具有全局特征的空间注意力机制

张连超，乔瑞萍，党祺玮，翟沛源，孙红帅
（西安交通大学信息与通信工程学院，710049，西安）

摘要：为了获得卷积神经网络特征图中不同特征点之间的长距离依赖关系，使卷积神经网络更好地区分前景目标和背景信息，提出了一种具有全局特征的空间注意力机制。通过通道融合层将多通道的原始特征图组合成单通道的特征融合图，消除了通道间信息分布对获取空间注意力权重的影响；将特征融合图经过全局特征获取处理，获得能够反映特征融合图中某特征点与特征融合图中所有点之间相关性的全局特征图；全局特征图与初始值为 0 的可学习变量相乘，并且在通道域复制自身，扩展为原始特征图大小，将扩展后的全局特征图与原始特征图对应元素相加，获得具有注意力机制的特征图。在不同卷积神经网络中加入具有全局特征的空间注意力机制进行实验，结果表明：在脑电波二分类任务中，所提注意力机制的分类准确率最高提升了 0.839%；在 CIFAR-10 数据集多分类任务中，所提注意力机制的分类准确率最高提升了 0.484%；在夜间车辆单类别检测中，在交并比阈值大于 0.5 的平均精度评判标准下，所提注意力机制最高提升了 3.860%，在交并比阈值大于 0.75 的平均精度评判标准下，所提注意力机制最高提升了 11.726%；在 voc2007 数据集多类别检测中，在交并比阈值大于 0.5 的平均精度评判标准下，所提注意力机制最高提升了 0.778%，在交并比阈值大于 0.75 的平均精度评判标准下，所提注意力机制最高提升了 1.232%。

关键词：卷积神经网络；空间注意力机制；全局特征；特征融合；目标分类；目标检测
中图分类号：TP391.4 **文献标志码**：A
DOI：10.7652/xjtuxb202011016 **文章编号**：0253-987X(2020)11-0129-10

OSID 码

Spatial Attention Mechanism with Global Characteristics

ZHANG Lianchao, QIAO Ruiping, DANG Qiwei, ZHAI Peiyuan, SUN Hongshuai
(School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To obtain the long-distance dependence between different feature points in the feature map of the convolutional neural network, so that the convolutional neural network can better distinguish the foreground target and background information, a spatial attention mechanism with global features is proposed. Combining the multi-channel original feature map into a single-channel feature fusion map through the channel fusion layer, the influence of the information distribution between channels on obtaining spatial attention weights is eliminated; the feature fusion map through global feature acquisition processing to obtain the global feature map, which represents the correlation between a feature point and all points in the feature fusion map; the global feature map is multiplied by the learnable variable with an initial value of 0, and copies itself to the size of the original feature map in channel domain. The expanded global feature map is added to the original feature map to obtain a feature map with attention mechanism. After

收稿日期：2020-01-06。 作者简介：张连超(1993—)，男，硕士生；乔瑞萍(通信作者)，女，副教授。 基金项目：陕西省重点研发计划资助项目(2020GY-074)。
网络出版时间：2020-06-24 网络出版地址：http://kns.cnki.net/kcms/detail/61.1069.T.20200623.1332.002.html

adding the spatial attention mechanism with global features to different convolutional neural networks, the experimental results show that in the brain wave two-classification task, the algorithm classification accuracy is heightened by the maximum of 0.839%; in the CIFAR-10 data set multi-classification task, the classification accuracy of the algorithm is improved by the maximum of 0.484%; in the single-category detection of night vehicles, the algorithm is heightened by the maximum of 3.860% with the evaluation standard of the average accuracy value of Intersection over union (IoU) greater than 0.5, and the algorithm is improved by the maximum of 11.726% with the evaluation standard of the average accuracy value of IoU greater than 0.75; in the multi-category detection of the voc2007 data set, the algorithm has improved by the maximum of 0.778% with the evaluation criterion of the average precision value of IoU greater than 0.5, and the algorithm has improved by the maximum of 1.232% with the evaluation criterion of the average precision value of IoU greater than 0.75.

Keywords: convolutional neural networks; spatial attention mechanism; global features; feature fusion; target classification; target detection

注意力机制广泛应用于各种任务中,例如机器翻译[1-3]和视觉识别领域[4-7]。在网络中添加注意力机制可以有效地提高网络的表达能力。目前,视觉识别领域主要有 3 种添加注意力机制的方式:空间注意力机制、通道注意力机制和混合注意力机制。通道注意力机制是在卷积网络不同通道间添加注意力机制,例如压缩激励块[8]。压缩激励块通过全局池化层对同一特征图的不同通道进行空间域压缩,然后沿通道域进行重新加权,权重大小表示各个通道的重要程度。在压缩激励块的基础上引入全局信息从而得到全局上下文块[9]。空间注意力机制是在卷积网络的空间域内添加注意力机制,例如借鉴非局部均值法的非局部神经网络[10],利用特征图中所有位置的特征加权和计算出一个位置的响应。空间注意力机制和通道注意力机制可以同时加入到网络中,代表性的混合注意力机制有卷积块注意力模块[11]。卷积块注意力模块是通过全局最大池化和全局平均池化提取空间和通道权重,并将两种权重融合的注意力机制。空间和通道压缩激励块是一种将压缩激励块的注意力机制从通道域扩展到空间域得到通道压缩和空间激励块,然后将通道压缩和空间激励块与压缩激励块结合的注意力机制[12]。通道压缩和空间激励块通过卷积层将多通道特征压缩为单通道特征,消除通道间信息分布对空间注意力机制的影响,然后通过 sigmoid 函数归一化空间权重,最终将空间权重信息和输入特征图对应元素相乘。

