一种双注意力模型引导的目标检测算法

冀中，孔乾坤，王建*
天津大学电气自动化与信息工程学院，天津 300072

摘要 为了解决对小目标物体识别精度较差的问题，提出了一种双注意力模型引导的目标检测算法。该方法基于单阶段检测算法的实现原理，通过引入两种注意力模型来提升检测性能，尤其是对小目标物体。首先在卷积神经网络中引入一个多尺度特征级联注意力模块，对原始卷积神经网络的特征图中各区域进行不同重要程度的关注，降低特征图的背景及负样本信息的干扰，特别是在浅层特征图中可对小目标物体进行有效的关注。此外，密集连接的方式缓解了网络反向传播过程中梯度消失的问题。其次，对融合后的特征引入显著通道自注意力模块，区分特征图不同通道，筛选出有用的通道信息，使待检测的特征更具表征性。在目标检测基准数据集 COCO 上进行测试，验证了所提方法的有效性和先进性。

关键词 图像处理；目标检测；卷积神经网络；小目标；注意力模型
中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/LOP57.061008

Object Detection Algorithm Guided by Dual Attention Models

Ji Zhong, Kong Qiankun, Wang Jian*
School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, *China*

Abstract In order to solve the problem of inferior recognition accuracy for small objects, a object detection algorithm guided by dual attention models is proposed. The method is based on the realization principle of single-stage detection algorithms, and introduces two attention models to improve the detection performance, especially for small objects. Specifically, a multi-scale feature cascade attention module is first introduced into the convolutional neural network, which weights the importance on different regions of the original convolutional neural network's feature map to reduce the interference of background and negative object information in the feature map, especially highlighting the small objects effectively in the shallow feature map. Besides, dense connection alleviates the problem of gradient disappearance in the process of back propagation. A salient channel self-attention module is introduced for the fused features, which focuses on the difference among different channels of the feature map so as to screen out useful channel information, thus making the feature map to be detected more representative. In addition, the experiments on COCO benchmark dataset of object detection verify the effectiveness and advancement of the proposed method.

Key words image processing; object detection; convolutional neural network; small objects; attention model
OCIS codes 100.3008; 100.4996; 100.2000

1 引 言

目标检测是计算机视觉领域十分重要且具有挑战性的研究课题之一。其要求计算机在一幅含有多目标物体的图像中，对感兴趣的目标物体进行分类，并且通过边界框返回每个目标在图像中的位置。在实际应用中，目标检测在目标侦查、精确制导、智能监控、视觉导航、人机交互、空间遥感及医疗辅助诊断等方面具有非常重要的研究意义与价值。

随着深度学习的发展，卷积神经网络（CNN）能够从大量数据中自动学习图像的高层次特征，从而在计算机视觉任务上取得了重大突破，如目标检测[1-5]、图像分割[6-7]、图像分类[8-11]和图像去雾[12-13]等。当前大多数目标检测器均直接采用在

收稿日期：2019-07-10；**修回日期**：2019-08-18；**录用日期**：2019-08-28
* **E-mail**：jianwang@tju.edu.cn

ImageNet[14]数据集上预训练得到的卷积神经网络，如 ResNet[8]和 VGG[11]等，对图片进行特征提取，并在 PASCAL VOC[15]和 COCO[16]等数据集上取得了较好的检测性能，但主要是针对大尺寸和中尺寸的目标物体具有较好的识别效果，对小尺寸目标物体的识别一直都是具有挑战性的任务，性能仍不甚满意。其原因主要有两点：1）由 CNN 直接提取的图像特征中含有大量背景以及负样本目标物体的信息，会干扰检测器对正样本目标物体的判别；2）随着网络层次的加深，小目标物体由于尺寸小，信息可能会因下采样操作而不断丢失。

