# 一种改进的单步多框目标检测算法

王燕妮[1]，刘祥[1]，刘江[2]
（1. 西安建筑科技大学信息与控制工程学院，710055，西安；2. 西安现代控制技术研究所，710065，西安）

**摘要**：针对单步多框目标检测算法(SSD)中存在的误检、漏检以及检测精度不够高等问题，提出了一种改进的 SSD 目标检测算法。该算法通过空洞卷积替换 conv4_3 卷积层及之前的两次标准卷积，扩大感受野，使用反卷积对不同尺度的特征图进行融合，使融合形成的特征图具有丰富的上下文信息，最后为特征图添加注意力模型，有效提取感兴趣区域的特征。仿真实验结果表明，改进算法在 VOC2007 数据集上较原算法检测精度提升 0.9%，检测结果更加准确，一定程度上改善了误检、漏检等问题，同时仍满足实时性的要求。

**关键词**：目标检测；单步多框目标检测算法；空洞卷积；反卷积；注意力机制
**中图分类号**：TP391 **文献标志码**：A
**DOI**：10.7652/xjtuxb202104016 **文章编号**：0253-987X(2021)04-0145-09

OSID 码

# An Improved Single Shot MultiBox Detector

WANG Yanni[1], LIU Xiang[1], LIU Jiang[2]
(1. School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China;
2. Xi'an Modern Control Technology Research Institute, Xi'an 710065, China)

**Abstract**: Aiming at the problems of false detection, missing detection and low detection accuracy in single shot multibox detector (SSD), an improved SSD object detection algorithm is proposed. The conv4_3 convolution layer and the previous two standard convolutions are replaced by dilated convolution to expand the receptive field. The deconvolution is used to fuse the feature maps of different scales, so that the feature maps formed by fusion contain rich context information. The attention model is added to the feature map to effectively extract the features of regions of interest. Simulation results show that the detection accuracy of the improved algorithm is 0.9% higher than that of the original algorithm, and the detection effect is better. To some extent, it solves the problems of false detection and missing detection, and still meets the requirements of real-time.

**Keywords**: object detection; single shot multibox detector; dilated convolution; deconvolution; attention mechanism

目标检测的任务是找出图像中的感兴趣目标，确定它们的类别和位置，是计算机视觉领域的核心问题之一。由于不同图像中待检测目标的类别、外观、数量、尺度、位置等不同，且存在遮挡等干扰因素，使目标检测在机器视觉领域成为最具挑战性的任务。目标检测在红外探测技术、智能视频监控、遥感影像目标检测、医疗诊断[1-4]以及智能建筑中的火灾、烟雾检测中都有广泛应用。

收稿日期：2020-10-27。 作者简介：王燕妮(1975—)，女，副教授，硕士生导师；刘祥(通信作者)，男，硕士生。 基金项目：陕西省自然科学基础研究资助项目(2020JM-499，2020JQ-684)。
网络出版时间：2020-12-30 网络出版地址：http://kns.cnki.net/kcms/detail/61.1069.T.20201230.0927.002.html

目标检测算法可以分为传统目标检测算法和基于深度学习的目标检测算法，传统目标检测算法使用手工设计特征来检测，首先在图像上选择候选区域，对这些区域进行特征提取，最后使用分类器进行分类。代表算法有尺度不变特征变换算法(SIFT)[5]和维奥拉-琼斯算法(V-J)[6]等，但该类算法时间复杂度高且鲁棒性较差。

基于深度学习的目标检测算法，根据是否有产生候选区域的机制分为两阶段目标检测算法和单阶段目标检测算法。两阶段目标检测算法有基于区域的卷积神经网络(R-CNN)[7]算法、快速基于区域的卷积神经网络(Fast R-CNN)[8]算法和更快的基于区域的卷积神经网络(Faster R-CNN)[9]算法等。R-CNN 算法通过选择性搜索算法对输入图像提取候选区域，利用卷积神经网络提取特征，最后通过支持向量机 SVM[10]进行分类；Fast R-CNN 在 R-CNN 的基础上，加入了感兴趣区域，通过多任务损失函数，实现端到端的训练；Faster R-CNN 采用候选区域生成网络(RPN)代替了选择性搜索算法，提高了算法的速度和准确率。单阶段目标检测算法利用回归的思想，直接在输入图像上回归出目标的类别及边框。单阶段目标检测算法有快速目标检测算法(YOLO)[11]和 SSD[12]：YOLO 采用全图信息进行预测，极大地提高了检测速率；SSD 算法借鉴了 Faster R-CNN 中的思想，在提高检测速度的同时也提高了检测精度。

目标检测算法中经常会存在误检、漏检等问题，因此本文以 SSD 算法为基础，对其进行改进，以提升检测效果，可以较好地改善误检、漏检情况。

# 1 SSD 算法

## 1.1 SSD 网络结构

SSD 目标检测算法使用视觉几何群网络 VGG-16[13]作为基础网络，并且将网络中原本的全连接层 fc6 和 fc7 转换为 3×3 的卷积层，将分类层替换为 4 组卷积层，用来获取更多特征图用于检测，网络结构如图 1 所示。由图 1 可以看出，SSD 采用多尺度特征图进行检测，其中大尺度特征图用于检测小目标，小尺度特征图用于检测中、大目标，这种检测策略可以提高识别的准确度，提高了对不同物体尺寸变化的泛化能力。

