第 59 卷 第 12 期/2022 年 6 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

融合自适应注意力机制的 Faster R-CNN 目标检测算法

王启胜[1,2,3]，王凤随[1,2,3*]，陈金刚[1,2,3]，刘芙蓉[1,2,3]

[1]安徽工程大学电气工程学院，安徽 芜湖 241000；

[2]检测技术与节能装置安徽省重点实验室，安徽 芜湖 241000；

[3]高端装备先进感知与智能控制教育部重点实验室，安徽 芜湖 241000

摘要 针对Faster R-CNN目标检测算法存在的定位和检测精度问题，设计了一种可嵌入Faster R-CNN目标检测算法并进行端到端训练的可移动的注意力(MA)模型。首先，为了获取更加精确的空间位置信息，MA采用两个自适应最大池化分别基于输入特征图的水平和竖直两个方向进行特征聚合，生成两个独立的方向感知特征图；其次，为了防止模型过拟合，使用Sigmoid激活函数增加网络非线性；最后，为了充分利用已经得到的空间位置信息，将具有非线性的两个特征图与输入特征图依次相乘以增强输入特征图的表征能力。实验结果表明：基于MA改进的Faster R-CNN目标检测算法有效地提升了网络对感兴趣目标的定位能力，并且平均检测精度也得到了明显的提升。

关键词 机器视觉；目标检测；Faster R-CNN；注意力机制；卷积神经网络；ResNet-50

中图分类号 TP181 **文献标志码** A **doi**：10.3788/LOP202259.1215016

Faster R-CNN Target-Detection Algorithm Fused with Adaptive Attention Mechanism

Wang Qisheng[1,2,3], Wang Fengsui[1,2,3*], Chen Jingang[1,2,3], Liu Furong[1,2,3]

[1]*School of Electrical Engineering, Anhui Polytechnic University, Wuhu* 241000, *Anhui, China*;

[2]*Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Wuhu* 241000, *Anhui, China*;

[3]*Key Laboratory of Advanced Perception and Intelligent Control of High-End Equipment, Ministry of Education, Wuhu* 241000, *Anhui, China*

Abstract To address the localization and the detection accuracy problems of the Faster R-CNN target-detection algorithm, a movable attention (MA) model that can be embedded in the algorithm and trained end-to-end is designed. First, to obtain more accurate spatial location information, MA uses two adaptive maximum pooling operations to aggregate features based on the horizontal and the vertical directions of the input feature and generates two independent directional-sensing feature maps. Second, to prevent model overfitting, the sigmoid activation function is used to increase network nonlinearity. Finally, to fully exploit the obtained spatial location information, the two nonlinear and input feature maps are multiplied successively to enhance the representational ability of the latter. The experimental results show that the improved Faster R-CNN target-detection algorithm based on MA can effectively enhance the network's ability to locate the target of interest, as well as considerably improve the average detection accuracy.

Key words machine vision; target detection; Faster R-CNN; attention mechanism; convolutional neural network; ResNet-50

收稿日期：2021-08-02；**修回日期**：2021-08-25；**录用日期**：2021-08-31

基金项目：安徽省自然科学基金(2108085MF197，1708085MF154)、安徽高校省级自然科学研究重点项目(KJ2019A0162)、检测技术与节能装置安徽省重点实验室开放基金(DTESD2020B02)

通信作者：*fswang@ahpu.edu.cn

1 引　　言

Faster R-CNN目标检测算法主要用于图像和视频任务中的感兴趣目标的定位和识别，而Faster R-CNN目标检测算法对感兴趣目标的定位识别能力与主干卷积神经网络的性能息息相关[1-5]。为了使卷积神经网络具有更好的性能，从具有开创性的AlexNet卷积神经网络模型[6]开始，许多研究不断展开[6-12]。