近年来，注意力模型[17]借鉴人类视觉注意力机制被广泛应用在计算机视觉任务中并取得了优异的效果。Xu 等[18]根据注意力模型的关注区域选择的不同，将注意力模型分为软注意力模型和硬注意力模型。软注意力模型分配注意力时，对输入的每个区域都给出一个概率，范围为 0～1，然后对其进行加权，会关注特征图的全部位置，只是不同位置的权重不同，因此其是参数化的、可微的。在最近的工作[19-20]中开发的软注意力模型可用于卷积神经网络进行端到端的训练。硬注意力模型指的是在每个时刻模型的序列只有一个取 1，其余全部为 0，即每次只关注一个位置，更有利于关注局部信息，但不可以直接求导，需要采用蒙特卡罗采样的方法估计模块的梯度从而实现反向传播。

本文结合注意力模型的思想改善卷积神经网络的特征，从而提升检测器的性能，尤其是提升了对小目标物体的检测性能。考虑到二阶段检测算法内存占用大、检测速度慢等缺点，提出了一个新的单阶段检测算法，称为双注意力引导模型（DAGM）的目标检测算法，包括多尺度特征级联注意力（MFCA）模块及显著通道自注意力（SCSA）模块。

2 相关工作

当前基于卷积神经网络的目标检测器根据是否生成预选框可划分为两类：一类是二阶段检测算法，如 Faster R-CNN[3]、FPN[21]、CoupleNet[22]和 SIN[23]等；另一类是单阶段检测算法，如 SSD[4]、YOLO[5]、YOLOv2[24]、YOLOv3[25]和 DSSD[26]等。

2014 年，Girshick 等[1]将 CNN 与目标候选区域机制相结合提出 R-CNN，这是基于区域的二阶段检测算法的开山之作；2015 年，Girshick 等[2]设计出感兴趣区域池化策略（RoI Pooling）改进 R-CNN，提出检测速度更快、精度更高的 Fast R-CNN；之后 Ren 等[3]将区域建议网络与 Fast R-CNN 算法相结合，提出了 Faster R-CNN 算法，进一步提升了检测的精度和速度。为了提升小目标物体的检测性能，2017 年 Lin 等[21]提出 FPN，通过自上而下的方式将高层次特征与低层次特征进行融合，构建了一个多尺度特征金字塔，从而大幅度提升了小目标物体的检测性能；Zhu 等[22]在 R-FCN[27]的基础上提出 CoupleNet，通过结合特征的局部、全局和上下文信息，提升了目标检测性能；2018 年，Liu 等[23]提出 SIN，结合场景信息和物体之间的联系，充分利用上下文信息，提出一个结构推理网络，针对目标间的关系对物体的定位发挥了重要作用。

与二阶段目标检测算法不同的是，单阶段目标检测算法是在待检测图像上直接进行采样锚框并预测类别，然后通过回归计算修正锚框的坐标。其中，SSD[4]和 YOLO[5]是具有代表性的单阶段检测算法，与二阶段检测算法相比，虽然检测精度较低，但速度得到显著的提升。由于 SSD[4]仅在网络的较高层使用小尺寸卷积核对多尺度特征图进行独立检测，所以丢失了大量浅层的视觉信息，而浅层的视觉信息对小目标物体的识别来说很重要。为了解决 SSD 检测算法检测小目标物体的难题，2017 年，Fu 等[26]提出 DSSD 算法，将 SSD 检测器中的卷积神经网络从 VGG-16[11]更改为 ResNet-101[8]，然后借鉴 FPN[21]的思想将深层特征图与浅层特征图进行融合，在融合后的特征图上预测目标类别的信息；2018 年，Redmon 等[25]在 YOLOv2[24]的基础上提出 YOLOv3，设计出更深层次的 DarkNet-53 网络结构，并在预测阶段借鉴多尺度 FPN[21]的思想，虽然检测速度略有下降，但对小目标物体的检测更加敏感。