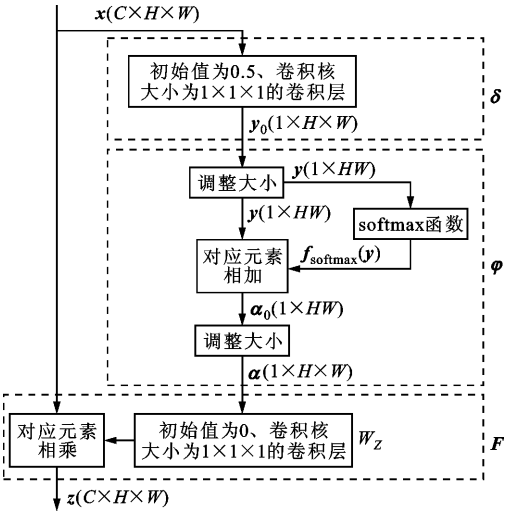
非局部神经网络借鉴非局部均值法[13]的思想,利用特征图中所有位置的特征加权信息计算出一个位置的响应,最终获得的像素级空间注意力权重反映了同一通道内不同位置间的依赖关系。正因为如此,非局部神经网络的计算量非常大。简化的非局部神经网络发现特征图中任意点的全局权重相似,通过共享全局权重的方式简化非局部神经网络的计算过程[9]。与非局部神经网络不同,简化的非局部神经网络最终获得通道注意力权重来表达各个通道的重要情况。

通道压缩和空间激励块沿通道域压缩,在空间域激活获得空间注意力权重。虽然融合了各通道信息,但是缺乏空间全局信息。非局部神经网络利用特征图中所有位置的特征加权和计算出一个位置的响应,计算复杂,并且无法兼容各通道信息。基于此,本文在融合各通道信息的基础上加入特征图的全局信息,提出一种新型轻量级注意力机制模块——具有全局特征的空间注意机制。比较通道压缩和空间激励块,具有全局特征的空间注意力机制增加了特征图的全局信息;比较非局部神经网络,具有全局特征的空间注意力机制融合了各通道信息,并且减少了计算成本。

1 具有全局特征的空间注意力机制

具有全局特征的空间注意力机制结构如图 1 所示。大小为 $C\times H\times W$(C 个通道,每个通道 H 行 W 列)的原始特征图 $\boldsymbol{x}$ 经过通道融合层 $\boldsymbol{\delta}$(初始值为 0.5、卷积核大小为 $1\times1\times1$ 的卷积层),得到大小为 $1\times H\times W$ 的特征融合图 $\boldsymbol{y}_0$。$\boldsymbol{y}_0$ 调整大小变为 $1\times HW$(1 行 HW 列)的特征融合矩阵 $\boldsymbol{y}$。$\boldsymbol{y}$ 经过 softmax 函数后,得到大小为 $1\times HW$ 的相关性矩阵 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y})$;$\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y})$ 和 $\boldsymbol{y}$ 对应元素点乘,得到大小

为 $1\times HW$ 的全局特征矩阵 $\boldsymbol{\alpha}_0$；$\boldsymbol{\alpha}_0$ 调整大小变为 $1\times H\times W$ 的全局特征图 $\boldsymbol{\alpha}$；$\boldsymbol{\alpha}$ 中每个元素乘上初始值为 0 的可学习变量 W_Z（初始值为 0 的 $1\times1\times1$ 卷积核权重），然后复制自身，大小变为 $C\times H\times W$，最后与 $\boldsymbol{x}$ 中的对应元素相加，获得大小为 $C\times H\times W$ 的具有注意力机制的特征图 $\boldsymbol{z}$。



$\boldsymbol{\varphi}$—全局特征获取；$\boldsymbol{F}$—全局特征融合。

图 1　具有全局特征的空间注意力机制结构

由图 1 可知，具有全局特征的空间注意力机制结构分为 3 个流程：①通道融合层获取；②全局特征获取；③全局特征聚合。具有全局特征的空间注意力机制可以抽象为

$$\boldsymbol{z}=\boldsymbol{F}(\boldsymbol{x},W_Z\boldsymbol{\varphi}(\boldsymbol{\delta}(\boldsymbol{x}))) \tag{1}$$

1.1　通道融合层获取

非局部神经网络和简化的非局部神经网络直接在原始特征图上进行全局特征提取，而具有全局特征的空间注意力机制在获取全局特征前加入通道融合层。本文借鉴通道压缩和空间激励块中的通道压缩方式进行通道融合，也就是通过一个 $1\times1\times1$ 卷积核的卷积层将多通道特征压缩为单通道特征，从而获得全通道信息。这样做的目的主要有两点：①减少运算量，单通道特征比多通道特征的计算量会大幅度降低；②增加鲁棒性，不同通道只能表示同一个位置的不同特性，通过卷积层将不同通道进行线性组合，获得的通道融合层可以有效融合同一位置的不同特性，从而增加了特征图表达的鲁棒性。通道融合层中的数学关系可以表达为

$$y_{ij}=\sum_{N=1}^{C}w_{\mathrm{g}}x_{ij}^{N},\ 1\leqslant i\leqslant H,1\leqslant j\leqslant W \tag{2}$$

式中：x_{ij}^{N} 表示 $\boldsymbol{x}$ 中第 N 通道位置 (i,j) 的元素；w_{g} 表示通道融合层中 $1\times1\times1$ 卷积核的权重；y_{ij} 表示 $\boldsymbol{y}_0$ 中的元素。以 $\boldsymbol{y}_0$ 中位置(1,1)处的点 y_{11} 为例，由式(3)得到 y_{11} 为

$$y_{11}=w_{\mathrm{g}}\sum_{N=1}^{C}x_{11}^{N} \tag{3}$$

通道融合层如图 2 所示。

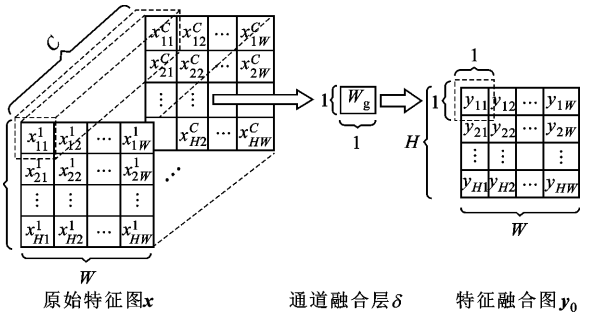


图 2　通道融合层示意图

原始特征图 $\boldsymbol{x}$ 有 C 个通道，说明其上层卷积拥有 C 个卷积核。如果将卷积核类比为滤波器，则 $\boldsymbol{x}$ 第 N 个通道就是浅层特征经过上层卷积中第 N 个卷积核所表达的特征。对于式(3)中的 x_{ij}^{N}，当 $i=j=1$ 时，其表示图 2 中的长方体框柱部分$[x_{11}^{1},x_{11}^{2},\cdots,x_{11}^{N},\cdots,x_{11}^{C}]$，$x_{11}^{N}$ 是位置(1,1)处的第 N 种特征。特征融合图 $\boldsymbol{y}_0$ 在位置(1,1)处的值 y_{11} 是由 C 种 x_{11}^{N} 线性组合而成。因此，y_{11} 表达位置(1,1)处的信息就兼顾了多种特征，而 x_{11}^{N} 表达位置(1,1)处的信息只关注了第 N 种特征。相比较单一通道获取全局特征的方法，具有全局特征的空间注意力机制中多特征融合产生的全局特征鲁棒性更强。