近年来，将注意力机制融入卷积神经网络得到了广泛的关注，并且取得了一定的成果[13-16]。2018年，Hu等[13]提出了通道注意力机制SE-Net，SE-Net通过建模特征通道间的动态、非线性关系，有效地提升了卷积神经网络对重要信息的关注。基于SE-Net，2018年，Woo等[14]提出卷积注意力机制(CBAM)，该注意力机制在给定中间特征图的情况下，先后经过通道和空间两个独立的维度对特征信息进行加权得到最终的特征图，并在实验中取得了更好的结果；2020年，Wang等[15]提出了高效的通道注意力机制ECA-Net，与SE-Net不同的是，ECA-Net采用一维卷积代替了SE-Net的全连接层组成的瓶颈结构，避免了由于降维造成的细节信息损失问题；2021年，Hou等[16]针对SE-Net注意力机制只考虑了通道间的信息编码而忽略了位置信息的问题提出了新的注意力机制(CA)，该注意力机制使用两个一维编码代替二维全局池化，分别沿着两个空间方向聚合特征，以增强关注对象的表示，提高了卷积神经网络对感兴趣目标的定位和检测精度。

然而，CBAM和CA注意力机制通过降低输入张量的通道维数来建模通道注意力的做法会导致细节信息的丢失；ECA-Net注意力机制只考虑通过建模通道关系来衡量每个通道的重要性，没有考虑空间位置信息，而空间位置信息在计算机视觉任务中对物体精准定位具有重要的作用。

基于以上问题，为了提高卷积神经网络性能继而提高Faster R-CNN目标检测算法对感兴趣目标的定位和识别能力，设计了一种可移动的注意力(MA)模型。首先，采用两个自适应池化核分别基于输入特征图的水平(宽)和竖直(高)两个方向进行特征聚合，生成两个独立的具有方向感知的注意力特征图；然后，利用Sigmoid激活函数增加模型的非线性；最后，将经过Sigmoid激活函数的两个特征图与输入特征图依次相乘，增强输入特征图的表征能力，并得到最终的注意力特征图。

2 MA模型

MA可以看成是一个计算单元，旨在提高卷积神经网络对重要特征信息的学习能力。MA能够以任意特征张量 $\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}\in\mathbf{R}^{C\times H\times W}$ 作为输入，并输出具有远程依赖关系和精确位置信息的变换张量 $\boldsymbol{Y}=\{\boldsymbol{y}_1,\boldsymbol{y}_2,\ldots,\boldsymbol{y}_n\}\in\mathbf{R}^{C\times H\times W}$，其结构如图1所示。

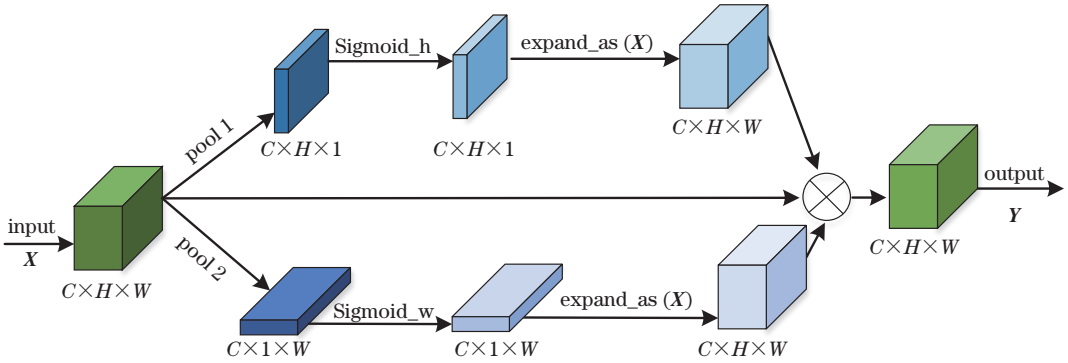

图1　MA模块结构图

Fig. 1　Structure diagram of MA

在注意力机制中，通常使用全局池化操作对空间信息进行编码，但是全局池化很难保留位置信息，而空间位置信息可以有效地提高目标检测算法对目标的定位和识别能力。因此，所设计的MA采用两个自适应最大池化核分别基于输入特征图的水平和竖直两个方向进行特征聚合，生成两个独立的方向感知注意力特征图。具体来说，对于输入特征图 $\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$，分别使用尺寸为(H,1)和(1,W)的自适应最大池化核沿着特征图的水平和竖直两个方向进行编码生成注意力特征图。沿着水平方向对输入特征图的每个通道进行编码操作可描述为

$$\boldsymbol{Z}_c^h(h)=\frac{1}{w}\sum_{j=1}^{w}\boldsymbol{X}_c(h,j),\tag{1}$$

式中：$\boldsymbol{Z}_c^h(h)\in\mathbf{R}^{C\times H\times 1}$表示竖直维度不变，对第$C$个通道按水平方向进行依次加权平均或取最大值获得的注意力张量。