3 本文算法

DAGM 网络结构模型如图 1 所示。该框架与当前大多数单阶段检测算法一样，采用端到端的训练方式，主要包括特征提取单元和特征预测单元。在特征提取单元中，增加 MFCA 模块和 SCSA 模块。其中，MFCA 模块会在网络每一阶段的特征提取过程中产生一个注意力张量，对该阶段特征图的各个区域进行加权求和操作；SCSA 模块对 CNN 各阶段输出的特征图各通道进行重要性区分。通过这两个模块为特征预测单元提供表征性更强、语义信息更加丰富的特征。特征预测单元主要包括目标分类模块和边界框坐标回归模块。实验采用轻量级的

全卷积网络进行运算，具体操作过程为：目标分类模块使用 4 个卷积层，卷积核的尺寸为 3×3×256，最后一层采用 3×3×d 的卷积核，通道数目 d 为 $k\times n$，其中，k 为目标物体的种类数，n 为在每一阶段特征图上预测的锚框数，使用非线性激活层计算各个锚框的预测概率。回归模块与分类模块大致相同，不同点在于 4 个卷积层操作后，最后一层采用一个 3×3×m 的卷积核，通道数目 m 为 $4\times n$。

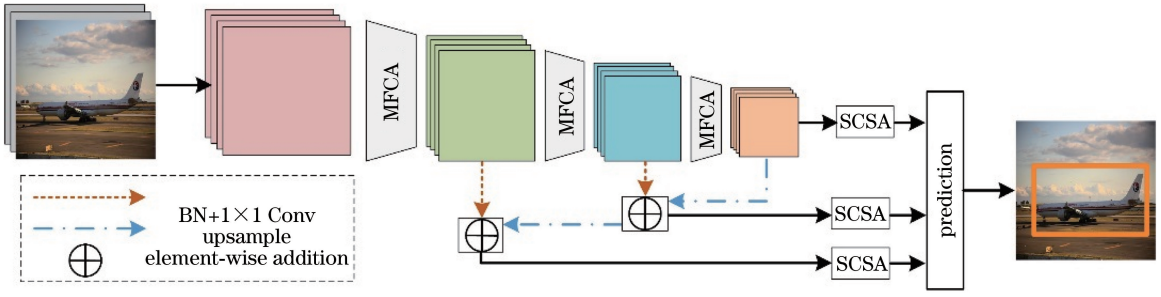

图 1 本文 DAGM 模型结构

Fig. 1 Architecture of the proposed DAGM model

3.1 MFCA 模块

遵循注意力机制思想以及空洞卷积[7]思想，本文方法在 CNN 特征提取阶段引入了 MFCA 模块，在不过度增加模型参数量的前提下，提升网络的特征表达能力。主要操作包括卷积块（Conv block）、空洞块（dilated block）、上采样操作（upsample）以及级联模块（C）等，各部件具体连接方式如图 2 所示。

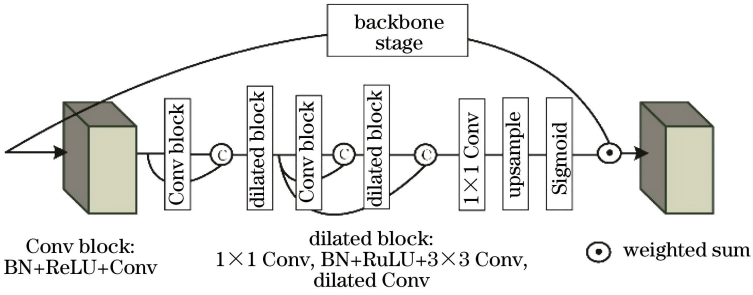

图 2 MFCA 模块结构图

Fig. 2 Architecture of MFCA model

为了减少 MFCA 模块的参数量，该结构通过紧密连接的级联方式使特征图被重复利用，控制模型参数量的同时将各层中的信息最大限度地结合起来。此外，这种紧密连接的结构也规避了网络反向传播过程中梯度消失的风险。该模块计算过程为

$$f[W_k(\boldsymbol{X}),F_k(\boldsymbol{X})]=W_k(\boldsymbol{X})\odot F_k(\boldsymbol{X})=[1+W_k(\boldsymbol{X})]\cdot F_k(\boldsymbol{X}),\tag{1}$$

