen 等[11] 通过融合多尺度和反卷积获得了更高的准确性。Li 等[12] 使用生成对抗网络 (GAN)[13] 和低分辨率特征作为 GAN 的输入来生成高分辨率特征。但是这些方法在检测中仍然存在不能有效提取目标特征的缺陷。

特征金字塔网络 (feature pyramid networks, FPN) 将高层特征图进行上采样, 然后与浅层特征进行加法合并操作, 得到新的表达能力更强的特征图。但是 FPN[14] 在融合之前, 不同的特征之间存在的语义信息差距较大, 直接进行融合会降低多尺度特征的表达能力; 其次, 在融合过程中, 信息自高层向低层进行传播时会丢失一些信息, 而且 FPN 在进行特征提取时, 忽略了高层信息与低层信息之间的关系, 导致检测的精确度有所降低。

本文提出了双向特征融合与注意力机制结合的目标检测方法。该方法首先对 SSD 模型深层特征层进行双线性插值放大与浅层特征层进行特征融合, 然后将得到的特征通过降采样的方式与深层信息进行融合, 这样可以充分结合深层和浅层信息以提升浅层特征网络对小目标的表达能力, 同时, 引入通道注意力机制 SeNet 结构对特征融合后的特征图进行更新, 提取出更有利于检测特征的通道。最后, 针对目标检测中小目标漏检的情况, 优化了正负样本的判定策略, 使得筛选出的框的数量增加, 进一步提高小目标检测的精度。

1　相关技术和理论

SSD 对目标进行分类和回归的过程: 首先, 输入大小为 300 像素×300 像素的图片, 然后使用卷积神经网络提取同卷积层中目标的特征, 本文采用的是 VGG 作为特征提取网络; 其次, 提取到的特征根据不同特征层 anchor 的数目, 对 anchor 进行分类, 利用边界框回归对其进行位置调整, 得到最终的建议框; 然后, 再使用非极大抑制算法对这些建议框进行筛选, 得到检测结果。

特征金字塔网络 (feature pyramid networks, FPN) 对高层特征图进行上采样, 然后与浅层特征进行加法合并操作, 得到新的表达能力更强的特征图, 这种多尺度的特征图在面对不同尺寸的物体时, 具有更好的鲁棒性[15-17]。此外, 这种特征金字塔结构是一种通用的特征提取结构, 可以应用到不同的网络框架中, 例如 Faster-RCNN、SSD 等。针对小目标的检测, 可以检测出更加丰富的特征, 进而提升其准确度。

2　基于改进 SSD 算法的目标检测

本模型采用 VGG16 作为骨干网络, 使用双向特征网络完成特征提取, 首先, 从骨干网络中提取出 Conv3_3、Conv4_3、Conv5_3、FC7、Conv8_2 和 Conv9_2 特征图, 然后分两个部分进行双向特征融合, 如图 1 所示。

2.1　双向特征融合过程

2.1.1　自顶向下特征融合

图 1 中自顶向下的特征融合分为 6 个步骤:

1) 对 256 通道的特征图 Conv9_2 使用 1×1 卷积升维成 512 通道的特征图 P_1;

2) 对 P_1 进行 2 倍上采样后与 Conv8_2 进行融合, 对融合结果使用一个 3×3 的卷积层消除上采样过程中的混叠效应, 使用 SeNet 对消除混叠效应后的特征图进行通道更新得到特征图 P_2;

3) 对 1 024 通道的特征图 FC7 使用 1×1 卷积降维成 512 通道, 对 P_2 进行 2 倍上采样后与降维后的 FC7 进行融合, 使用 3×3 卷积层消除混叠

效应，接着，使用 SeNet 进行通道更新，得到特征图 P_3；

4) 由于特征图 Conv5_3 通道数为 512，并且特征图尺寸与 P_3 大小一致，因此直接对 Conv5_3 和 P_3 进行相加融合，对融合结果同样使用 3×3 的卷积和 SeNet 得到特征图 P_4；

5) 对 P_4 进行 2 倍上采样，由于 Conv4_3 特征图尺寸为 38×38 比较大，因此在进行 P_4 与 Conv4_3 的特征融合时需要先对 P_4 进行归一化，接着使用 3×3 的卷积和 SeNet 得到特征图 P_5；