## 1.2 SSD 候选框

如 1.1 节中所述，SSD 采用多尺度特征图进行检测，分别选取 conv4_3、conv7、conv8_2、conv9_2、conv10_2、conv11_2 共 6 个特征图。SSD 借鉴了 Faster R-CNN 中锚点框的思想，在每个特征图上设置不同尺寸、长宽比的候选框，候选框的尺寸计算公式如下

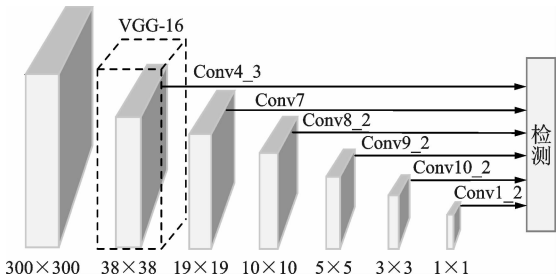



图 1 SSD 网络结构图

Fig. 1 SSD network structure

$$s_n = s_{\min} + \frac{s_{\max} - s_{\min}}{e-1}(n-1),\ n \in [1,e] \quad (1)$$

式中：$s_n$ 表示候选框与图片的比例；$s_{\max}$和 $s_{\min}$代表比例的最大值和最小值，分别取值为 0.9 和 0.2；$e$ 代表特征图的个数。

候选框的长宽比 $\alpha_r$ 的取值范围为 $\alpha_r \in \left\{1,2,3,\frac{1}{2},\frac{1}{3}\right\}$，候选框的宽度和高度计算公式如下

$$w_n^a = s_n\sqrt{\alpha_r};\ h_n^a = s_n/\sqrt{\alpha_r} \quad (2)$$

当候选框的宽高比为 1 时，会增加一个尺度为 $s'_n = \sqrt{s_n s_{n+1}}$的候选框，因此共设置 6 种候选框，其中 conv4_3、conv10_2、conv11_2 仅使用长宽比为 $\alpha_r \in \left\{1,1,2,\frac{1}{2}\right\}$的 4 种候选框。

## 1.3 SSD 算法损失函数

SSD 算法中的损失函数定义为位置误差 $L_{\text{loc}}$和置信度误差 $L_{\text{conf}}$的加权和，计算公式为

$$L(x,c,l,g) = \frac{1}{N}(L_{\text{conf}}(x,c) + \alpha L_{\text{loc}}(x,l,g)) \quad (3)$$

式中：$N$ 为匹配的候选框的数量；$c$ 为类别置信度预测值；$g$ 为真实框的位置参数；$l$ 为预测框的位置预测值；$\alpha$ 权重系数设置为 1；$x_{ij}^z \in \{1,0\}$表示候选框是否与真实框匹配，$x_{ij}^Z = 1$ 表示第 $i$ 个候选框与第 $j$ 个真实框匹配，且类别为 $Z$。

采用 Smooth Ll loss[14]计算 SSD 算法中的位置误差，对候选框的中心$(x_c, y_c)$及宽度$(w)$、高度$(h)$的偏移量进行回归，公式如下

$$L_{\text{loc}}(x,l,g) = \sum_{i\in P_{\text{os}}}^{N}\sum_{m\in\{x_c,y_c,w,h\}} x_{ij}^K \text{smooth}_{\text{L1}}(l_i^m - \hat{g}_j^m) \quad (4)$$

$$\hat{g}_j=(g_j^{x_c}-r_i^{x_c})/r_i^w;\ \hat{g}_j^{y_c}=(g_j^{y_c}-r_i^{y_c})/r_i^h \tag{5}$$

$$\hat{g}_j^w=\log\left(\frac{g_j^w}{r_i^w}\right);\ \hat{g}_j^h=\log\left(\frac{g_j^h}{r_i^h}\right) \tag{6}$$

式中：$\hat{g}$ 为真实框的编码值；$r$ 为候选框，通过对正样本集 $P_{os}$ 进行回归得到最终位置误差。

SSD 算法中的置信度损失，使用 softmax loss 表示，$N_{eg}$ 代表负样本集，公式如下

$$\left.\begin{aligned}L_{conf}(x,c)&=-\sum_{i\in P_{os}}^{N}x_{ij}^P\log(\hat{c}_i^P)-\sum_{i\in N_{eg}}\log(\hat{c}_i^0)\\ \hat{c}_i^P&=\frac{\exp(c_i^P)}{\sum_P\exp(c_i^P)}\end{aligned}\right\} \tag{7}$$