式中：⊙表示加权和；$\boldsymbol{X}$ 为输入特征图；$F_k(\boldsymbol{X})$ 为卷积神经网络中第 k 阶段提取的特征图，$F_k(\boldsymbol{X})\in\mathbb{R}^{w\times h\times d}$，$\mathbb{R}^{w\times h\times d}$ 表示列为 w、行为 h、通道数为 d 的实数空间；$W_k(\boldsymbol{X})$ 为第 k 阶段特征级联后产生的注意力权值，$W_k(\boldsymbol{X})\in\mathbb{R}^{w\times h\times d}$，取值范围为(0,1)。

该模块可被添加到任意的卷积神经网络中，实验采用的卷积神经网络为 ResNet-101。该网络共包括 5 个阶段，记为{C1，C2，C3，C4，C5}。由于 C1 阶段特征图的空间分辨率较高，考虑到网络模型参数以及运算效率，从 C2 阶段开始加入 MFCA 模块。具体操作流程为：先对 ResNet-101 网络中前一阶段的特征图进行 1×1 小卷积核的卷积运算来降低其空间分辨率，再通过线性整流函数（ReLU）对所得的特征图进行激活操作，主要作用是增强网络的稀疏性，减少网络中参数的相互依赖关系，减少过拟合问题的发生；再通过卷积块和特征图的级联操作，使高层次特征图中保留低层次特征图的信息，同时重用了前几层特征图，减少了网络运算参数；然后对级联后的特征图进行空洞块（dilated block）操作。空洞块的设计主要有两个作用：1）以 1×1×d/2 的卷积操作对级联后的特征图的通道进行降维，使级联后的特征图的通道数目恢复为输入特征图的通道数目；2）采用空洞卷积操作，保证全局信息不丢失的前提下，增大特征图的感受野，获取更丰富的上下文信息。最后，采用双线性插值法扩大所得特征图的空间分辨率，并采用 Sigmoid 函数将特征图上各个区域的值映射到(0,1)之间，得到与该阶段特征图维度

一致的注意力张量，如(1)式所示与输入特征图进行加权求和，从而使得特征图上的每一个区域得到不同重要程度的关注，这有助于削弱背景以及负样本目标的信息干扰。

3.2 SCSA 模块

FPN[21]结构将卷积神经网络提取的低分辨率、强语义信息的高层次特征图与高分辨率、弱语义信息的低层次特征图通过自上而下的方式进行融合，从而使各层次的特征都包含了丰富的语义信息，提高了目标检测的准确性，尤其是对小目标物体的检测。

众所周知，随着卷积神经网络层次的加深，特征图的通道数目也逐渐增多，以 ResNet-101 为例，{C1，C2，C3，C4，C5}5 个阶段特征的通道维度分别是{64，256，512，1024，2048}。FPN 采用 1×1×256 卷积操作对各阶段特征图通道进行通道压缩。但各通道具有的视觉信息与语义信息均不相同，并且包含的目标物体的信息量也不均等，直接用来预测目标物体缺少判别性。例如，部分通道的特征图包含的目标物体信息较弱，背景或者负样本目标物体的信息较强，不利于检测。实验将卷积神经网络 ResNet-101 的 C3 阶段特征图进行可视化处理，如图 3 所示。输入图像的维度为 640×349×3，C3 阶段输出的特征图维度为 160×88×256，可以发现每一个通道的特征图对目标的描述均不相同，差异较大。