6) 对 P_5 进行 2 倍上采样，Conv3_3 特征图尺寸为 75×75，先对 P_4 进行归一化，再与 Conv3_3 特征相加融合，接着使用 3×3 的卷积和 SeNet 得到特征图 P_6。

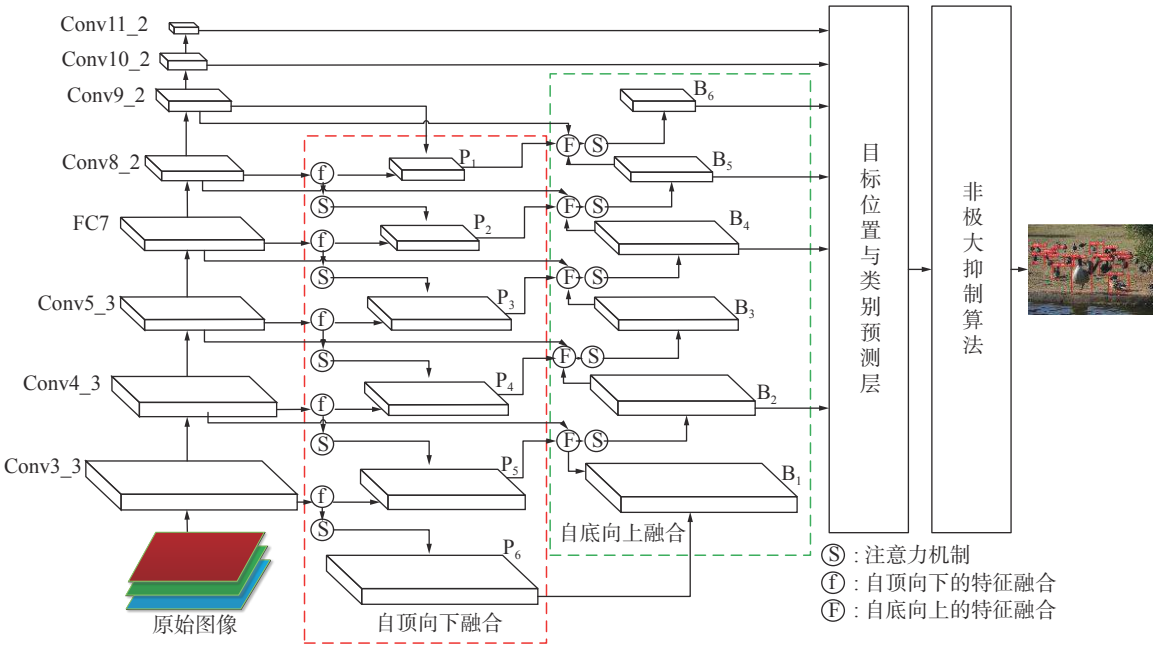

图 1　双向特征融合与注意力机制结合的模型

Fig. 1　Target model based on the combination of bidirectional feature fusion and an attention mechanism

2.1.2　自底向上特征融合

图 1 中自底向上的特征融合分为 5 个步骤：

1) 将自顶向下融合中步骤 6) 中的 P_6 作为起始特征图 B_1，然后，对 B_1 进行 2 倍降采样，接着对降采样后的特征图进行归一化，对 Conv4_3、P_5 和归一化后的特征图进行相加融合，对融合后的结果使用 3×3 的卷积和 SeNet 得到特征图 B_2；

2) 对 B_2 进行 2 倍降采样，由于 Conv4_3 特征图尺寸为 38×38，比较大，因此，在 B_2 与 Conv5_3、P_4 融合时，需要先对 B_2 进行 L_2 归一化，接着使用 3×3 的卷积和 SeNet 得到特征图 B_3；

3) 由于特征图 Conv5_3 通道数为 512，并且特征图尺寸与 B_3 大小一致，因此直接对 B_3 与 FC7、P_3 进行相加融合，对融合后的结果同样使用 3×3 的卷积和 SeNet 得到特征图 B_4；

