

## 卷积神经网络中的注意力机制综述

张宸嘉, 朱磊, 俞璐

陆军工程大学 通信工程学院, 南京 210007

**摘要:** 注意力机制因其优秀的效果与即插即用的便利性, 在深度学习任务中得到了越来越广泛的应用。主要着眼于卷积神经网络, 对卷积网络注意力机制发展过程中的各种主流方法进行介绍, 并对其核心思想与实现过程进行提取与总结, 同时对每种注意力机制方法进行实现, 针对同型号辐射源设备实测数据进行对比实验与结果分析, 并依据主流方法的思想与实验的结果总结并阐述了卷积网络中的注意力机制的研究现状与未来其发展方向。

**关键词:** 注意力机制; 卷积神经网络; 发展方向

**文献标志码:** A **中图分类号:** TP391 **doi:** 10.3778/j.issn.1002-8331.2105-0135

### Review of Attention Mechanism in Convolutional Neural Networks

ZHANG Chenjia, ZHU Lei, YU Lu

School of Communication Engineering, Army Engineering University of PLA, Nanjing 210007, China

**Abstract:** Attention mechanism is widely used in deep learning tasks because of its excellent effect and plug and play convenience. This paper mainly focuses on convolution neural network, introduces various mainstream methods in the development process of convolution network attention mechanism, extracts and summarizes its core idea and implementation process, realizes each attention mechanism method, and makes comparative experiments and results analysis on the measured data of the same type of emitter equipment. According to the main ideas and experimental results, the research status and future development direction of attention mechanism in convolutional networks are summarized.

**Key words:** attention mechanism; convolutional neural network; development direction

随着信息技术的发展, 海量繁杂的信息向人们不断袭来, 信息无时无刻充斥在四周。然而人类所能接收的信息则是有限的, 科研人员发现人类视觉系统在有限的视野之下却有着庞大的视觉信息处理能力。在处理视觉数据的初期, 人类视觉系统会迅速将注意力集中在场景中的重要区域上, 这一选择性感知机制极大地减少了人类视觉系统处理数据的数量, 从而使人类在处理复杂的视觉信息时能够抑制不重要的刺激, 并将有限的神经计算资源分配给场景中的关键部分, 为更高层次的感知推理和更复杂的视觉处理任务(如物体识别、场景分类、视频理解等)提供更易于处理且更相关的信息<sup>[1]</sup>。借鉴人类视觉系统的这一特点, 科研人员提出了注意力机制的思想。对于事物来说特征的重要性是不同的, 反映在卷积网络中即每张特征图的重要性是具有差异性的<sup>[2]</sup>。注意力机制的核心思想是通过一定手段获取到每张特征图重要性的差异, 将神经网络的计算资源更多地投入更重要的任务当中, 并利用任务结果反向指导特征图的权重更新, 从而高效快速地完成相应任务<sup>[3]</sup>。

近两年, 注意力模型(Attention Model)被广泛使用在自然语言处理、图像识别、语音识别等各种不同类型的深度学习任务当中, 如 SANet<sup>[4]</sup>(Structure-Aware Networks)与 VSG-Net<sup>[5]</sup>(Visual-Spatial-Graph Network)等, 是深度学习技术中最值得关注与深入了解的核心技术之一<sup>[6]</sup>。

图1展示了人类在看到一幅图像时是如何高效分配有限注意力资源的, 其中红色区域表明视觉系统更关

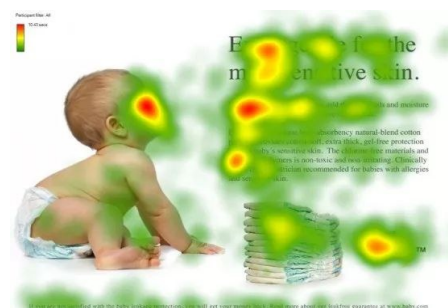


图1 视觉注意力机制

Fig.1 Mechanism of visual attention

**作者简介:** 张宸嘉(1996—), 男, 硕士研究生, 主要研究方向为辐射源识别、深度学习、模式识别; 朱磊(1973—), 通信作者, 男, 博士生导师, 主要研究方向为网络运维、网络规划, E-mail: 1075649789@qq.com。

**收稿日期:** 2021-05-11 **修回日期:** 2021-07-09 **文章编号:** 1002-8331(2021)20-0064-09

注的目标。很明显对于图1所示的场景,人们会把注意力更多投入到人的脸部、文本的标题以及文章首句等位置。

深度学习中的注意力机制从本质上讲和人类的选择性视觉注意力机制类似,核心目标也是从众多信息中选择出对当前任务目标更为关键的信息<sup>[7]</sup>。

1 卷积网络中注意力机制研究现状

注意力机制最初在2014年作为RNN(Recurrent Neural Network)中编码器-解码器框架的一部分来编码长的输入语句,后续被广泛运用在RNN中<sup>[8]</sup>。例如在机器翻译中通常是用一个RNN编码器读入上下文,得到一个上下文向量(RNN的最后一个隐状态),一个RNN解码器以这个隐状态为起始状态,依次生成目标的每一个单词<sup>[9]</sup>。但这种做法的缺点是:无论之前的上下文有多长,包含多少信息量,最终都要被压缩成一个几百维的向量。这意味着上下文越大,最终的状态向量会丢失越多的信息。输入语句长度增加后,最终解码器翻译的结果会显著变差。事实上,因为上下文在输入时已知,一个模型完全可以在解码的过程中利用上下文的全部信息,而不仅仅是最后一个状态的信息,这就是注意力机制的基础思想。

但是在CNN(Convolutional Neural Networks)中,没有了状态,注意力机制作用在特征图之上,试图获取特征图中的可用注意力信息,达到更好的任务效果<sup>[10]</sup>。接下来本文将对CNN中的注意力获取手段进行详细介绍和分析。