## 2　改进 SSD 算法

为改善 SSD 算法中存在的误检漏检等问题，本文结合空洞卷积[15]、反卷积[16]和注意力机制，提出一种改进的 SSD 目标检测算法。算法首先使用连续 3 次空洞卷积替代原本 conv4_3 卷积层及之前的 2 次卷积操作；然后结合反卷积的思想，对基础网络提取出的不同尺度的特征图进行特征融合；最后在每个用于检测的特征图后添加注意力模型，以达到提高检测能力的效果。

### 2.1　空洞卷积和反卷积

空洞卷积也称为膨胀卷积，是一种改进的图像卷积方法，在标准卷积的基础上注入空洞，以此扩大感受野。在空洞卷积中，增加了一个扩张率参数 $d$，$(d-1)$ 为卷积核之间填充的空洞数量，扩张率越大，卷积核之间的空洞数量越多，感受野越大，通过调整扩张率参数控制空洞卷积中空洞填充的大小。扩张率参数为 1 时，即为标准卷积。标准卷积和空洞卷积示意图如图 2 所示。

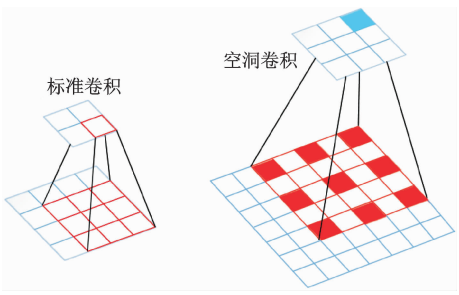



图 2　标准卷积和空洞卷积操作示意图

Fig. 2　Operation diagram of convolution and dilated convolution

加入空洞卷积后卷积核尺寸计算公式为

$$q=k+(k-1)*(d-1) \tag{8}$$

式中：$q$ 为加入空洞卷积后的卷积核尺寸；$k$ 为加入空洞卷积之前的卷积核尺寸。

加入空洞卷积后特征图尺寸的计算公式为

$$o=\left[\frac{a+2p-k-(k-1)*(d-1)}{s}\right]+1 \tag{9}$$

式中：$o$ 为加入空洞卷积后的特征图尺寸；$a$ 为输入特征图尺寸；$s$ 为步长；$p$ 为填充。

反卷积也称为转置卷积，是标准卷积操作的逆运算，用来对特征图尺寸进行还原，解决经过一系列卷积池化操作等运算后特征图分辨率变小的问题，扩大感受野。标准卷积和反卷积的操作示意图如下。

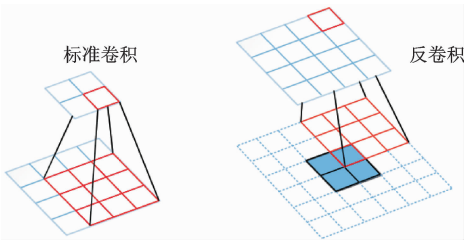



图 3　标准卷积和反卷积操作示意图

Fig. 3　Operation diagram of convolution and deconvolution

使用反卷积输出特征图尺度的计算公式如下

$$O=s(a-1)+k-2p \tag{10}$$

式中：$O$ 为反卷积输出特征图尺寸。

### 2.2　注意力机制

计算机视觉借鉴了人类视觉系统中的注意力机制，对一幅图像中的感兴趣区域进行重点关注，抑制其他干扰区域带来的影响。在深度学习中，通过对特征图添加掩码来实现注意力机制，就是对特征图添加权重，将感兴趣区域的特征标识出来，通过神经网络的训练，让网络学到每一幅图像中需要重点关注的感兴趣区域，这样就形成了深度学习中的注意力，以此提高神经网络处理信息的能力。从注意力域的角度，可以分为空间域、通道域和混合域。

空间域将图像中的空间信息变换到另一个空间中并保留了关键信息。Jaderberg 等提出空间变换网络的网络模型，模型结构如图 4 所示[17]，$H$ 为特征图的高，$W$ 为特征图宽，$C$ 为通道数。对图像中的空间域信息做对应的空间变换，提取特征图中的关键信息。在卷积操作后，不同卷积核产生的通道

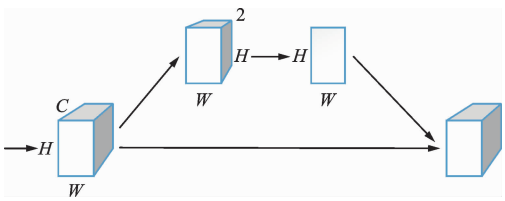



图 4　空间域模型结构图

Fig. 4　Spatial transform model structure

信息所包含的信息量及重要程度不同，故使用同样的变换器解释性不强，因而一般用于图像输入层之后。

通道域主要分布于通道中，表现在图像上就是对不同图像通道的关注程度不同。Hu 等提出通道域注意力机制[18]结构如图 5 所示。图中 $X$ 为输入特征，通过卷积操作生成新的特征 $U$，通过全局平均池化操作将其挤压成通道数为 $C$ 的一维向量，通过激励函数获得权重，最后对特征图上每个像素点加权。