4) 对 B_4 进行 2 倍降采样，然后与 Conv8_2、P_2 进行相加融合，对融合后的结果同样使用 3×3 的卷积和 SeNet 得到特征图 B_5；

5) 对 B_5 进行 2 倍降采样，然后与 Conv9_2、P_1 进行相加融合，对融合后的结果同样使用 3×3 的卷积和 SeNet 得到特征图 B_6。

2.1.3　双向特征融合模块

文献 [18-22] 中提出了不同的特征融合的方法可以增强语义信息的表达能力，本文提出的双向特征融合机制首先将深层特征层进行双线性插值放大与浅层特征层进行特征融合，然后将得到的特征通过降采样的方式与深层信息进行融合，这样可以充分结合深层和浅层信息进行特征提取，进而提升浅层特征网络对目标语义信息的表达能力。双向特征融合过程细节如图 2 所示，图 1 中的自顶向下的方式进行特征融合如图 (a) 所示，自底向上的方式进行特征融合如图 (b) 所示，图中以 Conv8_2 和 Conv9_2 为例进行特征融合的表示。如图 2 所示，首先，将 SSD 提取到的高层特征 Conv9_2 使用双线性插值的方法与 SSD 提取的原始特征 Conv8_2 通过自顶向下的方式进行融合得到 P_3；然后将 B_4 通过降采样的方式与 P_2 进行融合。在进行自底向上的特征融合过程中，为了得到表征能力更强的特征，同时融合了 SSD 提取的原始 Conv8_2 特征。为了防止出现梯度爆炸问题，在进行双向特征融合中，在自顶向下和自底向上都加入了 BN 层，经过 ReLU 激活函数和 SeNet 注

意力机制，得到最终的 B_5 特征，最后送入检测层。

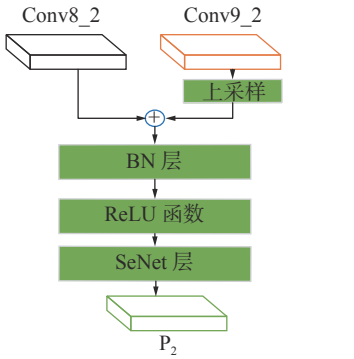

图 2　双向特征融合过程

Fig. 2　Bidirectional features fusion process

2.1.4　注意力机制模块

注意力机制是一种资源分配的策略，广泛应用在计算机视觉各个方向[23-24]。注意力机制让神经网络更多地去关注与特征相关的细节信息，这样可以加快信息的处理时间，进而一定程度上提高计算机处理信息的效率，最终提高特征的表达能力。本文选用的是 SeNet 注意力机制，该机制通过学习通道之间的关系，然后对卷积得到的特征图进行相应地处理，最终得到一个与通道数相同维度的向量，并且将此向量作为一个权重数，加到相应的通道上。注意力机制模型如图 3 所示。

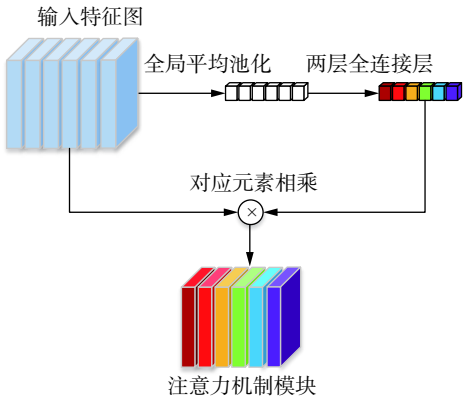

图 3　注意力机制模块

Fig. 3　Attention mechanism module

首先，对输入的特征图进行全局池化，再连接两个全连接层，第一个全连接层神经元个数是 $c/16$，第二个全连接层神经元个数为 c。然后，再接一个 Sigmod 层输出 1 ×1 ×c，得到各个通道对应的权重系数，最后与输入的特征图元素相乘。

2.2　目标位置与类别预测层

首先，将提取到的特征层输入到预测层，然后，根据不同特征层 anchor 的数目和类别，对 anchor 进行分类，利用边界框回归对其进行位置的调整，得到最终的建议框。