1.1 单路注意力

1.1.1 SE-Net(Squeeze and Excitation)

2018年内,CVPR(Computer Vision and Pattern Recognition)收录的论文中提出了SE-Net(Squeeze-and-Excitation Network),SE-Net是Momenta胡杰团队(WMW)提出的新的网络结构,该团队利用SE-Net,获得了ImageNet 2017竞赛图像分类任务的冠军,在ImageNet数据集上将top-5 error降低到2.251%,对比于以往的最好成绩2.991%有了较大的提升<sup>[11]</sup>。

SE-Net中的关键结构SE-Net block利用了注意力机制的思想,显式地建模特征图之间的相互依赖关系,并通过学习的方式来自适应地获取到每张特征图的重要性<sup>[12]</sup>,然后依照这个重要程度去对原数据进行更新。SE-Net通过这种方式提升有用的特征重要程度同时降低无用特征的重要性,并以不同通道的重要性为指导,将计算资源合理地投入不同通道当中。实验证明该方法在各种任务中都有着不错的表现。

通俗地说SENet的核心思想在于通过网络根据损失函数值loss去学习特征权重,使得对于任务更为效果明显的特征图权重变大,无效果或效果不明显的特征图权重变小的方式来训练模型从而达到更好的结果<sup>[13]</sup>。

SE-Net block并不是一个完整的网络结构,而是一个即插即用的轻量级模块,通过将此模块嵌入网络之中,可以在小幅度提升参数数量的代价下更加合理地分配神经网络的计算资源,大幅提升网络性能。

在SE-Net block中,每张特征图通过全局平均池化操作进行挤压,将每一张特征图挤压成一个实数(见公式(1)),这个实数具有特征图上的全局信息,每张特征图的挤压结果组合成一个向量作为每组特征图的权重,其中H和W分别为特征图的高和宽,u为卷积后的结果,z为对应特征图的全局注意力信息,将此向量通过全连接层与激活函数(见公式(2)),训练结果用来放大对于识别任务更加重要特征图的权重,缩小不重要特征图的权重,其中σ为relu激活函数,δ代表sigmoid激活函数,W<sub>1</sub>与W<sub>2</sub>代表两个不同的全连接操作。得到的向量s代表每张特征图的重要性程度。向量s通过公式(3)激励原特征图,指导特征图不断向着有利于识别任务的方向更新,挤压激励操作结构如图2所示。

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j)$$
$$s = \sigma(W_2 \delta(W_1 z))$$
$$x_c = s_c u_c$$

(1)

(2)

(3)

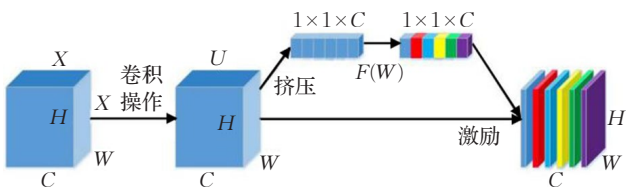


图2 挤压激励操作

Fig.2 Squeeze and excitation operation

1.1.2 ECA-Net(Efficient Channel Attention)

2020年,CVPR收录的论文中提出了ECA-Net<sup>[14]</sup>(Efficient Channel Attention Network)来对SE-Net进行改进,它实现了对SE-Net block的改进,提出了一种不降维的局部跨信道交互策略(ECA block)和自适应选择一维卷积核大小的方法,通过一维卷积层汇总跨信道信息的方法获取更加精确的注意力信息。ECA block的思想建立在作者认为跨通道的信息交互是很有必要的,而SE-Net block只注重通道内部信息的综合,没有考虑到相邻信道信息的重要性。ECA block的结构如图3所示。公式(4)表示最终的权重是综合了各个相邻通道的信息获得,其中σ为激活函数,y<sub>i</sub>代表通

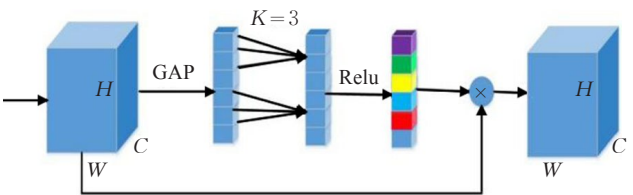


图3 ECA-Net模块

Fig.3 ECA-Net module

道,  $w_i$  为通道  $y_i$  的权重,  $\Omega$  代表与  $y_i$  相邻的  $k$  个通道,  $k$  的值是随着学习自适应变化的。为了实现这一想法, 作者利用了一维卷积层来进行实现, 通过核为  $k$  的一维卷积对通道与其相邻的  $k-1$  个通道信息进行综合, 如公式(5)所示,  $C1D_k$  表示核为  $k$  的一维卷积操作,  $y$  表示通道。

$$\omega_i = \sigma\left(\sum_{j=1}^k \omega_i^j y_i^j, y_i^j \in \Omega_i^k\right) \quad (4)$$

$$\omega = \sigma(C1D_k(y)) \quad (5)$$

## 1.2 多路注意力

### 1.2.1 SK-Net(Selective Kernel Networks)

2019年, CVPR收录的论文中提出了SK-Net(Selective Kernel Networks), SK-Net基于卷积核的注意力机制, 即卷积核的重要性, 即不同的图像经过不同卷积核的重要性是不同的, 其结构如图4所示。整个SK-Net结构由Split、Fuse、Select三部分组成<sup>[15]</sup>。Split的任务是将输入的特征图  $X$  进行不同卷积核大小的卷积操作。如图4所示, 对  $X$  进行 Kernel  $3 \times 3$  和 Kernel  $5 \times 5$  的卷积操作, 得到输出  $U_1$  与  $U_2$ 。而在Fuse部分将对  $U_1$  与  $U_2$  进行 element-wise summation, 得到输出特征图  $U$ , 通过全局平均池化与全连接层获取特征图的注意力信息, 并创建了一个紧凑的特征  $z \in \mathbb{R}^{d \times 1}$ , 以便为精确和自适应选择提供指导如公式(6)所示, 其中  $\delta$  是 ReLU 函数,  $B$  表示批量标准化,  $W$  表示全连接层且  $W \in \mathbb{R}^{d \times C}$ 。公式(7)表明为了研究  $d$  对模型效率的影响, 文章使用下降参数  $r$  来控制其值,  $L$  表示  $d$  的最小值。在Select部分中, 将这个紧凑特征  $z$  向量重新分为两个(本文情况)或多个(更多的情况)特征向量, 然后分别与相应的split之后的特征图进行相应通道的相乘操作, 然后再通过这种加权共同构成输入到下一个神经元的特征向量。两个特征向量  $a_c$ 、 $b_c$  的生成如公式(8)、(9)所示, 其中  $A, B \in \mathbb{R}^{C \times d}$ ,  $a, b$  分别表示  $U_1$  与  $U_2$  的注意力向量,  $A_c \in \mathbb{R}^{1 \times d}$  表示  $A$  的第  $c$  行,  $a_c$  表示  $a$  的第  $c$  个元素值, 对于向量  $B$  同理。