类似地，当沿着竖直方向对输入特征图的每个通道进行编码操作时，有

$$\boldsymbol{Z}_c^w(w)=\frac{1}{h}\sum_{i=1}^{h}\boldsymbol{X}_c(i,w),\tag{2}$$

式中：$\boldsymbol{Z}_c^w(w)\in\mathbf{R}^{C\times 1\times W}$表示水平维度不变，对第$C$个通道按竖直方向进行依次加权平均或取最大值获得的注意力张量。

为了更好地利用得到的具有方向感知和细节位置信息的特征张量，所设计的MA应满足4个标准。1）由于Faster R-CNN本身网络的复杂性，应保证MA整体结构的简单性；2）要避免降维操作导致的信息损失问题；3）应该增加注意力机制的非线性表达能力，防止过拟合；4）应充分利用已经得到的空间位置信息，从而提高对感兴趣目标的定位和识别能力。具体设计如下：首先，避免对式（1）、（2）输出的注意力张量进行通道上的降维操作；然后，直接对式（1）、（2）得到的聚合特征图利用Sigmoid激活函数增加非线性，防止过拟合。即

$$\boldsymbol{f}^h=\sigma(\boldsymbol{Z}^h),\tag{3}$$

$$\boldsymbol{f}^w=\sigma(\boldsymbol{Z}^w),\tag{4}$$

式中：σ表示Sigmoid激活函数；$\boldsymbol{f}^h\in\mathbf{R}^{C\times H\times 1}$和$\boldsymbol{f}^w\in\mathbf{R}^{C\times 1\times W}$表示经过Sigmoid激活函数处理得到的具有非线性的两个独立的方向感知中间注意力权重张量。

为了充分利用已经得到的具有空间位置信息的注意力权重张量。首先，将通过式（3）和式（4）得到的两个中间注意力权重张量分别扩展成和输入特征图$\boldsymbol{X}$一样的维度，即图1中Expand_as($\boldsymbol{X}$)操作，获得与输入特征图$\boldsymbol{X}$一样的维度并分别包含不同方向信息的特征张量$\boldsymbol{f}_c^h(i,j)\in\mathbf{R}^{C\times H\times W}$和$\boldsymbol{f}_c^w(i,j)\in\mathbf{R}^{C\times H\times W}$；然后，将输入特征图张量与$\boldsymbol{f}_c^h(i,j)$、$\boldsymbol{f}_c^w(i,j)$依次相乘，使输入特征图张量的每一个值都与权值进行相乘，得到包含细节信息的输出特征图$\boldsymbol{Y}_c(i,j)\in\mathbf{R}^{C\times H\times W}$。

$$\boldsymbol{Y}_c(i,j)=\boldsymbol{X}_c(i,j)\times\boldsymbol{f}_c^h(i,j)\times\boldsymbol{f}_c^w(i,j)。\tag{5}$$

3　实验及结果分析

3.1　实验和环境配置

由于Faster R-CNN主干卷积神经网络ResNet-50[10]的浅层和深层网络输出特征图的表达能力不同，实验中在保证MA注意力机制结构不变的情况下，设置了不同的池化核对浅、深层输出特征图进行编码。具体来说，当输入特征图为卷积神经网络浅层输出时，即当输入特征图具有较多纹理信息时，分别使用尺寸为(H,1)和(1,W)的自适应最大池化核沿着特征图水平和竖直两个方向进行编码生成注意力特征图；当输入特征图为卷积神经网络深层输出时，即当输入特征图具有较多语义信息时，分别使用尺寸为(H,1)和(1,W)的自适应平均池化核沿着特征图水平和竖直两个方向进行编码生成注意力特征图。实验中，为了更好地训练改进的网络模型并取得好的测试结果，均使用迁移学习[17]的方法进行模型训练，并且两种使用不同池化核的注意力机制在ResNet-50中的具体位置如表1所示，其中MA*代表深层网络注意力机制，MA代表浅层网络注意力机制。此外，实验环境配置如表2所示。