2.2.1　对 anchor 进行的改进

在网络模型划分正负样本阶段，anchor 和真实框的交并比 (intersection over union, IoU) 值作为正负样本划分的依据。低于通常使用的阈值作为负样本，高于通常使用的阈值作为正样本。当 IoU 低于通常使用的阈值时存在两种情况。第 1 种情况，anchor 大于真实框。但是，anchor 中的上下文信息在目标检测中也有很大的作用，通过上下文信息可以有效地检测目标的信息，进而改善目标的漏检情况。第 2 种情况，anchor 小于真实框，但是这些 anchor 包含被检测目标更多的局部特征。针对这两种情况，本文提出了一种改进的正负样本判定策略，保留低于 IoU 阈值但与目标相关的 anchor 作为正样本。将式 (1) 用于正负样本的判断中：

$$a = \frac{S_P \cap S_T}{\min(S_P, S_T)} \tag{1}$$

式中：S_P 代表先验框集合；S_T 代表真实框集合；$S_P \cap S_T$ 代表先验框和真实框的交集；$\min(S_P, S_T)$ 代表先验框和真实框中面积较小的一个。由经验可得 a 的值为 0.9。第一种情况下若 $a > 0.9$，表示 anchor 覆盖住 90% 以上的真实目标，就判断为正样本。第 2 种情况下，表示 anchor 有 90% 以上和目标真实框覆盖，就判定为正样本。

2.2.2　损失函数

网络的损失函数由两部分组成，分别是置信度损失和位置损失，如式 (2)~(5) 所示：

$$L_f(t,c) = -\sum_{i \in p}^{N} t_{ij} \log(\hat{c}_i^p) - \sum_{i \in n} \log(\hat{c}_i^0) \tag{2}$$

$$L(t,c,l,g) = \frac{1}{N}(L_f(t,c) + \alpha L_l(t,l,g)) \tag{3}$$

$$\hat{c}_i^p = \frac{\exp(c_i^p)}{\sum_p \exp(\hat{c}_i^p)} \tag{4}$$

$$L_l(t,l,g) = \sum_{i \in p}^{N} \sum_{m \in (x,y,w,h)} t_{ij}^k \text{smooth}_{\text{L1}}(l_i^m - \hat{g}_j^m) \tag{5}$$

式中：t 为输入图像；L_f 表示置信度损失；L_l 表示位置损失；i 表示属于正样本的个数；j 表示属于负

样本的个数；x 表示所选框的横坐标；y 表示所选框的纵坐标；w 表示所选框的宽；h 表示所选框的高；m 是框的位置的集合；N 为匹配成功的先验框个数。

2.3　非极大抑制算法

非极大抑制算法实现步骤：首先根据设定的目标框的置信度阈值排列候选框列表，然后选取置信度最高的框添加到输出列表，并将其从候选框列表中删除。最后计算置信度最高的框与候选框列表中的所有框的 IoU 值，删除大于阈值的候选框。重复上述过程，直到候选框列表为空。

3　实验以及结果分析

3.1　实验设备以及参数设置

本实验 GPU 使用的是 NVIDIA GeForce GTX 1080Ti。Batch-size 设置的是 16，防止模型过拟合和模型收敛，先将学习率设置为 0.001，训练 8 万次，再将学习率降低 10 倍设置为 0.000 1 训练 2 万次，最后再降低 10 倍设置为 0.000 01 训练 2 万次，共 12 万次。实验中的深度学习框架使用的是 pytorch1.1。

VOC 数据集有 20 类，包括训练集、验证集和测试集。为了防止过拟合，本文采用 VOC2007+VOC2012 数据集进行训练，然后再采用 VOC2007 测试集进行测试。

3.2　实验评价指标

本文实验采用平均准确率 (average precision, AP)、召回率 (average recall, AR) 和所有类别的均值平均精度 mAP (mean average precision) 作为精度评价指标；f/s(frame per second) 作为速度评价指标，代表每秒内可以处理的图片数量。其中，AP 代表正确检测为正占全部检测为正的比例[25]。