$$z = F_{fc}(s) = \sigma(B(Ws)) \quad (6)$$

$$d = \max(C/r, L) \quad (7)$$

$$a_c = \frac{e^{A_c z}}{e^{A_c z} + e^{B_c z}} \quad (8)$$

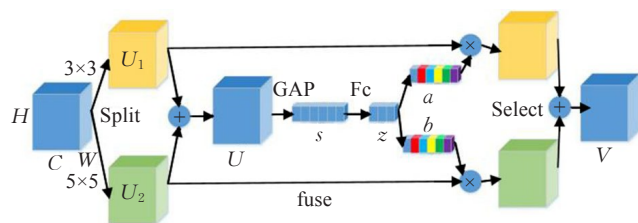


图4 SK-Net 模块

Fig.4 SK-Net module

$$b_c = \frac{e^{B_c z}}{e^{A_c z} + e^{B_c z}} \quad (9)$$

### 1.2.2 ResNeSt

2020年, 亚马逊、加州大学戴维斯分校的张航、李沐、Smola 等研究者进一步改进了ResNet<sup>[16]</sup>(Deep Residual Network), 提出了ResNeSt, 其中利用ResNet、SE-Net与SK-Net的思想, 提出了Split-Attention block<sup>[17]</sup>。在ResNeSt block中, 整体大框架运用了残差网络的结构, 通过将网络的输入Input输入  $k$  个Cardinal分支当中, 公式(10)表述了每个Cardinal的输入, 其中  $R$  代表每个Cardinal中split后的分支数,  $k$  代表第  $k$  个Cardinal,  $U$  代表着split后每个分支的输入。公式(11)表述了每个Cardinal模块的输出,  $V$  代表携带了通道权重的Cardinal输出,  $a(c)$  是由 softmax 计算得到的权重, 计算方法如公式(12)所示, 其中  $G$  代表每个split的权重。在经过Cardinal模块后对最后的  $k$  个输出进行拼接, 以达到综合  $k$  个Cardinal输出信息的目的, 如公式(13)所示, 并将拼接后的输出与原本的输入进行 element-wise summation, 得到最后的输出, 其结构如图5所示。

$$\hat{U}^k = \sum_{i=R(k-1)+1}^{Rk} U_i \quad (10)$$

$$V_c^k = \sum_{i=1}^R a_i^k(c) U_{R(k-1)+i} \quad (11)$$

$$a_i^k(c) = \begin{cases} \frac{\exp(G_i^c(s^k))}{\sum_{j=1}^R G_j^c(s^k)}, & R > 1 \\ \frac{1}{1 + \exp(-G_i^c(s^k))}, & R = 1 \end{cases} \quad (12)$$

$$V = \text{Concat}\{V^1, V^2, \dots, V^k\} \quad (13)$$

在每个Cardinal中, 又利用了SE-Net与SK-Net的思想, 使用Split模块对每个Cardinal的输入切分为  $r$  个分支, 通过SE-Net中的挤压激励操作获取每个分支的注意力信息作为Fuse模块, 最后在Select模块中使得具有注意力信息的向量与其对应的分支特征图相乘, 并通过 element-wise summation 综合  $r$  个分支的输出作为最终的Cardinal输出, Cardinal结构如图6所示。

### 1.2.3 CBAM(Convolutional Block Attention Module)

2018年, ECCV(European Conference on Computer Vision)收录的论文中提出了卷积注意力模块CBAM(Convolutional Block Attention Module Network), 它的创新在于, 它认为对于卷积网络中的特征图来说, 不仅通道中蕴含着丰富的注意力信息<sup>[18]</sup>, 通道内部, 即特征图像素点间也具有大量的注意力信息, 而以往的注意力机制只关注了通道上的注意力信息, 这对于空间上的注意力信息是一种浪费<sup>[19]</sup>。CBAM通过构建两个子模块, 空间注意力模块SAM(Spatial Attention Module), 通道注意力模块CAM(Channel Attention Module)分别