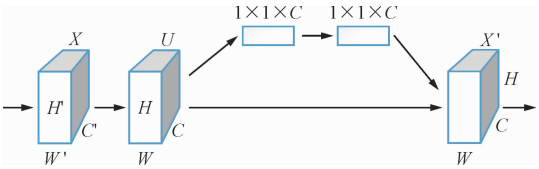



图 5 通道域模型结构图

Fig. 5 Squeeze-and-excitation networks model structure

混合域是结合空间域和通道域的注意力机制。Fei 等提出了基于注意力机制的残差学习方式[19]，其添加的掩码借鉴了残差网络的思想，不只根据当前网络层的信息添加掩码，还添加了上一层的信息，防止了掩码后信息量过少引起的网络层数不能堆叠很深的问题。这可以看做每一个特征元素的权重，通过给每个特征元素找到其对应的注意力权重，就同时形成了空间域和通道域的注意力机制。

### 2.3 改进 SSD 算法

在 SSD 目标检测算法中，采用多尺度特征图的方法对目标进行检测，其中底层特征图对应细节信息，高层特征图对应抽象的语义信息。底层特征图上的小目标物体在经过一系列卷积、池化等操作后，获得的高层特征图上保留的信息将变得更少，所以对小目标物体的检测更不敏感。因此，在 SSD 算法中，底层特征图用于对小目标进行检测，高层特征图用于对中、大目标进行检测。

感受野是指在卷积神经网络中，每一层特征图像素在输入图像上映射区域的大小。在 SSD 算法中，选择 conv4_3、conv7、conv8_2、conv9_2、conv10_2、conv11_2 这 6 个特征图用于检测，conv4_3 为底层特征图，包含的细节信息丰富，但语义信息较少，同时感受野较小；其他 5 个为高层特征图，包含的细节信息较少，语义信息丰富，感受野较大。只有 conv4_3 一个底层特征图用于对小目标进行检测，细节信息不足且没有利用到高层的语义信息，导致 SSD 算法对小目标物体检测效果较差。因此，引入空洞卷积，通过对扩张率进行调节，使具有相同尺寸的卷积核感知到更广泛的感受野，防止局部信息丢失，挖掘多尺度信息。

如 1.1 节中所述，SSD 算法采用 VGG-16 为基础网络进行特征图提取，通过叠加使用 3 次 3×3 卷积核替代 7×7 卷积核，有效减少了参数量，保证了检测算法的实时性。在感受野相同的情况下，引入了更多层的非线性函数，一定程度上加深了网络深度，使网络学到更复杂的特征，提升检测效果。为了增大卷积核的感受野，更容易获得待检测物体的全局特征，本文使用 3 次扩张率 $d=1,2,4$ 的空洞卷积，替换原本 conv4_3 卷积层及之前的 2 次卷积核尺寸为 3×3 的标准卷积，以达到扩大感受野的作用。引入空洞卷积结构与原始标准卷积结构对比如图 6 所示，空洞卷积感受野计算公式如下

$$v = 2*(d-1)*(k-1)+k \tag{11}$$

式中：$v$ 为空洞卷积感受野范围。

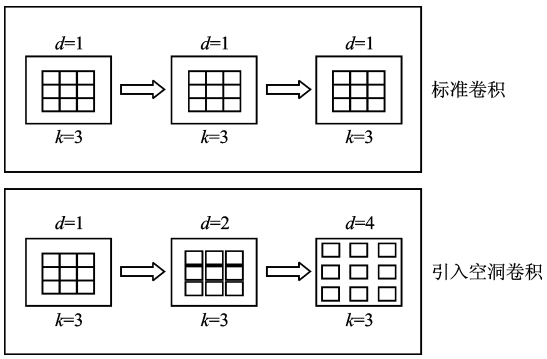



图 6 引入空洞卷积结构与原始标准卷积结构对比

Fig. 6 Comparison between dilated convolution structure and original standard convolution structure

通过叠加使用 3 次空洞卷积，使得每次卷积的输出都包含较大范围的信息，特征图的感受野扩大至 15×15，而原算法中叠加使用 3 次 3×3 标准卷积的感受野为 7×7，空洞卷积和标准卷积感受野范围对比如图 7 所示。同时，将 3 次标准卷积替换为 3 次空洞卷积后的特征图尺度和参数量对比如表 1

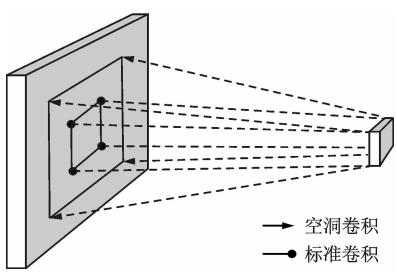



图 7 感受野对比图

Fig. 7 Contrast map of receptive field

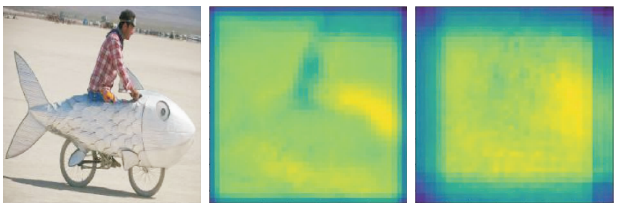
所示，可以看出，特征图尺度没有变化，参数量也没有增加，没有给网络增加额外的计算量。