表1　基于MA改进的ResNet-50结构对比

Table 1　Comparison of modified ResNet-50 structure based on MA

Layer name	Output size	ResNet-50	ResNet-50_MA*	ResNet-50_MA
Conv1	112×112	7×7, 64, stride 2		
Attention 0	112×112			MA
		3×3 max pool, stride 2		
Conv2_x	56×56	$\begin{bmatrix}1\times1,64\\3\times3,64\\1\times1,256\end{bmatrix}\times3$		
Conv3_x	28×28	$\begin{bmatrix}1\times1,128\\3\times3,128\\1\times1,512\end{bmatrix}\times4$		
Conv4_x	14×14	$\begin{bmatrix}1\times1,256\\3\times3,256\\1\times1,1024\end{bmatrix}\times6$		
Conv5_x	7×7	$\begin{bmatrix}1\times1,512\\3\times3,512\\1\times1,2048\end{bmatrix}\times3$		
Attention 1	7×7		MA*	
	1×1	Average pool, fc, softmax		

表 2 实验环境配置

Table 2 Experimental environment configuration

Name	Model number
System	Windows 10
CPU	Core i9-10900 @ 3.7 GHz
GPU	Nvidia GeForce RTX 2080Ti(11 GB)
Framework	Pytorch 1.2.0
Language	Python

3.2 数据集和评价指标

所有实验均在PASCAL VOC数据集[18]上进行，利用PASCAL VOC2007和PASCAL VOC2012的训练集合并得到的16551张图片进行训练，并使用PASCAL VOC2007的测试集进行测试，具体情况如表3所示。

表 3 PASCAL VOC数据集训练和测试数据统计

Table 3 PASCAL VOC dataset training and test data statistics

Data	Trainval		Test	
	Images	Objects	Images	Objects
VOC2007	5011	12608	4952	12032
VOC2012	11540	27450	0	0
Total	16551	40058	4952	12032

为了验证嵌入MA模块的Faster R-CNN目标检测算法的性能，用检测平均精度(AP)对目标检测输出的每一类别结果精确度进行评估，并使用所有类别AP和的平均值(mAP)来衡量整个模型的性能，其中AP由精度(P)和召回率(R)构成的曲线积分面积组成，P和R的表达式为

$$P=\frac{N_{TP}}{N_{TP}+N_{FP}}, \tag{6}$$

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

式中：N_{TP}指正样本被检测为正样本的框的数量；N_{FP}指负样本被误检为正样本的框的数量；N_{FN}指正样本被检测为负样本，即检测错误的框的数量。

3.3 实验结果与分析

首先，为了确定最终的注意力机制模型，分别将MA*、MA嵌入Faster R-CNN目标检测算法主干卷积神经网络ResNet-50对应位置中进行实验，每类目标的AP和衡量整个模型性能指标的mAP如表4所示，其中FR表示基于ResNet-50复现的Faster R-CNN目标检测算法；FR+MA*、FR+MA分别表示基于MA*和MA改进的Faster R-CNN目标检测算法；加粗数据表示相应的改进算法相对于原算法涨点的数据；“+”表示改进算法相对于原算法的增长数，“−”表示改进算法相对原算法的减少数。