1.2　全局特征获取

卷积神经网络通过共享卷积核进行特征提取，一般情况下卷积核大小为 3×3，这样会导致网络特征图的局部感受野较小，卷积神经网络只能学习到目标的局部特征。为了解决这个问题，可以通过增大卷积核来获取更大的感受野，例如 AlexNet 中采用 11×11 和 5×5 的卷积核[14]，但是这样会导致运算量和参数量的急剧增加，不利于模型深度的增加。VGGNet 提出通过利用多个 3×3 卷积核替代大型卷积核[15]，但浅层区域依然没有全局信息。基于此，诞生了借鉴非局部均值法的非局部神经网络，非局部神经网络通过计算一点与其他全部点的相关度提取全局特征。

1.2.1　*通道融合层的全局特征提取*　借鉴非局部神经网络的方法获取全局特征。经过 1.1 小节中的通道融合层后，得到了 H 行 W 列的特征融合图 $\boldsymbol{y}_0$。将 $\boldsymbol{y}_0$ 调整大小为 1 行 HW 列的矩阵 $\boldsymbol{y}$。$\boldsymbol{y}_0$ 中位置 (i,j) 的点，就是 $\boldsymbol{y}$ 中位置 $(1,iW+j)$ 的点 y_{ij}。y_{ij} 点

对应的全局特征是利用 y_{ij} 与 $\boldsymbol{y}$ 中所有特征点计算得出的，y_{ij} 点对应的全局特征可以表示为

$$\alpha_{ij}=\frac{1}{C'(\boldsymbol{y})}\sum_{a}\sum_{b}f(y_{ij},y_{ab})g(y_{ab}) \tag{4}$$

式中：α_{ij} 表示 y_{ij} 点对应的全局特征；y_{ab} 表示 $\boldsymbol{y}$ 中的一点；$f(y_{ij},y_{ab})$ 用于衡量 y_{ij} 和 y_{ab} 之间的相关性；$C'(\boldsymbol{y})$ 表示归一化函数；$g(y_{ab})$ 是对 y_{ab} 进行线性变化，非局部神经网络中是为了匹配通道数和获得各通道之间的依赖关系，简化的非局部神经网络中是为了获得各通道之间的依赖关系。$g(y_{ab})$ 采用 1×1 卷积核，其形式为

$$g(y_{ab})=W_{\rho}y_{ab} \tag{5}$$

式中 W_{ρ} 表示卷积权重。

$f(y_{ij},y_{ab})$ 采用高斯函数计算两点间的相关性，其形式为

$$f(y_{ij},y_{ab})=\mathrm{e}^{y_{ij}y_{ab}} \tag{6}$$

当位置 (i,j) 为定值时，$C'(\boldsymbol{y})$ 的形式为

$$C'(\boldsymbol{y})=\sum_{a}\sum_{b}f(y_{ij},y_{ab}) \tag{7}$$

由式(6)(7)可以推导出

$$\frac{1}{C'(\boldsymbol{y})}f(y_{ij},y_{ab})=\frac{\mathrm{e}^{y_{ij}y_{ab}}}{\sum_{a}\sum_{b}\mathrm{e}^{y_{ij}y_{ab}}}=f_{\mathrm{softmax}}(y_{ij}y_{ab}) \tag{8}$$

将式(8)和式(5)代入式(4)可得

$$\alpha_{ij}=\sum_{a}\sum_{b}f_{\mathrm{softmax}}(y_{ij}y_{ab})W_{\rho}y_{ab} \tag{9}$$

1.2.2 *全局特征提取简化* 简化的非局部神经网络经过实验发现，一幅图中每个点的全局特征相似。例如：对于特征融合矩阵 $\boldsymbol{y}$，通过余弦相似度和 JS 散度进行距离测算可以发现，任意 y_{ij} 与 $\boldsymbol{y}$ 的相关性矩阵 $\boldsymbol{f}_{\mathrm{softmax}}(y_{ij}\boldsymbol{y})$ 相似。

相关性矩阵的计算过程可以分为两步：①$HW\times 1$ 的 $\boldsymbol{y}^{\mathrm{T}}$ 与 $1\times HW$ 的 $\boldsymbol{y}$ 相乘得到 $HW\times HW$ 的二维矩阵；②$HW\times HW$ 的二维矩阵每一行分别经过 softmax 变换得到相关性矩阵 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y})$。

y_{11} 与 $\boldsymbol{y}$ 的相关性矩阵 $\boldsymbol{f}_{\mathrm{softmax}}(y_{11}\boldsymbol{y})$ 可以理解为 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y})$ 中的第一列。根据简化的非局部神经网络实验得出，不同 y_{ij} 的 $\boldsymbol{f}_{\mathrm{softmax}}(y_{ij}\boldsymbol{y})$ 差别很小。因此，$\boldsymbol{f}_{\mathrm{softmax}}(y_{ij}\boldsymbol{y})$ 与 y_{ij} 的取值关系不大。令 $y_{ij}=1$，则 $\boldsymbol{f}_{\mathrm{softmax}}(y_{ij}\boldsymbol{y})$ 可以简化为 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y})$。$\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y})$ 可由 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y})$ 近似表达，$\boldsymbol{\alpha}$ 由 $\boldsymbol{W}_{\rho}\boldsymbol{y}$ 与 $\boldsymbol{f}^{\mathrm{T}}_{\mathrm{softmax}}(\boldsymbol{y})$ 矩阵相乘得出，因此简化的非局部神经网络中的 α_{ij} 变为

$$\alpha_{ij}=\sum_{a}\sum_{b}f_{\mathrm{softmax}}(y_{ab})W_{\rho}y_{ab} \tag{10}$$