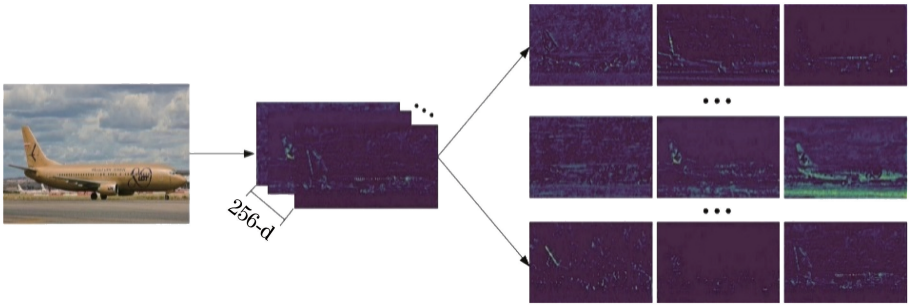

图 3 特征图的可视化

Fig. 3 Visualization of the feature map

直接通过特征金字塔融合操作得到的特征图没有关注到各通道间的区别信息，因此提出一个显著通道自注意力模块，对特征图不同通道采用 0～1 不同的权值用于代表不同通道的重要程度，使特征图能够更充分有效地被利用。该模块具体结构如图 4 所示。

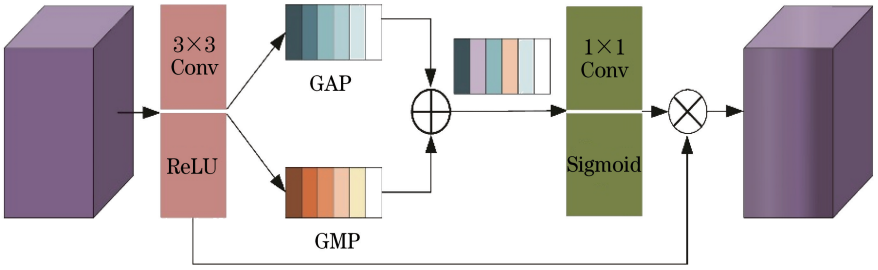

图 4 SCSA 模块原理结构图

Fig. 4 Architecture of SCSA model

SCSA 模块对融合后的特征图先进行 3×3 的卷积操作以及非线性激活操作，然后进行两路的池化操作，包括全局最大池化(GMP)和全局平均池化(GAP)。再对所得向量进行求和运算，最后通过1×1卷积操作以及激活操作，为输入特征图的各通道赋予一个有意义的权重。为了减少计算量，只对 ResNet-101 网络中的 C3，C4，C5 阶段具有强语义信息的特征图进行显著通道注意力操作，根据输入的特征图各通道间的差异性，分配不同的注意力权值 $\boldsymbol{W}_k$：

$$\boldsymbol{W}_k = \mathrm{Sigmoid}(\boldsymbol{W}_{\mathrm{GAP}} + \boldsymbol{W}_{\mathrm{GMP}}), \tag{2}$$

$$\boldsymbol{P}_k = \boldsymbol{S}_k \cdot \boldsymbol{W}_k, \tag{3}$$

式中：$\boldsymbol{S}_k$ 为第 k 阶段输入的特征图张量；$\boldsymbol{W}_k$ 为第 k 阶段通道注意力分布张量；$\boldsymbol{W}_{\mathrm{GAP}} \in \mathbb{R}^{1\times1\times d}$ 和 $\boldsymbol{W}_{\mathrm{GMP}} \in \mathbb{R}^{1\times1\times d}$ 表示特征图经过 3×3 卷积操作及 ReLU 非线性激活操作后分别经过全局平均池化和全局最大池化得到的张量。

4 实验结果与分析

4.1 数据集

为了验证所提模型的有效性，在 COCO 2017 基准数据集上进行实验。COCO 数据集是当前用于评价目标检测算法性能的权威数据集。该数据集中针对目标检测任务，共包括 80 类目标物体。训练集包括 118287 张图片，用于训练模型的参数；验证集包括 5000 张图片，用于验证各模块的有效性；测试集约包括 20000 张图片，用于测试模型的优劣，但需要到官方服务器上进行性能指标的测试。根据目标的像素占比，该数据集将目标分为小目标（$S_{area}<32\times32$，S_{area}为目标面积）、中目标（$32\times32<S_{area}<96\times96$）和大目标（$S_{area}>96\times96$），其数量占比分别为 41%，34%和 25%。针对这三类尺寸目标的检测性能提出了相应的检测指标 AP_S，AP_M 和 AP_L。根据交并比（IoU）的不同取值，还有两个主要的测试指标，为 AP_{50}和 AP_{75}，二者分别表示 IoU 值取 0.50 和 0.75 时所有类别的平均检测精度。此外，最重要的一个指标为 AP，是对 10 个阈值下 80 个目标类别的平均检测精度求平均值，该指标是决定检测算法优劣的关键。