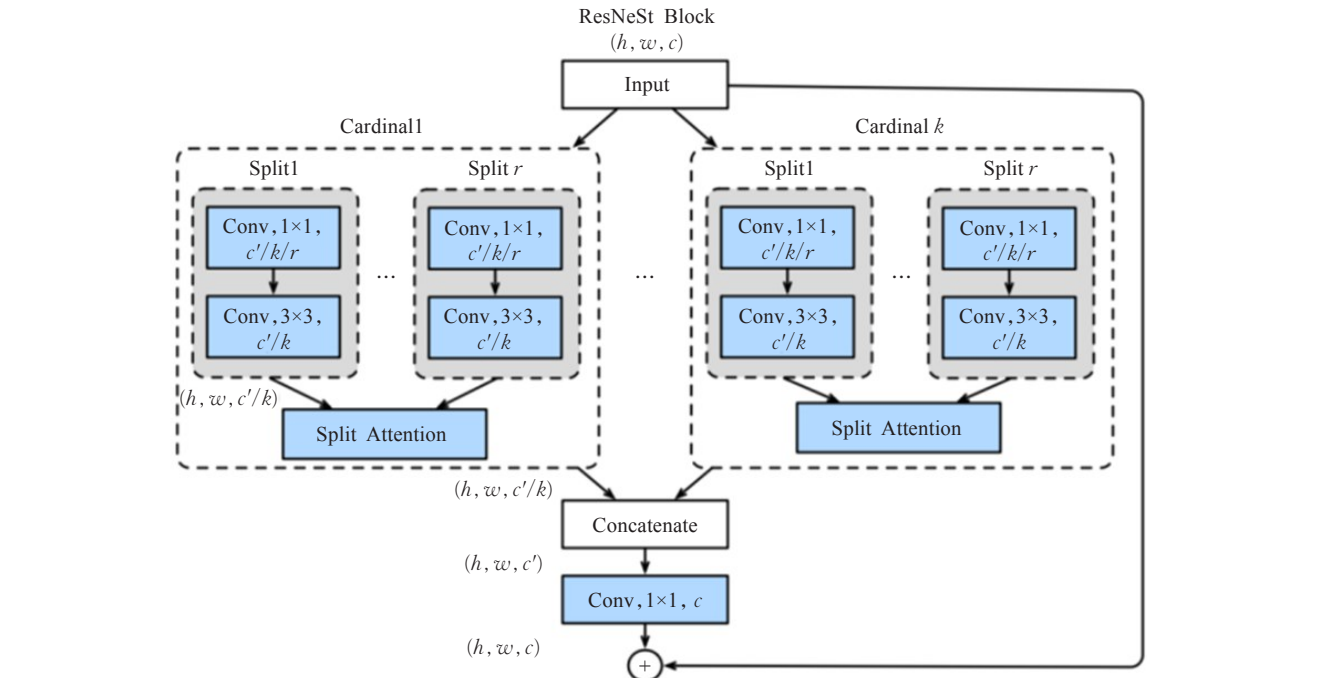


图5 ResNeSt block 结构

Fig.5 ResNeSt block structure

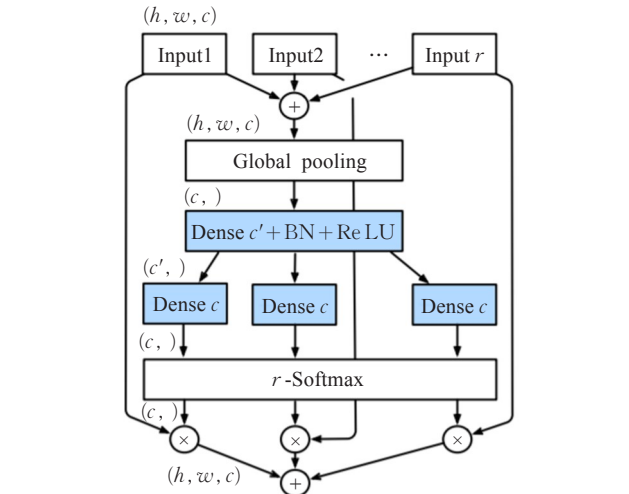


图6 Cardinal 结构

Fig.6 Cardinal structure

汇总空间和通道两方面的注意力信息,并将信息进行一定程度的综合,从而获得更全面可靠的注意力信息<sup>[20]</sup>,对计算资源的分配进行更合理的指导,其结构如图7所示。

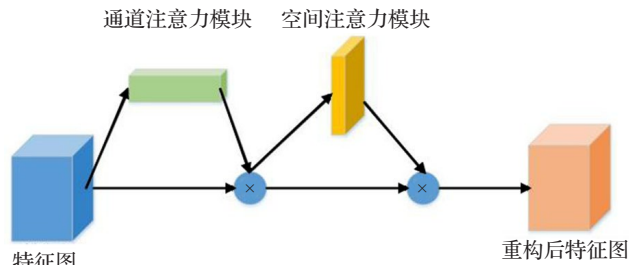


图7 CBAM 结构

Fig.7 CBAM structure

通道注意力模块将输入的特征图  $F(H \times W \times C)$  分别经过基于宽和高的 global max pooling(全局最大池化)和 global average pooling(全局平均池化),得到两个  $1 \times 1 \times C$  的特征图,接着,再将它们分别送入一个两层的神经网络(MLP),第一层神经元个数为  $C/r$  ( $r$  为减少率),激活函数为 Relu,第二层神经元个数为  $C$ ,这个两层的神经网络是共享的。而后,将 MLP 输出的特征进行基于 element-wise 的加和操作,再经过 sigmoid 激活操作,生成最终的通道注意力特征,即公式(14)中的  $M_c(F)$ ,其中  $F_{avg}$  为特征图经过全局平均池化的结果,  $F_{max}$  为经过全局最大池化的结果,整个通道门注意力模块结构如图8所示。

$$M_c(F)=\sigma(W_1(W_0(F_{avg}^c))+W_1(W_0(F_{max}^c))) \tag{14}$$

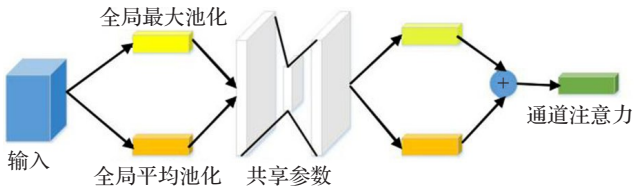


图8 CAM 结构

Fig.8 CAM structure

空间注意力模块将通道注意力模块输出的特征图  $F$  作为本模块的输入特征图<sup>[21]</sup>。首先做一个基于通道的全局最大池化和全局平均池化,得到两个尺寸为  $H \times W \times 1$  的特征图,然后将这两个特征图基于通道做拼接操作。然后经过一个  $7 \times 7$  卷积操作,降维为  $H \times W \times 1$ ,再经过 sigmoid 生成空间注意力特征,即公式(15)中的  $M_s(F)$ ,其中  $f$  代表卷积操作,  $[ ]$  代表通道拼接操作,最