从式(10)可以看出，同一通道内无论位置 (i,j) 如何变化，α_{ij} 都相同。因此，简化的非局部神经网络是通道注意力机制。但是，非局部神经网络得到的是像素级注意力权重，空间领域较通道领域更能体现注意力权重的细节。本文欲使简化后的非局部神经网络是一种空间注意力机制。因此，简化过程有两步：①利用 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y})$ 取代 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y})$；②将 $W_{\rho}\boldsymbol{y}$ 与 $\boldsymbol{f}_{\mathrm{softmax}}(\boldsymbol{y})$ 点乘。简化后的 α_{ij} 为

$$\alpha_{ij}=f_{\mathrm{softmax}}(y_{ab})W_{\rho}y_{ab} \tag{11}$$

对比式(10)和式(11)可以发现，对于同一通道的不同位置，简化的非局部神经网络得到的 α_{ij} 相同，表达的是不同通道的重要性，而具有全局特征的空间注意力机制得到的 α_{ij} 不同，表达的是空间域内各像素点之间的依赖关系。因此，在目标定位方面，加入空间注意力机制(具有全局特征的空间注意力机制)，比加入通道注意力机制(简化的非局部神经网络)可以起到更好的作用。

简化的非局部神经网络的输入特征图没有经过通道融合层，是一个多通道特征，需要一个卷积层通过线性组合的方式捕捉各个通道间的依赖关系。为了消除通道间信息分布对空间注意力机制获取的影响，具有全局特征的空间注意力机制在提取 α_{ij} 之前进行了通道融合，融合特征图 $\boldsymbol{y}_0$ 的通道数为 1，不需要获取通道间的依赖关系，因此本文省略了权重为 W_{ρ} 的卷积层，最终具有全局特征的空间注意力机制中的 α_{ij} 为

$$\alpha_{ij}=f_{\mathrm{softmax}}(y_{ab})y_{ab} \tag{12}$$

在非局部神经网络、简化的非局部神经网络和具有全局特征的空间注意力机制中，α_{ij} 生成的计算式分别为式(9)(10)(12)。

1.3 全局特征聚合

由于 softmax 函数本身有 $\sum_{i}\sum_{j}f_{\mathrm{softmax}}(y_{ij})=1$ 的特性，因此大量 α_{ij} 接近 0。如果 $\boldsymbol{\alpha}$ 与 $\boldsymbol{x}$ 通过点乘的方式进行全局特征聚合，势必导致 $\boldsymbol{z}$ 中元素 z_{ij}^{N} 过小，产生梯度消失，使梯度无法迭代。为了解决该问题，本文采用 $\boldsymbol{\alpha}$ 与 $\boldsymbol{x}$ 中对应元素相加来生成 $\boldsymbol{z}$。

$\boldsymbol{\alpha}$ 乘上初始值为 0 的可学习权重 W_Z，使得 $\boldsymbol{x}$ 向 $\boldsymbol{z}$ 平滑过渡，从而使网络更加容易迭代收敛。

具有全局特征的空间注意力机制的全局特征聚合可以表达为

$$z_{ij}^{N}=W_Z\alpha_{ij}+x_{ij}^{N} \tag{13}$$

式中：z_{ij}^{N} 表示 $\boldsymbol{z}$ 中 N 通道上位置 (i,j) 的点；x_{ij}^{N} 表示 $\boldsymbol{x}$ 中 N 通道上位置 (i,j) 的点。

2　实验结果及分析

为了评估本文提出的具有全局特征的空间注意力机制，对机器视觉中两类基本任务进行实验：①目标分类任务，在 chb_mit 数据集上进行脑电波二分类，在 CIFAR-10 数据集上进行多分类；②目标检测任务，在夜间车辆数据集上进行单目标检测，在 voc2007 数据集上进行多目标检测。

本文各实验环境相同，均为一台操作系统为 Ubuntu 16.04，编程语言为 python3.6、pytorch 1.0，显卡支撑为 CUDA9.0、Cudnn7.0，显卡为 1080Ti 单显卡的计算机。

2.1　chb_mit 数据集上的脑电波二分类

通过短时傅里叶变换将 chb_mit 数据集的脑电波数据转换为 23×128×8 的三维矩阵，其中 23 为采集通道数，128 表示脑电信号在 1～128 Hz 频率范围内，8 表示将 2 s 采集信号均分为 8 段时域。对三维矩阵填充数据 0 到 23×128×128 大小，作为输入数据。每个样本大小为 363.5 kB 左右。

在实验中，发现部分网络在添加具有全局特征的空间注意力机制后出现损失不收敛问题。通过数据分析发现，具有全局特征的空间注意力机制的通道融合图通过 softmax 函数时会出现部分数据过大，导致计算无法进行。这是因为通道融合图的参数通常是成百上千个单通道数据的线性组合，经过 softmax 函数中的指数函数时会产生巨大数。这些巨大数会导致两种情况发生：①32 bit 容纳不下巨大数，导致迭代无法进行；②巨大数经过 softmax 后会产生接近于 1 的结果，根据 softmax 函数所有输出结果总和为 1 的性质，通道融合图的其他数据通过 softmax 后结果接近于 0，发生信息丢失。为了避免产生巨大数，对通道融合层进行了适应性调整，即通道融合图 $\boldsymbol{y}_0$ 的数据统一除以调整参数 s。

实验采用 ResNet18、ResNet50、ResNet101 共 3 种网络，3 种网络的优化器、最大迭代数、学习率、训练批数、s 均分别设置为随机递度下降法、50、0.001、64、8。各网络无预训练权重。

目标分类标准指标有 top-1 准确率 T_1 和 top-5 准确率 T_5。T_1 表示概率最大的结果分类正确则判定预测正确；T_5 表示概率前 5 最大的结果如果存在分类正确则判定预测正确。对于二分类问题，$T_1=T_5$。

本小节实验基准线为没有添加注意力机制的 ResNet18、ResNet50、ResNet101 网络的 T_1，分别为 94.741%、96.570%、96.799%。

(1)不同注意力机制的实验结果。为了对比在所有残差单元加入不同注意力机制对目标分类任务的提升情况，在 ResNet18、ResNet50 和 ResNet101 所有残差单元加入简化的非局部神经网络、通道压缩和空间激励块、压缩激励块和具有全局特征的空间注意力机制，实验结果如表 1 所示。相比没有添加注意力机制的网络，分析可知：在 ResNet18 中加入通道压缩和空间激励块效果最佳，有提升效果的注意力机制包括通道压缩和空间激励块和具有全局特征的空间注意力机制；ResNet50 和 ResNet101 中加入具有全局特征的空间注意力机制效果最佳，有提升效果的注意力机制包括简化的非局部神经网络和具有全局特征的空间注意力机制；所有残差单元加入具有全局特征的空间注意力机制后，ResNet18、ResNet50 和 ResNet101 的 T_1 分别提升了 0.533%、0.839%和 0.763%。