表 1　特征图尺度和参数值的对比

Table 1　Parameter value comparison of different feature maps

| 层级名称 | 特征图尺度 | 数量 |
|---|---|---|
| Conv4_1 | 38×38 | 1 180 160 |
| Conv4_2 | 38×38 | 2 359 808 |
| Conv4_3 | 38×38 | 2 359 808 |
| Dila_conv4_1 | 38×38 | 1 180 160 |
| Dila_conv4_2 | 38×38 | 2 359 808 |
| Dila_conv4_3 | 38×38 | 2 359 808 |

图 8a 表示算法输入图像，图 8b 表示原算法进行标准卷积后的 conv4_3 特征图输出，图 8c 表示由空洞卷积替代标准卷积后的 conv4_3 特征图输出。对比图 8b 和图 8c，在使用空洞卷积替换原标准卷积后，扩大了输出的特征图感受野，但是细节信息较之前有所下降，可能对检测结果产生不利影响，故需要引入反卷积和注意力机制。通过反卷积将空洞卷积输出的底层特征图和高层特征图进行特征融合，可以同时具有较丰富的细节信息和语义信息；同时通过注意力机制，使网络对图像中复杂的信息进行再次整合。



(a)输入图像　(b)标准卷积后的特征图输出　(c)空洞卷积后的特征图输出

图 8　特征图对比

Fig. 8　Comparison of feature maps

在进行特征融合时，首先对 conv4_3 进行扩张率为 2 的空洞卷积以扩大感受野；然后对具有丰富语义信息的特征图 conv9_2 进行反卷积操作，对其进行特征图尺寸的还原，扩大感受野；最后通过元素相加的方式进行融合，直接对 2 个特征图进行相加，没有改变通道数，节省了计算量。通过融合将高层特征图中的语义信息映射到用于检测小目标物体的底层特征图中，相比融合前包含了更丰富的上下文信息，对小目标物体的检测更敏感，同时也增强了网络的辨识能力。

本文引入通道注意力机制，分为挤压、激励和注意 3 个步骤。

挤压操作的公式如下

$$Y=\frac{1}{HW}\sum_{I=1}^{H}\sum_{J=1}^{W}U(I,J) \tag{12}$$

式中：$H$、$W$ 分别为输入的高度、宽度；$U$ 为输入特征图；$Y$ 为挤压操作的输出；$C$ 为输入的通道数。式(12)将 $HWC$ 的输入转化为 $1\times1\times C$ 的输出，相当于进行了一次全局平均池化操作。

激励操作的公式如下

$$S=\text{h-Swish}(W_2\ \text{ReLU6}(W_1Y)) \tag{13}$$

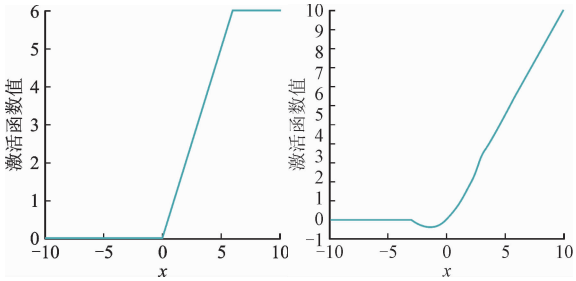
式中：$S$ 为激励操作的输出；$W_1$ 的维度为 $C/\beta\times C$，$W_2$ 的维度为 $C\times C/\beta$，$\beta$ 为一个缩放参数，本文取值为 4。$W_1$ 与 $Y$ 相乘代表全连接操作，然后经过 ReLU6 激活函数；再与 $W_2$ 相乘，也代表一次全连接操作，最后再经过 h-Swish 激活函数，就完成了激励操作。ReLU6 和 h-Swish 激活函数的图像如图 9 所示，公式如下

$$\left.\begin{aligned}&\text{ReLU6}=\min(6,\max(0,x))\\&\text{h-Swish}(x)=x\frac{\text{ReLU6}(x+3)}{6}\end{aligned}\right\} \tag{14}$$

注意操作的公式如下

$$X=SU \tag{15}$$

式中：$X$ 为添加注意力机制后的特征图。通过该方法为融合后的特征图和其他 5 个特征图分别添加注意力模型，为特征图通道上的信号添加权重。这个权重代表与感兴趣区域的相关度，使网络投入更多关注对感兴趣区域进行学习，忽视不重要的区域，使训练的模型学习能力提升，能更充分地学习到待测物体的多种特征信息，以提升检测效果。本文提出的改进 SSD 算法网络结构如图 10 所示。