后将该向量和该模块的输入特征图做乘操作,得到最终生成的特征。整个空间注意力模块结构如图9所示。

$$M_s(F) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (15)$$

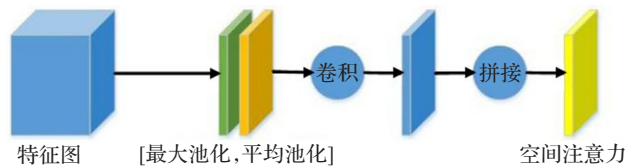


图9 SAM结构

Fig.9 SAM structure

#### 1.2.4 双注意力网络(Dual Attention Network)

2019年CVPR收录的论文中提出了DA-Net(Dual Attention Network),与CBAM相似的是,它的思想也是综合通道和空间两路的注意力信息,但不同的是CBAM的两路注意力信息的获取是串行的,而DA-Net中的两路注意力信息的获取是并行的,且获取注意力信息的方式也有很大差别。DA-Net从通道与空间两个分支通过对特征图进行矩阵操作构建特征图的相关性矩阵 $S$ 和 $X$ ,两个矩阵分别用来表征通道之间的相关性和通道内像素点之间的相关性,用此矩阵对特征图的更新进行引导,增大关键特征的权重,使得将更多的注意力放在更易于进行区分的优秀特征之上。双注意力模块结构如图10所示。

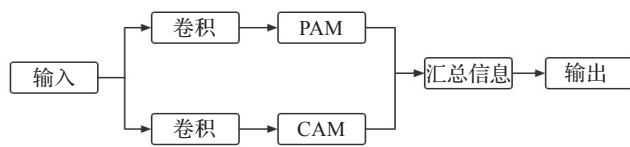


图10 双注意力网络

Fig.10 Dual attention network

其中PAM(Position Attention Module)是空间分支,其结构如图11所示,CAM(Channel Attention Module)是通道分支,其结构如图12所示,这两个分支通过对于

特征图的处理分别构建出了关于特征图通道与空间位置的相关性矩阵 $X$ (尺寸为 $C \times C$ )与 $S$ (尺寸为 $(H \times W) \times (H \times W)$ ),其中 $H$ 、 $W$ 、 $C$ 分别为特征图的高、宽与通道数,并用此两个相关性矩阵来引导特征图不同通道与空间位置权重的更新方向,DA-Net捕捉了空间和通道维度中的全局特征依赖关系,使用位置注意力模块来学习特征的空间相互依赖性,通道注意力模块来模拟通道相互依赖性<sup>[22]</sup>。

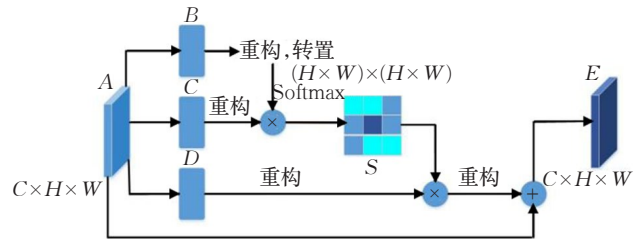


图11 空间分支

Fig.11 Spatial branching

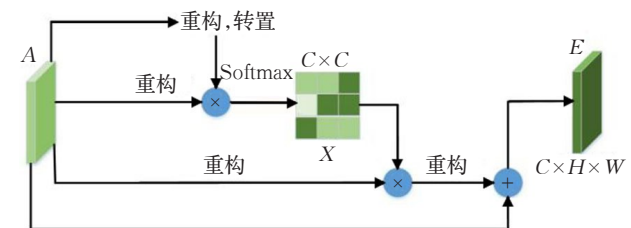


图12 通道分支

Fig.12 Channel branching

#### 1.2.5 金字塔特征注意力网络(Pyramid Feature Attention Network)

2019年CVPR收录的论文中提出了金字塔特征注意力网络,同样是利用特征图在通道间与通道内部像素点都富含大量的注意力信息的思想<sup>[23]</sup>,其结构如图13所示。其中CA模块(Channel Attention Module)为通道注意力模块,CA分支的结构与SE-Net的思想是相同的,都是通过全局平均池化提取通道注意力信息,利用

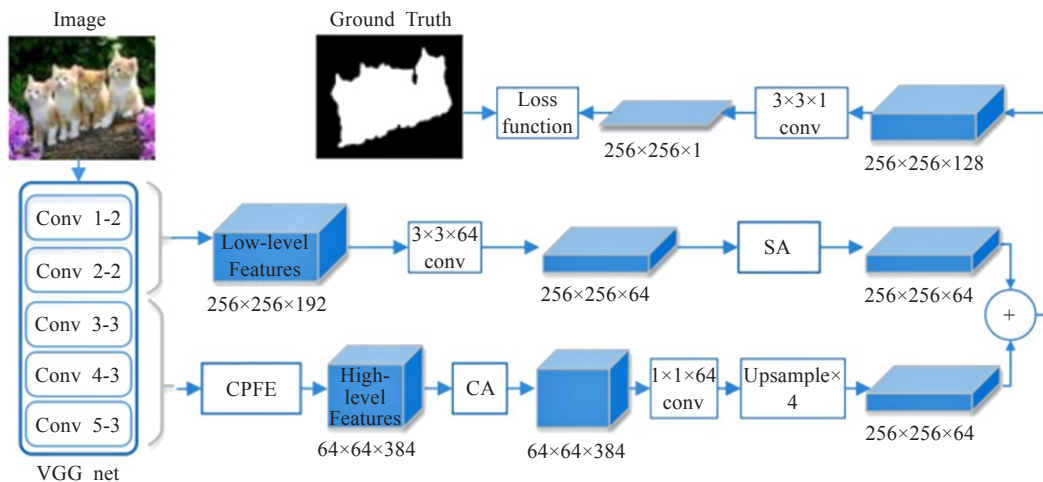


图13 金字塔特征注意力网络

Fig.13 Pyramid feature attention network

全连接获取各个通道的权重,如图14所示,SA模块(Spatial Attention Module)为空间注意力模块,它利用了交替的卷积核相同的卷积层来提取通道内部像素位置之间的注意力信息,获得通道内部不同像素位置之间的相关性与重要程度等信息<sup>[24]</sup>,其结构如图15所示。两个模块分别从通道与空间两个方向提取特征图中的注意力信息,提取不同通道与空间中不同像素位置的权重信息,并对特征图进行自适应的更新。