3.3　实验结果及对比

通过改进 SSD 模型，本文模型在 VOC2007 和 VOC2012 训练集上训练，在不同尺度上的平均准确率和召回率值如表 1 所示。

表 1　不同尺度的检测结果

Table 1　Test results of different scales

算法	平均准确率mAP/%	召回率AR/%	速度/f·s^{-1}
本文SSD300	80.6	90.2	33
本文SSD321	81.5	91.7	30
本文SSD512	82.5	93.5	12

由表 1 可知，在本文所提出的算法中，当输入尺寸大小为 300 像素×300 像素时，平均准确率为 80.6%，召回率为 90.2%，速度为 33 f/s；当输入大小为 321 像素×321 像素时，平均准确率为 81.5%，召回率为 91.7%，速度为 30 f/s，当输入大小为 512 像素×512 像素时，平均准确率为 82.5%，召回率为 93.5%，速度为 12 f/s。

为了进一步验证本文算法的有效性，进行了消融实验，其结果如表 2 所示。

表 2　不同方法对精度的影响

Table 2　Influence of different improvement methods on accuracy

网络类型	mAP/%	速度/f·s^{-1}
原始SSD算法	77.5	46
SSD+双向特征融合	79.3	42
SSD+双向特征融合+SeNet	79.7	40
SSD+anchor改进	77.9	37
SSD+双向特征融合+anchor+anchor改进+SeNet	80.6	33

由表 2 可知，本文所改进的方法中，在双向特征融合的基础上加入注意力机制可以使平均准确率提高 2.2%，anchor 的改进可以使平均准确率提高 0.4%，整体的平均准确率可以提高 3.1%。

本文模型在 VOC2007 和 VOC2012 训练集上训练，在 VOC 2007 测试集上的结果与主流算法 Faster-RCNN、YOLOV2、DSSD321 等实验结果对比如表 3 所示。

表 3　VOC2007 测试结果对比

Table 3　Comparison of test results of VOC2007

算法	基础网络	尺寸/像素×像素	mAP/%	速度/f·s^{-1}
Faster-RCNN[5]	VGGNet	600×1 000	73.2	7
Faster-RCNN[5]	ResNet-101	600×1 000	76.4	2.4
YOLOV2[6]	Darknet-19	352×352	73.7	81
SSD300[7]	VGGNet	300×300	77.5	46
Ours300	**VGGNet**	**300×300**	**80.6**	**33**
DSSD321[8]	ResNet-101	321×321	79.5	9.5
Ours321	**VGGNet**	**321×321**	**81.5**	**30**
SSD512[7]	VGGNet	512×512	78.5	19
YOLOV2[6]	Darknet-19	544×544	78.6	40
DSSD512[8]	ResNet-101	512×512	81.5	6.6
Ours512	**VGGNet**	**512×512**	**82.5**	**12**

由表 3 可知，输入图像尺寸为 300 像素×300 像素时，本文提出的模型平均准确率较 YOLOV2、SSD300、DSSD321 分别提高了 6.9%、3.1%、1.1%。同时，SSD 模型每秒检测 46 张图像，而本文改进后 SSD 模型每秒检测 33 张图像，相比于 SSD 模