表 4 20个目标类的检测精度对比

Table 4 Comparison of detection accuracy of 20 target classes unit: %

Category	FR	FR+MA*	FR+MA
Cat	87.23	**87.31$^{+0.08}$**	**89.28$^{+2.05}$**
Car	86.60	86.33$^{-0.27}$	85.85$^{-0.75}$
Horse	84.87	**86.14$^{+1.27}$**	**86.50$^{+1.63}$**
Dog	83.62	82.87$^{-0.75}$	**84.29$^{+0.67}$**
Bus	80.29	**82.55$^{+2.26}$**	**83.45$^{+3.16}$**
Train	82.49	**83.12$^{+0.63}$**	**82.96$^{+0.47}$**
Motorbike	83.48	**83.62$^{+0.14}$**	83.07$^{-0.41}$
Bicycle	79.56	**83.14$^{+3.58}$**	**82.72$^{+3.16}$**
Person	79.75	**80.09$^{+0.34}$**	**80.38$^{+0.63}$**
Aeroplane	79.75	76.09$^{-3.66}$	74.28$^{-5.47}$
Sheep	75.22	**78.26$^{+3.04}$**	73.20$^{-2.02}$
Bird	74.14	**76.45$^{+2.31}$**	**77.53$^{+3.39}$**
Cow	74.73	**78.78$^{+4.05}$**	74.70$^{-0.03}$
Tvmonitor	74.34	73.33$^{-1.01}$	73.09$^{-1.25}$
Diningtable	72.26	71.63$^{-0.63}$	**73.96$^{+1.70}$**
Sofa	70.44	**75.12$^{+4.68}$**	**73.31$^{+2.87}$**
Boat	65.69	62.92$^{-2.77}$	**66.60$^{+0.91}$**
Chair	54.19	51.53$^{-2.66}$	**56.31$^{+2.12}$**
Bottle	52.06	**57.03$^{+4.97}$**	**56.48$^{+4.42}$**
Pottedplant	45.96	43.85$^{-2.11}$	45.25$^{-0.71}$
mAP	74.25	**75.01$^{+0.76}$**	**75.16$^{+0.91}$**

由表4数据可知：在Faster R-CNN目标检测算法的主干卷积神经网络浅层输出后串接注意力机制MA有效地提升了65%目标类别的检测精度，高于在Faster R-CNN目标检测算法的主干卷积神经网络深层输出后串接注意力机制MA*提升的60%目标类别的检测结果；从单个目标类可以看出，嵌入MA注意力机制的Faster R-CNN目标检测算法有效地提升了大目标和纹理比较清晰的目标类的检测精度，例如公交车、自行车、沙发、瓶子等；此外，使用MA改进的Faster R-CNN目标检测算法的mAP提升了0.91个百分点，高于用MA*改进后测试得到的检测精度。综上所述，将使用自适应最大池化核的MA作为最终注意力模型。

其次，为了研究MA中水平、竖直两个方向的注意力权值对提升模型性能的影响，将MA进行拆分，分成两个分别只含水平或竖直方向的注意力模块并嵌入MA在Faster R-CNN中的对应位置进行实

验，即分别断开图 1 中的最下和最上面的支路进行实验，相关实验数据记录如表 5 所示，其中 MA1 表示只含水平注意力模块，MA2 表示只含竖直注意力模块。

表 5　MA 消融实验结果对比

Table 5　Comparison of experimental results of ablation of MA　　unit: %

Category	FR	FR+MA1	FR+MA2
Cat	87.23	$87.10^{-0.13}$	$90.07^{+2.84}$
Car	86.60	$85.94^{-0.66}$	$84.66^{-1.94}$
Horse	84.87	$85.20^{+0.33}$	$85.90^{+1.03}$
Dog	83.62	$86.41^{+2.79}$	$85.11^{+1.49}$
Bus	80.29	$83.17^{+2.88}$	$85.36^{+5.07}$
Train	82.49	$81.98^{-0.51}$	$82.09^{-0.40}$
Motorbike	83.48	$81.38^{-2.10}$	$81.37^{-2.11}$
Bicycle	79.56	$81.80^{+2.24}$	$83.38^{+3.82}$
Person	79.75	$78.64^{-1.11}$	$79.61^{-0.14}$
Aeroplane	79.75	$77.68^{-2.07}$	$76.20^{-3.55}$
Sheep	75.22	$75.62^{+0.40}$	$73.97^{-1.25}$
Bird	74.14	$75.58^{+1.44}$	$75.87^{+1.73}$
Cow	74.73	$78.09^{+3.36}$	$73.06^{-1.67}$
Tvmonitor	74.34	$72.20^{-2.14}$	$73.79^{-0.55}$
Diningtable	72.26	$70.78^{-1.48}$	$70.58^{-1.68}$
Sofa	70.44	$74.22^{+3.78}$	$73.24^{+2.80}$
Boat	65.69	$66.58^{+0.89}$	$66.29^{+0.66}$
Chair	54.19	$53.49^{-0.70}$	$51.81^{-2.38}$
Bottle	52.06	$51.17^{-0.89}$	$55.37^{+3.31}$
Pottedplant	45.96	$45.20^{-0.76}$	$46.21^{+0.25}$
mAP	74.25	$74.61^{+0.36}$	$74.70^{+0.45}$