(a)ReLU6 函数图像　(b)h-Swish 函数图像

图 9　激活函数图像

Fig. 9　Activation function images

## 3　实验结果

### 3.1　实验环境

本文的实验平台为 i7-8700 处理器，NVIDIA

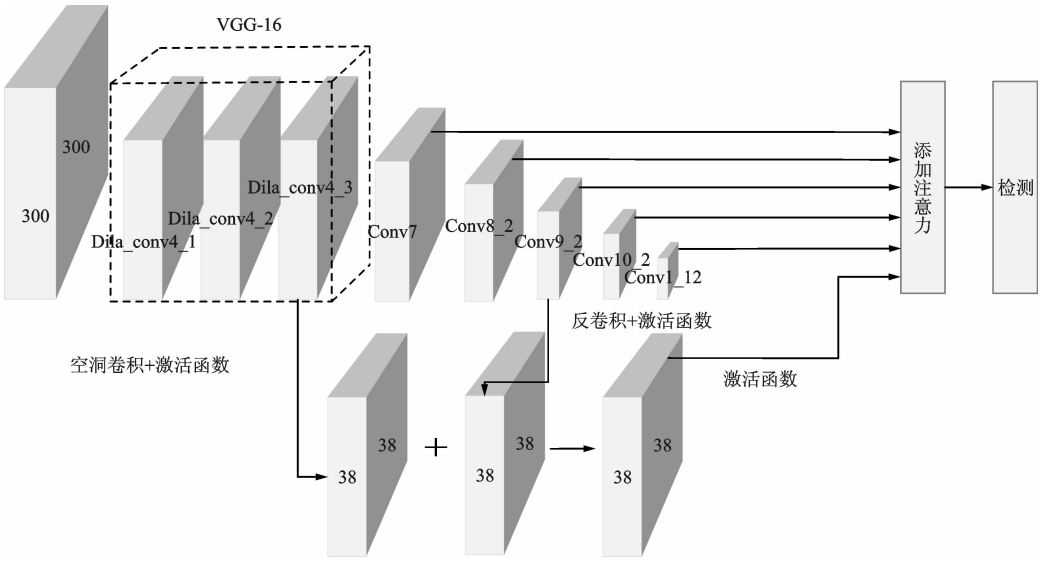



图 10 改进 SSD 算法网络结构图

Fig. 10 Improved algorithm network structure diagram

GeForce TX1070Ti 显卡，基于深度学习框架 Keras。使用 PASCAL VOC 数据集，包含 20 个类别，分别为人、鸟、猫、牛、狗、马、羊、飞机、自行车、船、巴士、汽车、摩托车、火车、瓶子、椅子、餐桌、盆栽植物、沙发、电视。本文使用 VOC2007 和 VOC2012 作为训练集，使用 VOC2007 作为测试集。训练时采用随机梯度下降法(SGD)，批量设置为 32，初始学习率设置为 0.001，动量参数设置为 0.9，学习率在迭代次数为 100 000 和 150 000 时调小 90%，共训练 200 000 次。

### 3.2 检测效果客观评价

本文使用平均检测精度(mAP)作为检测算法的评价指标。检测到的每一个类别都会得到由查准率和查全率构成的曲线，曲线下的面积就是单个类别的检测精度(AP)，对所有类别的检测精度求平均，即可得到平均检测精度。

以 VOC2007 和 VOC2012 数据集为训练集，VOC2007 为测试集，将本文算法与原始 SSD 算法、基于反卷积的单步多框目标检测 DSSD 算法[20]进行对比，对比结果如表 2 所示，将数据可视化后的对比情况如图 11 所示。

此外，为检验算法的实时性，本文对比了其他主流模型及本文提出的改进算法的检测速度，对比结果如表 3 和图 12 所示。

从表 2 和表 3 可以看出，本文提出的改进 SSD 算法与其他主流模型相比，检测精度均有一定提升；相比原始 SSD 算法检测精度提升了 0.9%，检测速

表 2 本文算法与其他算法检测结果对比

Table 2 Comparison of test results between the proposed algorithm and other algorithms

| 算法 | 基础网络 | 平均检测精度/% | 检测精度/% | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 飞机 | 自行车 | 鸟 | 船 | 瓶子 | 巴士 | 汽车 | 猫 | 椅子 | 牛 |
| SSD300 | VGG-16 | 74.3 | 75.5 | 80.2 | 72.3 | 66.3 | 47.6 | 83.0 | 84.2 | 86.1 | 54.7 | 78.3 |
| SSD300* | VGG-16 | 77.2 | 78.8 | 85.3 | 75.7 | 71.5 | 49.1 | 85.7 | 86.4 | 87.8 | 60.6 | 82.7 |
| DSSD321 | Resnet-101 | 78.6 | 81.9 | 84.9 | 80.5 | 68.4 | 53.9 | 85.6 | 86.2 | 88.9 | 61.1 | 83.5 |
| 本文算法 | VGG-16 | 78.1 | 80.4 | 85.9 | 77.3 | 72.6 | 49.5 | 84.9 | 87.1 | 87.6 | 62.6 | 81.2 |

| 算法 | 基础网络 | 平均检测精度/% | 检测精度/% | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 餐桌 | 狗 | 马 | 摩托车 | 人 | 盆栽 | 羊 | 沙发 | 火车 | 电视 |
| SSD300 | VGG-16 | 74.3 | 73.9 | 84.5 | 85.3 | 82.6 | 76.2 | 48.6 | 73.9 | 76.0 | 83.4 | 74.0 |
| SSD300* | VGG-16 | 77.2 | 76.5 | 84.9 | 86.7 | 84.0 | 79.2 | 51.3 | 77.5 | 78.7 | 86.7 | 76.3 |
| DSSD321 | Resnet-101 | 78.6 | 78.7 | 86.7 | 88.7 | 86.7 | 79.7 | 51.7 | 78.0 | 80.9 | 87.2 | 79.4 |
| 本文算法 | VGG-16 | 78.1 | 76.3 | 84.2 | 86.6 | 86.2 | 82.2 | 56.2 | 79.5 | 77.4 | 88.8 | 76.5 |