型，改进模型的检测速度略有下降，这是由于改进双向特征融合时模型计算量有所增加，从而影响模型的检测速度。

VOC2007 具有 20 类目标，其中每一类别目标与主流的目标检测算法的 AP 的对比结果如表 4 所示。

表 4　VOC2007 测试结果详细比较

Table 4　Detailed comparison of test results of VOC2007　　m AP/%

类别方法	Faster-Rcnn[5]	SSD300[7]	SSD512[7]	DSSD321[8]	DSSD513[8]	Ours300	Ours321	Ours512
飞机	76.5	79.5	84.8	81.9	86.6	84.4	85.5	**87.8**
自行车	79.0	83.9	85.1	84.9	86.2	85.7	86.7	**88.0**
鸟	70.9	76.0	81.5	80.5	82.6	78.2	81.5	**84.0**
船	66.5	69.6	73.0	68.4	74.9	76.3	77.5	**81.1**
瓶	52.1	50.5	57.8	53.9	52.5	**61.8**	58.9	61.5
公共汽车	83.1	87.0	87.8	85.6	**89.0**	87.9	85.8	82.1
小汽车	84.7	85.7	88.3	86.2	88.7	87.2	88.3	**88.8**
猫	86.4	88.1	87.4	88.9	88.8	88.8	89.0	**89.5**
椅子	52.0	60.3	63.5	61.1	65.2	65.9	67.6	**68.9**
奶牛	81.9	81.5	85.4	83.5	**87.0**	85.9	86.2	86.8
桌子	65.7	77.0	73.2	78.7	78.7	77.7	**80.4**	80.3
狗	84.8	86.1	86.2	86.7	**88.2**	86.2	87.6	87.4
马	84.6	87.5	86.7	88.7	**89.0**	88.5	87.6	87.7
摩托车	77.5	83.9	83.9	86.7	87.5	87.0	87.0	**87.8**
人	76.7	79.4	82.5	79.7	83.7	81.5	84.1	**84.7**
植物	38.8	52.3	55.6	51.7	51.5	59.1	60.3	**62.0**
山羊	73.6	77.9	81.7	78.0	**86.3**	78.5	83.3	83.3
沙发	73.9	79.5	79.0	80.9	81.6	81.2	82.3	**82.3**
火车	83.0	87.6	86.6	87.2	85.7	88.9	89.3	**89.2**
电视	72.6	76.8	80.0	79.4	**83.7**	80.6	81.2	80.2

由表 4 可知本文算法平均精度有一定的提高，尤其对小目标的提升更为显著。

为了验证本文算法的有效性，对原始 SSD 算法和本文算法的目标检测结果进行了可视化展示，如图 4、5 所示。本文算法检测框和目标贴合得更为紧密，同时对小目标的漏检和误检情况有一定程度的改善。

(a) 较小尺寸目标识别结果　(b) 被遮挡的目标识别结果　(c) 密集场景的小尺寸目标识别结果

图 4　SSD 检测效果

Fig. 4　SSD detection results

(a) 识别出较小尺寸目标　(b) 识别出被遮挡目标　(c) 识别出密集场景目标

图 5　改进的 SSD 检测效果

Fig. 5　Improved SSD detection results

4　结束语

针对 SSD 目标检测中存在的问题，提出了双向特征融合和注意力机制结合的目标检测方法。与传统的特征金字塔不同，该方法引入双向融合特征金字塔，充分考虑高层与低层信息之间的关系，进一步得到了语义信息更丰富的多尺度特征

图。同时，在自顶向下和自底向上的双向融合中加入了通道注意力机制，提高了特征融合的效率。最后，针对目标的漏检情况，本文提出了一种改进的正负样本判定策略，提取到被检测目标更多的局部特征。经过实验对比，本文所提出的算法模型相较于传统 SSD 算法，平均准确率方面提高 3.1%，表明了本文所提算法的有效性。

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. Columbus, USA, 2014: 580−587.

[2] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154–171.

[3] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, 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] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1440−1448.

[5] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137–1149.

[6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 779−788.

[7] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 6517−6525.

[8] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherland, 2016: 21−37.

[9] FU Chengyang, LIU Wei, RANGA A, et al. DSSD: deconvolutional single shot detector[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, USA, 2017: 2881−2890.

[10] SINGH B, DAVIS L S. An analysis of scale invariance in object detection-SNIP[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 3578−3587.

[11] 温静, 李雨萌. 基于多尺度反卷积深度学习的显著性检测 [J]. 计算机科学, 2020, 47(11): 179–185.
WEN Jing, LI Yumeng. Salient object detection based on multi-scale deconvolution deep learning[J]. Computer science, 2020, 47(11): 179–185.

[12] LI Jianan, LIANG Xiaodan, WEI Yunchao, et al. Perceptual generative adversarial networks for small object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 1951−1959.

[13] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada, 2014: 2672−2680.

[14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 936−944.

[15] 刘涛, 汪西莉. 采用卷积核金字塔和空洞卷积的单阶段目标检测 [J]. 中国图象图形学报, 2020, 25(1): 102–112.
LIU Tao, WANG Xili. Single-stage object detection using filter pyramid and atrous convolution[J]. Journal of image and graphics, 2020, 25(1): 102–112.