表 1　所有残差单元加入不同注意力机制的实验结果

网络	T_1/%			
	简化的非局部神经网络	通道压缩和空间激励块	压缩激励块	具有全局特征的空间注意力机制
ResNet18	92.950	**96.189**	90.930	95.274
ResNet50	96.723	95.998	94.512	**97.409**
ResNet101	97.142	73.895	96.684	**97.561**

注：表中加粗数据为最优值。

(2)不同残差单元加入具有全局特征的空间注意力机制的实验结果。为了讨论具有全局特征的空间注意力机制加入到特征提取网络的任意残差单元中的效果，分别在 ResNet50、ResNet18 和 ResNet101 的 $c1$、$c2$、$c3$、$c4$ 残差单元和 $c1$～$c4$ 残差单元中加入具有全局特征的空间注意力机制，实验结果如表 2 所示。相比没有添加注意力机制的网络，分析可知：在 ResNet18、ResNet50 和 ResNet101 大部分残差单元加入具有全局特征的空间注意力机制后，分类效果有所提升，所有残差单元中加入具有全局特征的空间注意力机制可以获得更好的分类结果；不同残差单元加入具有全局特征的空间注意力机制后，ResNet18、ResNet50 和 ResNet101 的 T_1 分别最大提升了 0.533%、0.839%、0.763%。

从本小节实验可以看出，在 chb_mit 数据集上进行脑电波二分类，网络 ResNet101＋具有全局特

表 2 不同残差单元加入具有全局特征的空间注意力机制的二分类实验结果

网络	T_1/%				
	c1 残差单元	c2 残差单元	c3 残差单元	c4 残差单元	所有残差单元
ResNet18	95.084	93.674	93.636	94.474	**95.274**
ResNet50	96.875	96.532	96.913	96.608	**97.409**
ResNet101	96.456	96.951	96.532	96.951	**97.561**

注：表中加粗数据为最优值。

征的空间注意力机制检测准确率最高。在较为简单的网络 ResNet18 中，通道压缩和空间激励块表现最好，在较为复杂的网络 ResNet50 和 ResNet101 中，具有全局特征的空间注意力机制表现最好。这是因为数据集单个样本信息较大，浅层网络表达能力弱，特征图包含的相关性不明显，此时加入全局相关性可能对网络产生负面影响。为了验证这一结论，选取小样本图片分类数据集 CIFAR-10 进行多分类任务。

2.2 CIFAR-10 数据集上的多分类任务

CIFAR-10 数据集由 10 类、60 000 张 3 通道 RGB 彩色图像组成，每张图像尺寸为 32×32，每类含有6 000张图像。该数据集共含有50 000张训练图像和10 000张测试图像。每张图像大小为 1.3 kB。

实验采用 ResNet18、ResNet50、ResNet101 共 3 种网络，除 3 种网络的最大迭代数设置为 20、ResNet18、ResNet50 的训练批数设置为 32、ResNet101 的训练批数设置为 16 外，其他参数设置均与 2.1 小节保持相同。

本小节实验基准线为没有添加注意力机制的 ResNet18、ResNet50、ResNet101 网络的 T_1，分别为 83.333%、83.401%、83.383%。

(1)不同注意力机制的实验结果。为了对比 c3 残差单元加入不同注意力机制对目标分类任务的提升情况，在 ResNet18、ResNet50 和 ResNet101 的 c3 残差单元中加入非局部神经网络、简化的非局部神经网络、通道压缩和空间激励块、压缩激励块和具有全局特征的空间注意力机制，实验结果如表 3 所示。相比没有添加注意力机制的网络，分析可知：在 ResNet18 中加入具有全局特征的空间注意力机制效果最佳，有提升效果的注意力机制包括通道压缩和空间激励块、压缩激励块和具有全局特征的空间注意力机制；ResNet50 中加入具有全局特征的空间注意力机制效果最佳，有提升效果的注意力机制包括非局部神经网络、简化的非局部神经网络、通道压缩和空间激励块和具有全局特征的空间注意力机制；ResNet101 中加入具有全局特征的空间注意力机制效果最佳，有提升效果的注意力机制包括压缩激励块和具有全局特征的空间注意力机制；c3 残差单元加入具有全局特征的空间注意力机制后，ResNet18、ResNet50 和 ResNet101 的 T_1 分别提升了 0.484%、0.043%和 0.173%。T_5 等于 1。

表 3 c3 残差单元加入不同注意力机制的实验结果

网络	T_1/%				
	非局部神经网络	简化的非局部神经网络	通道压缩和空间激励块	压缩激励块	具有全局特征的空间注意力机制
ResNet18	83.183	83.167	83.600	82.350	**83.817**
ResNet50	83.418	83.416	83.420	83.322	**83.444**
ResNet101	83.075	83.000	83.325	83.517	**83.556**

注：表中加粗数据为最优值。

(2)不同残差单元加入具有全局特征的空间注意力机制的实验结果。为了讨论具有全局特征的空间注意力机制加入到特征提取网络的任意残差单元中的效果，分别在 ResNet50、ResNet18 和 ResNet101 的 c1、c2、c3、c4 残差单元和 c1～c4 残差单元加入具有全局特征的空间注意力机制，实验结果如表 4 所示。相比没有添加注意力机制的网络，分析可知：在 ResNet18、ResNet50 和 ResNet101 大部分残差单元加入具有全局特征的空间注意力机制后，分类效果有所提升；不同残差单元加入具有全局特征的空间注意力机制后，ResNet18、ResNet50 和 ResNet101 的 T_1 分别最大提升了 0.484%、0.059%和 0.349%。T_5 等于 1。

表 4 不同残差单元加入具有全局特征的空间注意力机制的多分类实验结果

网络	T_1/%				
	c1 残差单元	c2 残差单元	c3 残差单元	c4 残差单元	所有残差单元
ResNet18	83.619	83.717	**83.817**	83.609	83.683
ResNet50	**83.460**	83.380	83.444	83.400	83.325
ResNet101	83.528	83.512	83.556	**83.732**	83.658