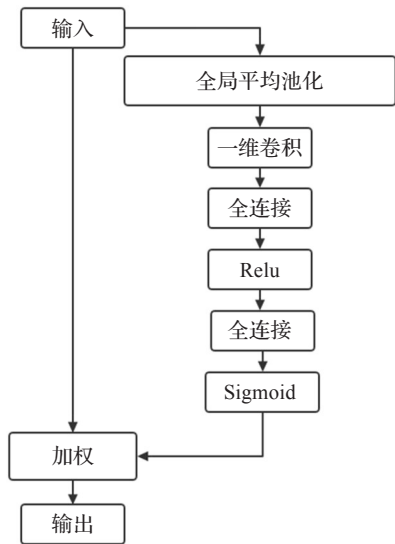


图14 通道注意力模块  
Fig.14 Channel attention module

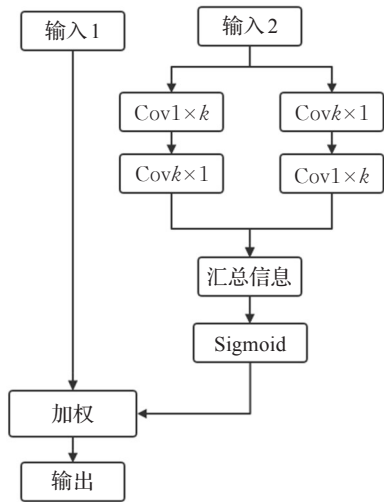


图15 空间注意力模块  
Fig.15 Spatial attention module

2 卷积网络中注意力机制实验分析

2.1 实验数据集采集

本次实验中采用的是实验室采集的5类同型号电台辐射源实测数据,实验目标是运用此数据进行同型号的辐射源个体识别任务。

在实验室环境当中,利用6台同型号USRP(Universal Software Radio Peripheral)设备,其中5台作为发送设

备,1台作为接收设备,5台辐射源设备同时发送主频为1 GHz的辐射源信号,将所接收到的5台辐射源设备的信号通过LabVIEW软件进行数据处理,接收端接收的信号如图16所示,其中白线为I路信号振幅,红线为Q路信号振幅。

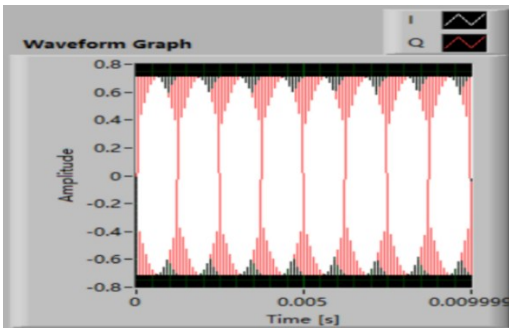


图16 接收端接收的信号  
Fig.16 Signal received by receiver

将经过处理后的5类辐射源信号作为数据集,分别以传统手动特征提取方法,普通深度神经网络方法,本文提出的深度注意力神经网络方法进行辐射源个体识别实验。输入网络中数据的尺寸为2×128,样本数为100 000,其中90%作为训练集,10%作为测试集,网络的batch size为128,训练100轮次,损失函数为交叉熵函数,使用Adam算法来优化网络。

2.2 深度网络模型搭建

基于注意力机制的轻量与即插即用的特性,设计实验所用的深度网络模型为数据预处理层、特征提取层与数据分类层的组合,其结构如图17所示。其中在数据预处理层对信号进行降噪处理,特征提取层通过卷积层与注意力模块的组合实现信号特征的合理有效提取。保持网络结构不变,仅替换不同的注意力模块,实现对5类同型号电台辐射源信号的分类任务。

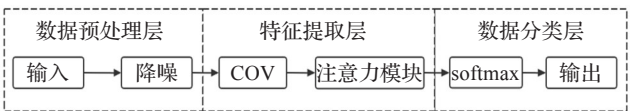


图17 网络模型结构  
Fig.17 Network model structure

2.3 仿真实验与分析

本文针对五个同型号电台辐射源信号作为数据集进行仿真实验,利用本文介绍的七种不同注意力模块,在相同网络条件下分别对每一种注意力模块与不使用注意力模块对数据集进行分类任务,实验结果如图18和表1所示。

从实验结果中可以看出,对比普通神经网络,当在网络中使用注意力模块时,分类任务将更快地达到收敛,且具有更优的分类效果,不使用注意力模块时识别率为83.52%,而在使用注意力模块后分类效果都有着



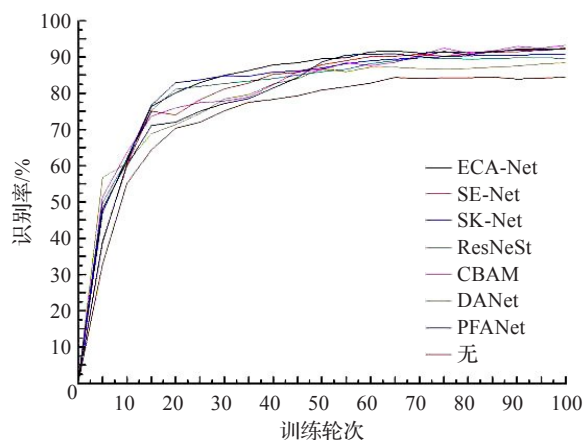


图 18 实验结果

Fig.18 Experimental results

表 1 实验结果

Table 1 Experimental results

注意力方法	识别率/%	训练时间/s
无	84.52	11
SE-Net	92.21	11
ECA-Net	92.28	11
SK-Net	90.66	13
ResNeSt	89.70	13
CBAM	93.23	12
DANet	88.37	12
PFANet	92.47	12