4.2 实验环境及训练细节

实验基于 Keras 深度学习平台，运行在 Ubuntu 14.04 系统环境下，中央处理器为 3.3 GHz Intel Core(TM) CPU i9-7900x，内存 32 GB，显卡型号是 NVIDIA GeForce GTX 1080ti，加速库为 CUDA8.0 和 CUDNN6.0。

训练的主干网络是在 ImageNet 上参数初始化 ResNet-101，采用 Adam 优化器进行优化，指数衰减率 α 取 0.9，β 取 0.999，学习率采用周期性学习率，初始学习率设为 5.0×10^{-6}，最大学习率设为 1.5×10^{-5}。由于计算资源有限，实验中 batch size 设置为 1。

4.3 对比实验

选取当下比较流行的目标检测算法进行对比，包括二阶段检测算法 Faster R-CNN[3] 及其扩展算法[8,28]、CoupleNet[22]、SIN[23]、DeNet[29] 和 MLKP[30]，还有单阶段检测算法如 SSD513[4]、YOLOv2[24]、YOLOv3[25]、DSSD513[26]、RON[31] 和 Relation Network[32]，实验结果均是在 COCO 测试集上所得，对比结果如表 1 所示。

表 1 不同检测算法的检测结果对比

Table 1 Comparison of detection results with different detection algorithms %

Method	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Faster R-CNN+++[8]	ResNet-101-C4	34.9	55.7	37.4	15.6	38.7	50.9
Faster R-CNN (FR) by G-RMI[28]	ResNet-101-C4	34.7	55.5	36.7	13.5	38.1	52.0
YOLOv2[24]	DarkNet-19	21.6	44.0	19.2	5.0	22.4	35.5
SSD513[4]	ResNet-101-SSD	31.2	50.4	33.3	10.2	34.5	49.8
DSSD513[26]	ResNet-101-DSSD	33.2	53.3	35.2	13.0	35.4	51.1
RON[31]	VGG-16	27.4	27.1	49.5	—	—	—
DeNet[29]	DeNet-101	33.8	53.4	36.1	12.3	36.1	50.8
CoupleNet[22]	ResNet-101	33.1	53.5	35.4	11.6	36.3	50.1
YoLov3[25]	DarkNet-53	33.0	57.9	34.4	18.3	35.4	41.9
SIN[23]	VGG-16	23.2	44.5	22.0	7.3	24.5	36.3
Relation Network[32]	ResNet-50	32.5	54.0	33.8	—	—	—
MLKP[30]	ResNet-101	26.9	48.4	26.9	8.6	29.2	41.1
MFCA (ours)	ResNet-101	36.2	54.5	38.7	18.5	39.2	47.6

通过表 1 可得，当前大部分检测算法对小目标物体的检测精度较低；基于 Faster R-CNN 的检测算法[8,28]等二阶段检测算法检测精度比单阶段检测算法的检测精度高。而所提 DAGM 检测算法在平均检测精度方面优于二阶段检测算法的同时，针对小目标物体的检测性能有最好的表现，这证明了所提算法的有效性。