注：表中加粗数据为最优值。

从本小节实验可以看出，在 CIFAR-10 数据集上进行多分类任务，较为简单的网络 ResNet18 中加入具有全局特征的空间注意力机制表现最好，较为复杂的网络 ResNet50 和 ResNet101 中所有注意力机制的提升效果均有限。这是因为 ResNet18 已经很好地表达了 CIFAR-10，更深的网络意味着更多的参数，在简单样本上不稳定。因此，各种注意力机制对 ResNet50 和 ResNet101 的提升效果有限。

2.3　夜间车辆数据集上的单目标检测

夜间车辆数据集是用于对夜间车辆单一目标进行检测的数据集，数据集由因特网上收集到的 200 段不等时长的夜间行车记录仪视频采样得到的 2 500张图像组成。其中训练数据集有2 000张图像，测试数据集有 500 张图像。每张图像尺寸为 1 080×1 920，数据集共包含11 920个夜间车辆目标，其中大型目标占样本总数 16.7%，中型目标占样本总数 60%，小型目标占样本总数 23.3%。

实验采用 YOLOv3 网络，优化器、最大迭代数、学习率、训练批数、s 分别设置为随机梯度下降法、100、0.000 5、1、8。网络无预训练权重。

目标检测指标主要有平均精度 $A_{50:90}$、A_{50} 和 A_{75}。平均精度是精准率作为 y 轴和召回率作为 x 轴的坐标轴上的精准率-召回率曲线与 x 轴和 y 轴围成图像的面积。A_{50} 表示当真实框与预测框的交并比大于阈值 50%时的平均精度。A_{75} 表示当真实框与预测框的交并比大于阈值 75%时的平均精度。$A_{50:90}$ 是 A_{50}、A_{60}、A_{70}、A_{80}、A_{90} 的平均值。精准率表示 A_{50} 下的平均精准率，召回率表示 A_{50} 下的平均召回率。

本小节实验基准线为没有添加注意力机制的 YOLOv3 网络[16] 的 $A_{50:90}$、精准率、召回率、A_{50}、A_{75}，分别为 42.282%、80.981%、69.001%、73.809%、43.090%。

(1)不同注意力机制的实验结果。在 YOLOv3 的特征图层 3 的 4 个残差单元中分别添加非局部神经网络、简化的非局部神经网络、通道压缩和空间激励块、压缩激励块和具有全局特征的空间注意力机制，实验结果如表 5 所示。相比没有添加注意力机制的网络，分析可知：以 A_{50}、A_{75} 和 $A_{50:90}$ 为评价标准，YOLOv3 中加入具有全局特征的空间注意力机制效果最好，有提升效果的是压缩激励块和具有全局特征的空间注意力机制，通道压缩和空间激励块在 A_{50} 评价标准下对网络有负面作用，非局部神经网络和简化的非局部神经网络加入后网络不收敛，说明本任务采用的数据集和 YOLOv3 对仅具有全局特征提取的注意力机制不友好；特征图层 3 所有残差单元加入具有全局特征的空间注意力机制后，以 A_{50} 为评价标准，最大提升了 2.104%，以 A_{75} 为评价标准，最大提升了 6.979%。

表 5　YOLOv3 特征图层 3 加入不同注意力机制的实验结果　%

注意力机制	$A_{50:90}$	精准率	召回率	A_{50}	A_{75}
通道压缩和空间激励块	43.536	75.326	76.672	72.176	44.695
压缩激励块	45.884	77.667	76.902	73.998	45.400
具有全局特征的空间注意力机制	**47.817**	79.721	75.282	**75.913**	**50.069**

注：表中加粗数据为最优值。

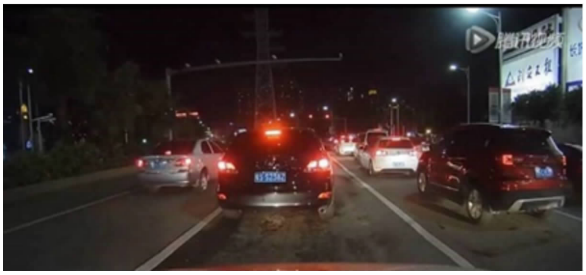
(2)不同特征图层加入具有全局特征的空间注意力机制的实验结果。分别在 YOLOv3 卷积神经网络中特征图层 1、特征图层 2、特征图层 3 和全部特征图层中的残差单元加入具有全局特征的空间注意力机制，结果如表 6 所示。相比没有添加注意力机制的网络，分析可知：YOLOv3 不同特征图层加入具有全局特征的空间注意力机制效果均有所提升；以 A_{50} 和 A_{75} 为评价标准，所有特征图层都加入具有全局特征的空间注意力机制效果最好；所有特征图层都加入具有全局特征的空间注意力机制后，以 A_{75} 为评价标准，提升了 3.860%，以 A_{75} 为评价标准，提升了 11.726%。

表 6　YOLOv3 不同特征图层加入具有全局特征的空间注意力机制的实验结果　%

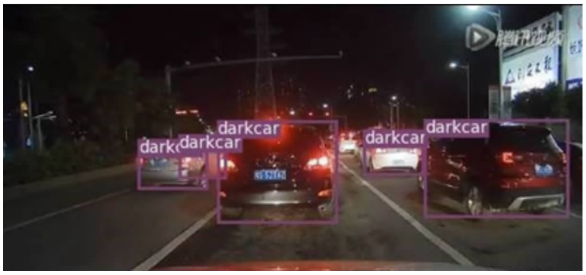
图层	$A_{50:90}$	精准率	召回率	A_{50}	A_{75}
特征图层 1	47.395	81.974	75.003	75.593	50.250
特征图层 2	**50.060**	80.450	76.981	76.716	52.392
特征图层 3	47.817	79.721	75.282	75.913	50.069
所有特征图层	48.486	85.355	74.470	**77.669**	**54.816**