不同程度的提升,其中当使用CBAM模块时效果最优为93.23%(+9.71%)。而在提升分类任务识别效果与收敛速度的同时,注意力模块因其轻量的特征,仅给网络带来少量参数量的提升,整体的训练速度也几乎保持不变。

综上所述,注意力机制在为网络带来极小负担的条件下对网络的速度与任务效果带来大幅度的提升,同时其即插即用的特性为程序设计人员带来了极大的便利性,是深度学习中高效且实用性较高的机制之一。

3 卷积网络中注意力机制展望

近几年,注意力机制的思想被广泛应用在各种深度学习任务中,如计算机视觉、图像分割、自然语言处理等<sup>[25]</sup>。大量实验证明了注意力机制是行之有效且节省资源的,当注意力机制的思想运用于卷积神经网络中时,需要着眼于卷积网络中所特有的特征图中的关键信息。当前注意力机制的主流方法是将特征图中的潜在注意力信息进行深度挖掘,最常见的是通过各种手段获取各个特征图通道间的通道注意力信息与特征图内部像素点之间的空间注意力信息,获取的方法也包括但不仅限于卷积操作,矩阵操作构建相关性矩阵等,其共同的目的是更深层次、更全面地获取特征图中完善的注意力信息<sup>[26]</sup>,于是如何更深地挖掘,从哪里去挖掘特征图的注意力信息,将极有可能成为未来注意力方法发展的方向之一。

目前,获取注意力的方法基本基于通道间的注意力信息、空间像素点之间的注意力信息和卷积核选择的注意力信息,是否能够从新的方向去获取特征图更丰富的注意力信息,或者以新的方式或手段去获取更精准的注意力信息也是未来需要关注的一个重点<sup>[27]</sup>。

ECA-Net论文中的实验证明了跨通道的信息交互对于注意力信息的获取是有积极作用的,这也从侧面验证了不同通道之间并不是相互独立的,其内部是存在许多有利的有价值的信息的<sup>[28]</sup>,那么着眼于不同通道内部的其他信息的提取,如不同特征图中像素点的空间注意力信息或其他跨通道信息是否对于获取更加精准的注意力分布有着正确的导向作用也是一个值得探索的方向<sup>[29]</sup>。

注意力机制作为一个轻量级的模块<sup>[30]</sup>,有着即插即用的特点,但是即使其本身参数量并不高<sup>[31]</sup>,在深度学习一些任务当中,注意力模块往往会被反复多次的调用,当注意力模块调用次数过多时仍然会对网络整体造成一定的负担<sup>[32]</sup>,如何优化模块结构,降低模块参数量或减少模块调用次数,更快地获取更精准的注意力信息,对于以后注意力机制在其他任务中的推广有着举足轻重的作用<sup>[33]</sup>,也是未来需要研究的重要内容之一。

卷积网络中的注意力机制的核心在于深度挖掘特征图中所含有的信息<sup>[34]</sup>,而目前所发现的注意力获取渠道相对较少,但是注意力机制已经被广泛证明其针对大量深度学习任务不仅具有参数量小<sup>[35]</sup>,即插即用的便捷性<sup>[36]</sup>,还可以较为明显地提升任务效果。说明未来对于注意力机制的深度研究是必要且意义非凡的<sup>[37]</sup>,将对深度学习任务产生重大的影响。

随着信息技术的不断发展,人类必将面临着大量而繁杂的信息,针对如此庞杂的信息去完成各项深度学习任务将变得更为困难<sup>[38]</sup>。当数据量无法任意改变的情况下,如何高效率地完成任务就变得尤为重要<sup>[39]</sup>。注意力机制便是提升深度学习任务效率的重要方法之一。当深度学习方法较为低效时,在深度学习任务中引入注意力机制将会实现“曲线救国”<sup>[40]</sup>,利用其低成本、高收益<sup>[41]</sup>的特点,大幅提升信息处理的效率,在未来的深度学习任务中大放异彩<sup>[42]</sup>。

4 结束语

注意力机制在深度学习任务中得到了广泛的使用,运用注意力机制进一步提升深度学习的任务效果是常用且有效的一种手段,也成为了当前的研究热点之一<sup>[43]</sup>。本文针对卷积网络中注意力机制的相关研究方法进行了简要总结,并对该技术的实现方法以及其创新之处进行了详细介绍。同时本文对于注意力机制的未来发展方向与可挖掘的创新方式进行了分析<sup>[44]</sup>,指出了需要当前亟待解决的瓶颈与问题,为后续卷积神经网络中注意力机制的进一步研究提供了指导作用<sup>[45]</sup>。