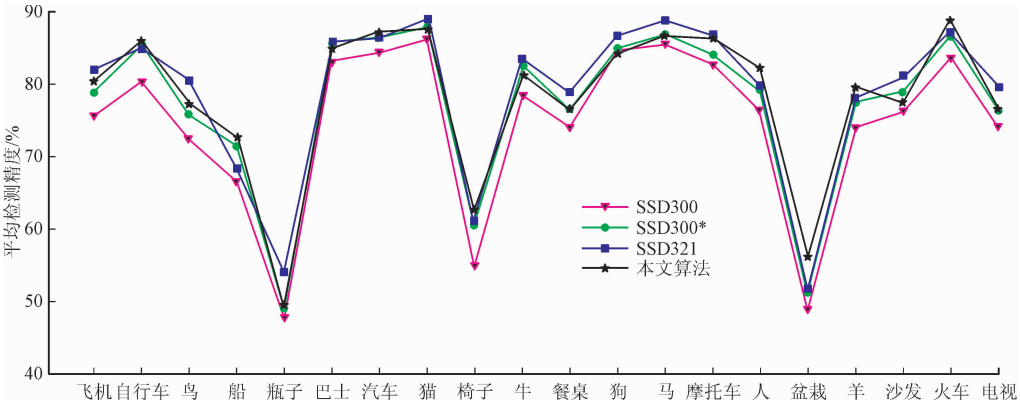



图 11　VOC2007 测试集上各类别平均检测精度对比

Fig. 11　mAP comparison of classes on VOC2007 test set

度略低于 SSD 算法，相比 DSSD 算法检测精度略有下降，但检测速度有较大提升。因此，本文提出的算法在提高检测精度的同时，满足了实时性的需求。

表 3　检测速度对比

Table 3　Comparison of detection speed

| 算法 | 基础网络 | 平均检测精度/% | 检测速度/(帧·$s^{-1}$) |
|---|---|---|---|
| Faster R-CNN | VGG-16 | 73.2 | 7.0 |
| Faster R-CNN* | Resnet-101 | 76.4 | 2.4 |
| YOLO | GoogleNet | 63.4 | 45.0 |
| SSD300 | VGG-16 | 74.3 | 46.0 |
| SSD300* | VGG-16 | 77.2 | 46.0 |
| DSSD321 | Resnet-101 | 78.6 | 9.5 |
| 本文模型 | VGG-16 | 78.1 | 39.0 |

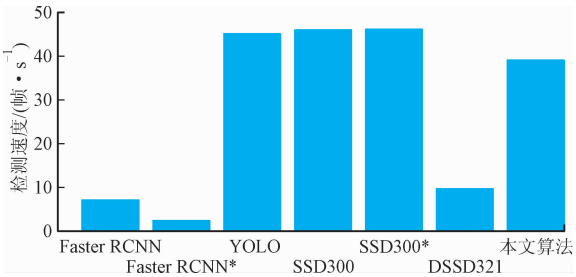



图 12　VOC2007 测试集上检测速度对比图

Fig. 12　Detection speed comparison on VOC2007 test set

### 3.3　检测效果主观评价

为了更直观地感受本文提出的改进算法检测能力，在 VOC2007 测试集中选取若干图片，分别用原始 SSD 算法和改进后的算法进行检测并对比，限于篇幅限制，本节共选取 6 张图片进行对比，对比结果如图 13～图 18。

从图 13～16 中可以看出，改进后的算法相比原始 SSD 算法可以检测到更多目标，有效改善了漏检的问题。如图 17 所示，改进后的算法检测出更多目标的同时对目标的分类也更加准确，改善了误检的问题。如图 18 所示，在待检测目标发生遮挡现象时，改进后的算法检测更加准确，仍可以检测出重叠目标的类别。

(a)SSD 检测结果　　(b)改进算法检测结果

图 13　两种算法的瓶子检测结果对比

Fig. 13　Comparison of bottles detection results between the SSD and proposed algorithms

(a)SSD 检测结果　　(b)改进算法检测结果

图 14　两种算法的飞机检测结果对比

Fig. 14　Comparison of planes detection results between the SSD and proposed algorithms

(a)SSD 检测结果　　(b)改进算法检测结果

图 15　两种算法的船检测结果对比

Fig. 15　Comparison of boats detection results between the SSD and proposed algorithms