AP 是判定检测算法在 COCO 数据集上性能优劣的决定性指标。上述各算法的 AP 检测结果如图 5 所示，可以更直观地表示各检测算法性能。

4.4 模块有效性验证实验

对所提的两个注意力模块 MFCA 和 SCSA 分别进行实验来验证其有效性，检测结果如表 2 所示。

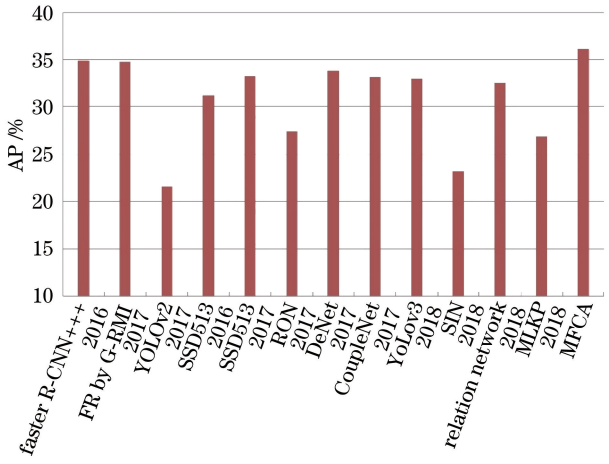

图 5 不同检测算法的检测平均精度

Fig. 5 AP histogram of different detection algorithms

表 2 各模块的检测结果对比

Table 2 Comparison of test results of each module %

Module	Backbone	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
Baseline	ResNet-101	35.0	52.6	37.7	17.1	38.8	48.0
Baseline+MFCA	ResNet-101	35.5	53.3	38.1	18.4	39.2	48.1
Baseline+MFCA+SCSA	ResNet-101	35.9	54.0	38.6	18.8	39.8	48.5

表 2 实验结果均是在 COCO 验证集上进行测试所得。与基准结构相比，加入 MFCA 模块后 AP 提升了 0.5 个百分点，针对小尺寸目标物体的 AP_S 提升了 1.3 个百分点，检测性能提升较为显著；然后在 MFCA 模块的基础上添加 SCSA 模块后，AP 又提升了 0.4 个百分点，并且针对大、中、小三种尺寸的目标物体的检测精度均有提升。实验设计的两个模块协调作用，与基准识别框架结构相比，所提算法模型 DAGM 在 COCO 数据集上的 AP，AP_{50} 和 AP_{75} 分别提升了 0.9 个百分点，1.4 个百分点和 0.9 个百分点。这与本文的动机相符，在特征的各个区域及各通道间采用有效的注意力，突出了特征图中目标物体的信息，相当于对特征图的通道进行了一次筛选，并且自上而下的特征融合丰富了特征图的上下文信息，更有利于对目标物体的检测，尤其对小目标物体的检测性能提升较为显著，证明了这两个模块的有效性。

5 结 论

提出了一种基于 DAGM 的目标检测算法。该算法包括两种注意力模块，一是在卷积神经网络提取特征的过程中，设计 MFCA 模块对特征图的各个区域赋予不同的重要性权值；二是对融合后的特征图引入 SCSA 模块，使得特征图的不同通道间形成差异，实现特征的自适应学习，从而强化了目标物体的信息，有效去除图像中的背景及负样本的信息干扰。对各模块的有效性进行了探究，实验结果表明所提算法具有更优的检测性能。

参 考 文 献

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 580-587.

[2] Girshick R. Fast R-CNN [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE, 2015: 1440-1448.

[3] 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 and Machine Intelligence, 2017, 39(6): 1137-1149.

[4] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M] // Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.

[5] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection [C] //

2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 779-788.

[6] Tan G H, Hou J, Han Y P, et al. Low-parameter real-time image segmentation algorithm based on convolutional neural network [J]. Laser & Optoelectronics Progress, 2019, 56(9): 091003.
谭光鸿, 侯进, 韩雁鹏, 等. 基于卷积神经网络的低参数量实时图像分割算法[J]. 激光与光电子学进展, 2019, 56(9): 091003.

[7] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[J/OL]. (2016-04-30)[2019-07-09]. https://arxiv.gg363.site/abs/1511.07122.

[8] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.

[9] Liu Y Z, Jiang Z Q, Ma F, et al. Hyperspectral image classification based on hypergraph and convolutional neural network [J]. Laser & Optoelectronics Progress, 2019, 56(11): 111007.
刘玉珍, 蒋政权, 马飞, 等. 基于超图和卷积神经网络的高光谱图像分类[J]. 激光与光电子学进展, 2019, 56(11): 111007.