由表 5 可知：用水平注意力模块 MA1 和竖直注意力模块 MA2 进行改进的 Faster R-CNN 目标检测算法分别提高了 45%、50% 目标类别的检测精度；从单个目标类别看，经 MA2 改进的 Faster R-CNN 目标检测算法最高提升了 5.07 个百分点的检测精度，经 MA1 改进的 Faster R-CNN 目标检测算法最高提升了 3.78 个百分点的检测精度；从衡量整个模型性能的 mAP 看，经 MA2 改进的 Faster R-CNN 目标检测算法取得了更好的检测精度。综上所述，MA 注意力机制的竖直方向注意力权值对提升模型性能更加有益。

最后，为了让 MA 的有效性表现得更加直观，随机抽取几张图片放在未改进和基于 MA 改进的 Faster R-CNN 目标检测算法中进行检测，结果如图 2 所示。

(a)　(b)

图 2　检测结果对比图。(a)原始算法检测结果；(b)基于 MA 改进的算法检测结果

Fig. 2　Comparison diagram of test results. (a) Original algorithm detection results; (b) Detection results of improved algorithm based on MA

由检测结果对比图可以看出，嵌入 MA 注意力机制的 Faster R-CNN 目标检测算法提高了对感兴趣目标的定位准确性，减少了漏检和误检情况。由第 1 行和第 2 行的对比图可以看到，改进的算法对感兴趣目标的定位变得更加准确，并提高了对应的置信度得分；由第 3 行的对比图可以发现，改进的算法降低了误检的概率，并提高了目标物在方向上的关联性，使得植物被有效检测；此外，由第 4 行的行人图可以看出，改进的算法提高了被遮挡物体的定位和识别精度。综合表 4 和图 2 可以发现，基于 MA 改进的 Faster R-CNN 目标检测算法在一定程度上降低了漏检和误检概率，提高了定位和检测精度。

5　结　　论

提出了一种提升 Faster R-CNN 目标检测算法性能的可移动注意力机制 MA。该注意力机制可以使特征图获得方向感知和位置感知信息，并且不降维的设计可以避免细节信息的损失。最终实验结果表明，与原算法相比，改进的 Faster R-CNN 目标检测算法提高了 65% 类目标的检测平均精度，其中单

个目标类的检测平均精度最高提升了4.42个百分点，此外，最终改进的Faster R-CNN目标检测算法使mAP得到了0.91个百分点的提升，有效地提升了模型对感兴趣目标的定位和检测精度。

参 考 文 献

[1] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.

[2] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[EB/OL]. (2015-12-08)[2021-01-02]. https://arxiv.org/abs/1512.02325.

[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, New York: IEEE Press, 2017, 39(6): 1137-1149.

[4] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23) [2021-05-08]. https://arxiv. org/abs/2004.10934.

[5] 姚群力，胡显，雷宏. 基于多尺度卷积神经网络的遥感目标检测研究[J]. 光学学报，2019，39(11): 1128002.
Yao Q L, Hu X, Lei H. Object detection in remote sensing images using multiscale convolutional neural networks[J]. Acta Optica Sinica, 2019, 39(11): 1128002.

[6] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[7] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015: 15523970.

[8] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, New York: IEEE Press, 1998, 86(11): 2278-2324.

[9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04) [2021-05-08]. https://arxiv. org/abs/1409.1556.

[10] 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, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[11] Li P H, Xie J T, Wang Q L, et al. Is second-order information helpful for large-scale visual recognition? [C]// 2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2089-2097.

[12] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7794-7803.

[13] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.

[14] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[EB/OL]. (2018-07-17)[2021-05-06]. https://arxiv.org/abs/1807.06521.

[15] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.

[16] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]// 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 13708-13717.

[17] 张雪松，庄严，闫飞，等. 基于迁移学习的类别级物体识别与检测研究与进展[J]. 自动化学报，2019，45(7): 1224-1243.
Zhang X S, Zhuang Y, Yan F, et al. Status and development of transfer learning based category-level object recognition and detection[J]. Acta Automatica Sinica, 2019, 45(7): 1224-1243.

[18] Everingham M, Eslami S M A, Gool L, et al. The pascal visual object classes challenge: a retrospective [J]. International Journal of Computer Vision, 2015, 111(1): 98-136.