(a)SSD 检测结果 (b)改进算法检测结果

图 16 两种算法的盆栽检测结果对比

Fig. 16 Comparison of pot plants detection results between the SSD and proposed algorithms

(a)SSD 检测结果 (b)改进算法检测结果

图 17 两种算法的牛检测结果对比

Fig. 17 Comparison of cattle detection results between the SSD and proposed algorithms

(a)SSD 检测结果 (b)改进算法检测结果

图 18 两种算法的摩托车检测结果对比

Fig. 18 Comparison of motorcycles detection results between the SSD and proposed algorithms

## 4 结 论

本文提出一种改进的 SSD 目标检测算法，通过连续 3 次空洞卷积替换原本 conv4_3 卷积层及之前的两次标准卷积卷积核为 3×3，以扩大特征图的感受野；结合反卷积，对不同尺度的特征图进行特征融合，使用于检测的特征图同时具有丰富的细节信息和上下文信息；最后结合注意力机制，对多尺度特征图添加注意力模块，提升算法对感兴趣区域的检测能力。通过理论分析和实验对比，改进后的算法相比原始 SSD 算法检测精度有明显的提升，有效改善了误检和漏检等问题。由于引入了额外计算量，导致检测速度有所下降，但仍能满足实时性需求。从对比图中也可以看出，改进后的算法仍存在一定漏检现象，在接下来的研究中，将对网络结构和损失函数部分进行优化，进一步提高算法各方面的性能。

## 参考文献：


[1] 王好贤，董衡，周志权. 红外单帧图像弱小目标检测技术综述 [J]. 激光与光电子学进展，2019，56(8)：1-14.
WANG Haoxian，DONG Heng，ZHOU Zhiquan. Review on dim small target detection technologies in infrared single frame images [J]. Laser & Optoelectronics Progress，2019，56(8)：1-14.

[2] 黄凯奇，陈晓棠，康运锋，等. 智能视频监控技术综述 [J]. 计算机学报，2015，38(6)：1093-1118.
HUANG Kaiqi，CHEN Xiaotang，KANG Yunfeng，et al. Intelligent visual surveillance：a review [J]. Chinese Journal of Computers，2015，38(6)：1093-1118.

[3] 欧攀，张正，路奎，等. 基于卷积神经网络的遥感图像目标检测 [J]. 激光与光电子学进展，2019，56(5)：051002.
OU Pan，ZHANG Zheng，LU Kui，et al. Object detection of remote sensing images based on convolutional neural networks [J]. Laser & Optoelectronics Progress，2019，56(5)：051002.

[4] 杨洁，陈灵娜，陈宇韶，等. 基于深度学习的医学目标检测与识别 [J]. 信息技术，2018，42(10)：81-87.
YANG Jie，CHEN Lingna，CHEN Yushao，et al. Target detection and recognition based on depth learning [J]. Information Technology，2018，42(10)：81-87.

[5] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision，2004，60(2)：91-110.

[6] VIOLA P，JONES M J. Robust real-time face detection [J]. International Journal of Computer Vision，2004，57(2)：137-154.

[7] GIRSHICK R，DONAHUE J，DARRELL T，et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ，USA：IEEE，2014：580-587.

[8] GIRSHICK R. Fast R-CNN [C]// Proceedings of 2015 IEEE International Conference on Computer Vision. Piscataway，NJ，USA：IEEE，2015：1440-1448.

[9] 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.

[10] LIN C F，WANG S D. Fuzzy support vector machines [J]. IEEE Transactions on Neural Networks，2002，

13(2)：464-471.

[11] 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. Piscataway，NJ，USA：IEEE，2016：779-788.

[12] LIU W，ANGUELOV D，ERHAN D，et al. SSD：single shot MultiBox detector [C]// Proceedings of the 14th European Conference on Computer Vision. Berlin，Germany：Springer，2016：21-37.

[13] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-09-26]. http：// arXiv. org/abs/1409. 1556v6.

[14] 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. Piscataway，NJ，USA：IEEE Computer Society，2014：580-587.

[15] YU F，KOLTUN V. Multi-scale context aggregation by dilated convolutions [EB/OL]. [2020-10-01]. https：// arXiv. org/abs/1511. 07122.

[16] PENG L Y，ZHANG L，ZHANG Y，et al. Deep deconvolution neural network for image super-resolution [J]. Journal of Software，2018，29(4)：927-934.

[17] JADERBERG M，SIMONYAN K，ZISSERMAN A，et al. Spatial transformer networks [C]// Advances in Neural Information Processing Systems. Montreal，Canada：[s. n.]，2015：2008-2016.

[18] HU Jie，SHEN Li，SUN Gang. Squeeze-and-excitation networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway，NJ，USA：IEEE，2018：7132-7141.

[19] WANG Fei，JIANG Mengqing. Residual attention network for image classification [EB/OL]. [2020-09-22]. https：// arXiv. org/abs/1704. 06904

[20] FU Chengyang，LIU Wei，RANGA A，et al. DSSD：deconvolutional single shot detector [EB/OL]. [2020-09-22]https：// arXiv. org/abs/1701. 06659.

（编辑　杜秀杰）