[10] Liu F, Wang X, Lu L X, et al. Landform image classification based on sparse coding and convolutional neural network[J]. Acta Optica Sinica, 2019, 39(4): 0410001.
刘芳, 王鑫, 路丽霞, 等. 基于稀疏编码和卷积神经网络的地貌图像分类[J]. 光学学报, 2019, 39(4): 0410001.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J/OL]. (2015-04-10)[2019-07-09]. https://arxiv.gg363.site/abs/1409.1556.

[12] Xu Y, Sun M S. Convolution neural network image defogging based on multi-feature fusion[J]. Laser & Optoelectronics Progress, 2018, 55(3): 031012.
徐岩, 孙美双. 基于多特征融合的卷积神经网络图像去雾算法[J]. 激光与光电子学进展, 2018, 55(3): 031012.

[13] Zhao J T. Single-image defogging algorithm based on deep learning[J]. Laser & Optoelectronics Progress, 2019, 56(11): 111005.
赵建堂. 基于深度学习的单幅图像去雾算法[J]. 激光与光电子学进展, 2019, 56(11): 111005.

[14] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C] // 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL, USA. New York: IEEE, 2009: 248-255.

[15] Everingham M, van Gool L, Williams C K I, et al. The Pascal visual object classes (VOC) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.

[16] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[M] // Fleet D, Pajdla T, Schiele B, et al. Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 740-755.

[17] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate [J/OL]. (2016-05-19)[2019-07-09]. https://arxiv.gg363.site/abs/1409.0473.

[18] Xu K, Ba J, Kiros R, et al. Show, attend and tell: neural image caption generation with visual attention [C] // International Conference on Machine Learning, July 6-11, 2015, Lille, France. USA: MIT Press, 2015: 2048-2057.

[19] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks [C] // Advances in Neural Information Processing Systems, December 7-12, 2015, Montreal, Quebec, Canada. Canada: NIPS, 2015: 2017-2025.

[20] Chen L C, Yang Y, Wang J, et al. Attention to scale: scale-aware semantic image segmentation [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 3640-3649.

[21] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 936-944.

[22] Zhu Y S, Zhao C Y, Wang J Q, et al. CoupleNet: coupling global structure with local parts for object detection[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 4146-4154.

[23] Liu Y, Wang R P, Shan S G, et al. Structure inference net: object detection using scene-level context and instance-level relationships [C] // 2018 IEEE/CVF Conference on Computer Vision and

Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 6985-6994.

[24] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 6517-6525.

[25] Redmon J, Farhadi A. Yolov3: an incremental improvement[J/OL]. (2018-04-08)[2019-07-09]. https://arxiv.gg363.site/abs/1804.02767.

[26] Fu C Y, Liu W, Ranga A, et al. Dssd: deconvolutional single shot detector[J/OL]. (2017-01-23)[2019-07-09]. https://arxiv.gg363.site/abs/1701.06659.

[27] Dai J, Li Y, He K, et al. R-FCN: object detection via region-based fully convolutional networks[C]// Advances in Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. Canada: NIPS, 2016: 379-387.

[28] Huang J, Rathod V, Sun C, et al. Speed/accuracy trade-offs for modern convolutional object detectors [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3296-3305.

[29] Tychsen-Smith L, Petersson L. DeNet: scalable real-time object detection with directed sparse sampling [C]// 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 428-436.

[30] Wang H, Wang Q L, Gao M Q, et al. Multi-scale location-aware kernel representation for object detection [C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 1248-1257.

[31] Kong T, Sun F C, Yao A B, et al. RON: reverse connection with objectness prior networks for object detection[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 5244-5252.

[32] Hu H, Gu J, Zhang Z, et al. Relation networks for object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 18-22, 2018, Salt Lake City, Utah. New York: IEEE, 2018: 3588-3597.