注：表中加粗数据为最优值。

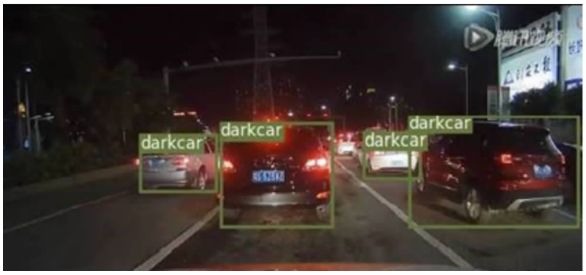
图 3 显示了 YOLOv3 加入不同注意力机制的检测结果，可以看出，相比较原始检测结果，加入压缩激励块可以解决误检测问题。加入具有全局特征的空间注意力机制同时解决了误检和漏检问题，并且车辆预测框更加接近真实车辆，这也映证了具有全局特征的空间注意力机制中包含的全局相关性信息有利于定位和区分目标。



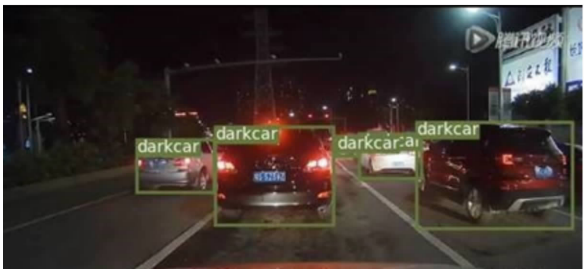
(a)原图



(b)YOLOv3 检测结果



(c)YOLOv3+压缩激励块检测结果



(d)YOLOv3+具有全局特征的空间注意力机制检测结果

图 3 YOLOv3 加入不同注意力机制的检测结果

2.4 voc2007 数据集上的多目标检测

实验采用 YOLOv3 网络,优化器、最大迭代数、学习率、训练批数、s 分别设置为随机梯度下降法、200、0.004、4、8。网络无预训练权重。

本小节实验基准线为没有添加注意力机制的 ResNet50 网络的 $A_{50:90}$、精准率、召回率、A_{50}、A_{75},分别为 49.622%、87.645%、86.961%、74.834%、49.284%。

(1)不同注意力机制的实验结果。在 ResNet50 的 $c4$ 残差单元分别加入非局部神经网络、简化的非局部神经网络、通道压缩和空间激励块、压缩激励块和具有全局特征的空间注意力机制,结果如表 7 所示。相比没有添加注意力机制的网络,分析可知:在 $A_{50:90}$、A_{50} 和 A_{75} 这 3 个评价标准中,加入具有全局特征的空间注意力机制效果都达到最好;以 A_{50} 为评价标准,有提升效果的注意力机制包括非局部神经网络、压缩激励块、具有全局特征的空间注意力机制,加入简化的非局部神经网络后网络不收敛;在 ResNet50 的 $c4$ 残差单元中加入具有全局特征的空间注意力机制后,以 A_{50} 为评价标准,最大提升了 0.778%。

(2)不同残差单元加入具有全局特征的空间注意力机制的实验结果。分别在 ResNet50 的 $c1$、$c2$、$c3$、$c4$ 残差单元和 $c1$~$c4$ 残差单元中加入具有全局特征的空间注意力机制,结果如表 8 所示。相比没有添加注意力机制的网络,分析可知:以 A_{50} 为评价标准,除 $c3$ 残差单元外,在 ResNet50 的其他残差单元中加入具有全局特征的空间注意力机制后,检测效果都有提升,其中提升最大的是在 $c4$ 残差单元加入具有全局特征的空间注意力机制,提升了 0.778%。

表 7 ResNet50 的 $c4$ 残差单元加入不同注意力机制的实验结果 %

注意力机制	$A_{50:90}$	精准率	召回率	A_{50}	A_{75}
全局上下文块	49.736	87.405	87.535	74.977	49.213
通道压缩和空间激励块	47.889	87.188	86.663	73.837	46.499
压缩激励块	50.155	87.359	87.511	75.049	49.041
具有全局特征的空间注意力机制	**50.160**	87.627	87.713	**75.612**	**50.516**

注:表中加粗数据为最优值。

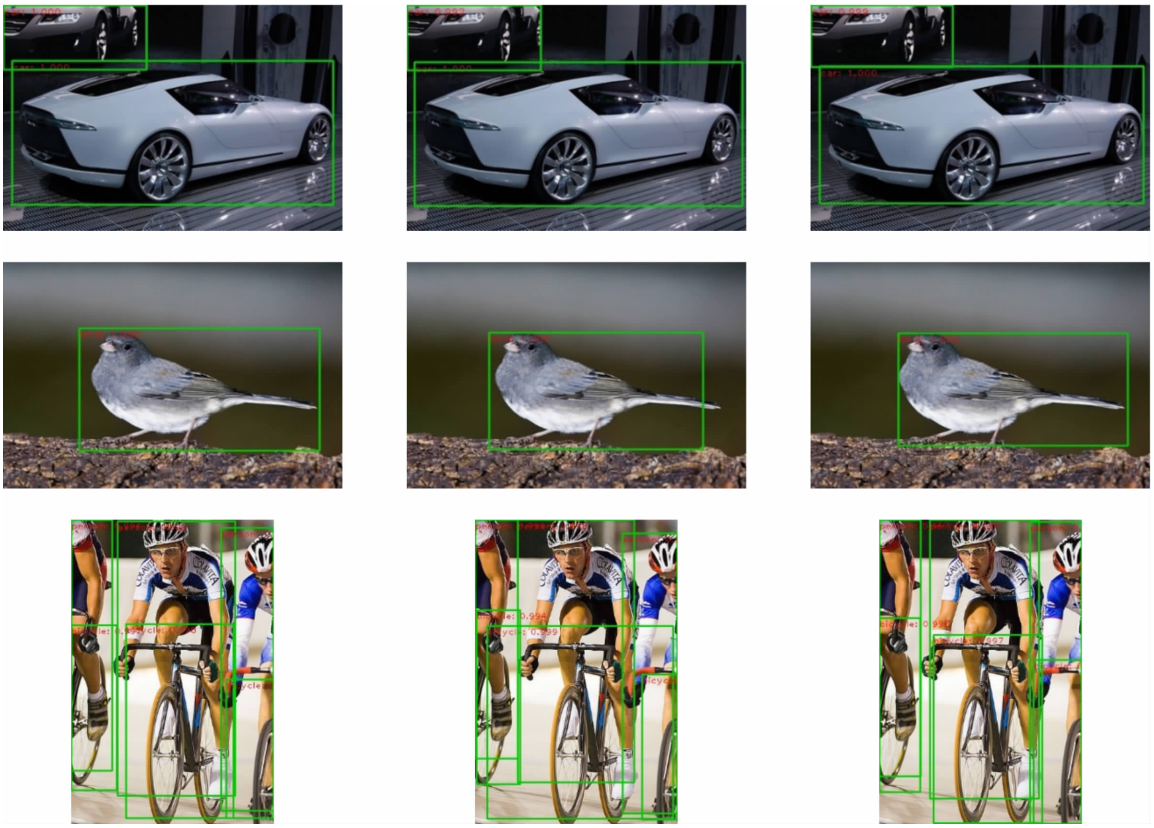
表 8 ResNet50 不同残差单元加入具有全局特征的空间注意力机制的实验结果 %