[16] 陈景明, 金杰, 王伟锋. 基于特征金字塔网络的改进算法 [J]. 激光与光电子学进展, 2019, 56(21): 211505.
CHEN Jingming, JIN Jie, WANG Weifeng. Improved algorithm based on feature pyramid networks[J]. Laser & optoelectronics progress, 2019, 56(21): 211505.

[17] 张涛, 张乐. 一种基于多尺度特征融合的目标检测算法 [J]. 激光与光电子学进展, 2021, 58(2): 0215003.
ZHANG Tao, ZHANG Le. Multiscale feature fusion-based object detection algorithm[J]. Laser & optoelectronics progress, 2021, 58(2): 0215003.

[18] 和超, 张印辉, 何自芬. 多尺度特征融合工件目标语义分割 [J]. 中国图象图形学报, 2020, 25(3): 476–485.
HE Chao, ZHANG Yinhui, HE Zifen. Semantic segmentation of workpiece target based on multiscale feature fusion[J]. Journal of image and graphics, 2020, 25(3): 476–485.

[19] 鞠默然, 罗江宁, 王仲博, 等. 融合注意力机制的多尺

度目标检测算法 [J]. 光学学报, 2020, 40(13): 1315002.
JU Moran, LUO Jiangning, WANG Zhongbo, et al. Multi-scale target detection algorithm based on attention mechanism[J]. Acta optica sinica, 2020, 40(13): 1315002.

[20] 张筱晗, 姚力波, 吕亚飞, 等. 双向特征融合的数据自适应 SAR 图像舰船目标检测模型 [J]. 中国图象图形学报, 2020, 25(9): 1943–1952.
ZHANG Xiaohan, YAO Libo, LV Yafei, et al. Data-adaptive single-shot ship detector with a bidirectional feature fusion module for SAR images[J]. Journal of image and graphics, 2020, 25(9): 1943–1952.

[21] 高杨, 肖迪. 基于多层特征融合的小目标检测算法 [J]. 计算机工程与设计, 2020, 41(7): 1905–1909.
GAO Yang, XIAO Di. Small object detection algorithm based on multi-feature fusion[J]. Computer engineering and design, 2020, 41(7): 1905–1909.

[22] 杨锐, 张宝华, 张艳月, 等. 基于深度特征自适应融合的运动目标跟踪算法 [J]. 激光与光电子学进展, 2020, 57(18): 181501.
YANG Rui, ZHANG Baohua, ZHANG Yanyue, et al. Moving object tracking algorithm based on depth feature adaptive fusion[J]. Laser & optoelectronics progress, 2020, 57(18): 181501.

[23] 赵文清, 程幸福, 赵振兵, 等. 注意力机制和 Faster RCNN 相结合的绝缘子识别 [J]. 智能系统学报, 2020, 15(1): 92–98.
ZHAO Wenqing, CHENG Xingfu, ZHAO Zhenbing, et al. Insulator recognition based on attention mechanism and faster RCNN[J]. CAAI transactions on intelligent systems, 2020, 15(1): 92–98.

[24] 徐诚极, 王晓峰, 杨亚东. Attention-YOLO: 引入注意力机制的 YOLO 检测算法 [J]. 计算机工程与应用, 2019, 55(6): 13–23.
XU Chengji, WANG Xiaofeng, YANG yadong. Attention-YOLO: YOLO detection algorithm that introduces attention mechanism[J]. Computer engineering and applications, 2019, 55(6): 13–23.

[25] 单义, 杨金福, 武随烁, 等. 基于跳跃连接金字塔模型的小目标检测 [J]. 智能系统学报, 2019, 14(6): 1144–1151.
SHAN Yi, YANG Jinfu, WU Suishuo, et al. Skip feature pyramid network with a global receptive field for small object detection[J]. CAAI transactions on intelligent systems, 2019, 14(6): 1144–1151.

作者简介:

赵文清, 教授, 博士, 主要研究方向为人工智能与图像处理。获河北省科技进步二等奖、三等奖各 1 项。发表学术论文 50 余篇。

杨盼盼, 硕士研究生, 主要研究方向为深度学习和目标检测。