残差单元	$A_{50:90}$	精准率	召回率	A_{50}	A_{75}
$c1$ 残差单元	49.851	88.067	86.421	75.285	49.660
$c2$ 残差单元	50.203	87.932	86.923	74.986	49.998
$c3$ 残差单元	49.784	87.327	86.915	74.728	49.451
$c4$ 残差单元	**50.160**	87.627	87.713	**75.612**	**50.516**
所有残差单元	49.992	87.361	87.211	75.159	50.144

注:表中加粗数据为最优值。

图 4 是 Faster R-CNN[17] 加入不同注意力机制的检测结果。采用 A_{50} 评价标准对比 Faster R-

CNN 和两个效果最好的网络 Faster R-CNN＋压缩激励块和 Faster R-CNN＋具有全局特征的空间注意力机制。第 1 行图像中，两辆车辆的预测框图不应该发生重叠，Faster R-CNN＋具有全局特征的空间注意力机制可以很好地解决这个问题；第 2 行图像的单目标检测，相比较 Faster R-CNN 较大的预测框和 Faster R-CNN＋压缩激励块较小的预测框，Faster R-CNN＋具有全局特征的空间注意力机制的预测框更贴近物体本身；第 3 行图像的多目标重叠问题，Faster R-CNN＋具有全局特征的空间注意力机制可以更精准地区分自行车和骑手。更加精准的预测框是因为具有全局特征的空间注意力机制中的特征图中加入了空间全局信息，可以有效地关联前景目标和背景环境，更好地定位目标位置。



(a)Faster R-CNN 检测结果　(b)Faster R-CNN＋压缩激励块检测结果　(c)Faster R-CNN＋具有全局特征的空间注意力机制检测结果

图 4　Faster R-CNN 加入不同注意力机制的检测结果

3　结　论

本文提出一种具有全局特征的空间注意力机制。具有全局特征的空间注意力机制通过通道融合层消除了通道信息分布对空间权重提取的影响，通过全局特征提取获得特征图各个点的空间依赖关系。加入具有全局特征的空间注意力机制有利于网络区分前景目标和背景环境，使网络对目标的定位更加精准。本文结论如下。

(1)具有全局特征的空间注意力机制加入通道融合层，通过一个一维卷积核将多通道压缩为单通道信息，消除了通道信息分布对空间权重的影响，并且使用多特征表达，比单特征表达更具鲁棒性。

(2)具有全局特征的空间注意力机制将简化的非局部神经网络中的矩阵乘法变为对应元素相乘，从而将简化的非局部神经网络从通道域注意力机制变为空间域注意力机制，更能体现注意力权重的细节。

(3)具有全局特征的空间注意力机制对通道融合图中的数据统一除以调整参数 s。解决了实验中产生巨大数的问题，使具有全局特征的空间注意力机制更加稳定。

(4)在目标分类和目标检测任务上，具有全局特征的空间注意力机制多数情况下表现优于简化的非局部神经网络和通道压缩和空间激励块，并且不会出现损失不收敛和准确率骤降的问题，与目前先进

的注意力机制压缩激励块相比，具有全局特征的空间注意力机制的计算量和参数量更小，稳定性和有效性更优。

今后将研究具有全局特征的空间注意力机制对目标分割任务的促进作用。

参考文献：

[1] GEHRING J, AULI M, GRANGIER D, et al. A convolutional encoder model for neural machine translation [C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: Volume 1 Long Papers. Stroudsburg, PA, USA: Association for Computational Linguistics, 2017: 123-135.

[2] GEHRING J, AULI M, GRANGIER D, et al. Convolutional sequence to sequence learning [C]// Proceedings of the 34th International Conference on Machine Learning. Princeton, NJ, USA: International Machine Learning Society (IMLS), 2017: 2029-2042.

[3] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]// Proceedings of the 31st Annual Conference on Neural Information Processing Systems. Vancouver, Canada: NIPS, 2017: 5999-6009.

[4] WANG F, JIANG M Q, QIAN C, et al. Residual attention network for image classification [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6450-6458.

[5] HU H, GU J Y, ZHANG Z, et al. Relation networks for object detection [C] // Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 3588-3597.

[6] HUANG Z L, WANG X G, HUANG L C, et al. CCNet: criss-cross attention for semantic segmentation [C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2019: 603-612.

[7] YUAN Yuhui, WANG Jingdong. OCNet: object context network for scene parsing [EB/OL]. (2019-01-22)[2019-12-06]. https://arxiv.org/abs/1809.00916.

[8] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7132-7141.

[9] CAO Yue, XU Jiarui, LIN S, et al. GCNet: non-local networks meet squeeze-excitation networks and beyond [EB/OL]. (2019-04-25)[2019-12-06]. https://arxiv.org/abs/1904.11492.

[10] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 7794-7803.

[11] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]// Proceedings of the 15th European Conference on Computer Vision. Cham, Germany: Springer, 2018: 3-19.

[12] ROY A G, NAVAB N, WACHINGER C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks [C]// Proceedings of the 21st International Conference on Medical Image Computing and Computer Assisted Intervention. Cham, Germany: Springer, 2018: 421-429.

[13] BUADES A, COLL B, MOREL J M. A non-local algorithm for image denoising [C]// Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2005: 60-65.

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

[15] SIMONYAN K, ZISSERMAN. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10)[2019-12-06]. https://arxiv.org/abs/1409.1556.

[16] REDMON J, FARHADI A. Yolov3: an incremental improvement [EB/OL]. (2018-04-08)[2019-12-06]. https://arxiv.org/abs/1804.02767.

[17] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C] // Proceedings of the 29th Annual Conference on Neural Information Processing Systems. Vancouver, Canada: NIPS, 2015: 91-95.

（编辑 陶晴）