## 参考文献:

- [1] 王文冠,沈建冰,贾云得.视觉注意力检测综述[J].软件学报,2019,30(2):416-439.  
WANG Wenguan, SHEN Jianbing, JIA Yunde. Review of visual attention detection[J]. Acta Sinica Sinica, 2019, 30(2):416-439.
- [2] 任欢,王旭光.注意力机制综述[J].计算机应用,2021,41(21):1-6.  
REN Huan, WANG Xuguang. Review of attention mechanism[J]. Computer Application, 2021, 41(21):1-6.
- [3] 冯兴杰,张志伟,史金钊.基于卷积神经网络和注意力模型的文本情感分析[J].计算机应用研究,2018,35(5):1434-1436.  
FENG Xingjie, ZHANG Zhiwei, SHI Jinchuan. Text sentiment analysis based on convolutional neural network and attention model[J]. Computer Application Research, 2018, 35(5):1434-1436.
- [4] ZHONG Z, LIN Z Q, BIDART R, et al. Squeeze-and-attention networks for semantic segmentation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020:13062-13071.
- [5] ULUTAN O, IFTEKHAR A S M, MANJUNATH B S. VSGNet: spatial attention network for detecting human object interactions using graph convolutions[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020:13614-13623.
- [6] 张卫锋.跨媒体数据语义分析技术研究[D].杭州:杭州电子科技大学,2019.  
ZHANG Weifeng. Research on semantic analysis of cross media data[D]. Hangzhou: Hangzhou University of Electronic Science and Technology, 2019.
- [7] 朱丰,胡晓峰,吴琳,等.从态势认知走向态势智能认知[J].系统仿真学报,2018,30(3):761-771.  
ZHU Feng, HU Xiaofeng, WU Lin, et al. From situation cognition to situation intelligent cognition[J]. Journal of System Simulation, 2018, 30(3):761-771.
- [8] 梁斌,刘全,徐进,等.基于多注意力卷积神经网络的特定目标情感分析[J].计算机研究与发展,2017,54(8):1724-1735.  
LIANG Bin, LIU Quan, XU Jin, et al. Target specific emotion analysis based on multi attention convolutional neural network[J]. Computer Research and Development, 2017, 54(8):1724-1735.
- [9] XU Tingting, ZHENG Guping. Semantic segmentation of street scene based on multi-scale and attention mechanism[J]. World Scientific Research Journal, 2021, 7(5).
- [10] 朱张莉,饶元,吴渊,等.注意力机制在深度学习中的研究进展[J].中文信息学报,2019,33(6):1-11.  
ZHU Zhangli, RAO Yuan, WU Yuan, et al. Research progress of attention mechanism in deep learning[J]. Chinese Journal of Information, 2019, 33(6):1-11.
- [11] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[J]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018:7132-7141.
- [12] 耿磊,邱玲,吴骏,等.结合深度可分离卷积与通道加权的全卷积神经网络视网膜图像血管分割[J].生物医学工程学杂志,2019,36(1):107-115.  
GENG Lei, QIU Ling, WU Jun, et al. Retinal vascular segmentation based on deep separable convolution and channel weighted convolution neural network[J]. Journal of Biomedical Engineering, 2019, 36(1):107-115.
- [13] 盛纾纬.基于弱监督学习的细粒度图像识别技术研究[D].成都:电子科技大学,2019.  
SHENG Shuwei. Research on fine-grained image recognition technology based on weak supervised learning[D]. Chengdu: University of Electronic Science and Technology, 2019.
- [14] WANG Q, WU B, ZHU P, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020:11531-11539.
- [15] LI X, WANG W, HU X, et al. Selective kernel networks[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019:510-519.
- [16] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016:770-778.
- [17] ZHANG H, WU C R, ZHANG Z Y, et al. ResNeSt: split-attention networks[J]. arXiv:2004.08955, 2020.
- [18] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[J]. arXiv:1807.06521v2, 2018.
- [19] WANG Yibin, WANG Haifeng, PENG Zhaohua. Rice diseases detection and classification using attention based neural network and Bayesian optimization[J]. Expert Systems with Applications, 2021, 178.
- [20] YU Y, LIU M, FENG H, et al. Split-attention multiframe alignment network for image restoration[J]. IEEE Access, 2020:39254-39272.
- [21] 李启行,廖薇,孟静雯.基于注意力机制的双通道DAC-RNN文本分类模型[J/OL].计算机工程与应用:1-9(2021-04-20)[2021-05-10].<http://kns.cnki.net/kcms/detail/11.2127.tp.20210420.1354.070.html>.  
LI Qixing, LIAO Wei, MENG Jingwen. Two channel DAC-RNN text classification model based on attention mechanism[J/OL]. Computer Engineering and Application: 1-9 (2021-04-20)[2021-05-10]. <http://kns.cnki.net/kcms/detail/11.2127.tp.20210420.1354.070.html>.
- [22] LI Zhongzhi, QU Na, LI Xiaoxue, et al. Partial discharge detection of insulated conductors based on CNN-LSTM of attention mechanisms[J]. Journal of Power Electronics, 2021.



- [23] LIU Yifan, ZHU Qigang, CAO Feng, et al. High-resolution remote sensing image segmentation framework based on attention mechanism and adaptive weighting[J]. ISPRS International Journal of Geo-Information, 2021, 10(4): 241.
- [24] ZHU Xiaoqing, HAN Junjun. Research on visual question answering based on deep stacked attention network[J]. Journal of Physics: Conference Series, 2021, 1873(1).
- [25] YIN Fulian, LI Sitong, JI Meiqi, et al. Neural TV program recommendation with label and user dual attention[J]. Applied Intelligence, 2021.
- [26] XIE L, HUANG C. A residual network of water scene recognition based on optimized inception module and convolutional block attention module[C]//6th International Conference on Systems and Informatics(ICSAI), Shanghai, China, 2019: 1174-1178.
- [27] FU J. Dual attention network for scene segmentation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Long Beach, CA, USA, 2019: 3141-3149.
- [28] ZHAO T, WU X. Pyramid feature attention network for saliency detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Long Beach, CA, USA, 2019: 3080-3089.
- [29] 石磊, 王毅, 成颖, 等. 自然语言处理中的注意力机制研究综述[J]. 数据分析与知识发现, 2020, 4(5): 1-14.
- SHI Lei, WANG Yi, CHENG Ying, et al. A review of attention mechanism in natural language processing[J]. Data Analysis and Knowledge Discovery, 2020, 4(5): 1-14.
- [30] 乔若羽. 基于注意力机制的神经网络预测模型[D]. 合肥: 中国科学技术大学, 2020.
- QIAO Ruoyu. Neural network prediction model based on attention mechanism[D]. Hefei: University of Science and Technology of China, 2020.
- [31] 陈龙杰, 张钰, 张玉梅, 等. 基于多注意力多尺度特征融合的图像描述生成算法[J]. 计算机应用, 2019, 39(2): 354-359.
- CHEN Longjie, ZHANG Yu, ZHANG Yumei, et al. Image description generation algorithm based on multi attention and multi-scale feature fusion[J]. Computer Applications, 2019, 39(2): 354-359.
- [32] 边继龙, 王厚博, 李金凤. 基于多尺度注意力网络的立体匹配方法[J/OL]. 北京邮电大学学报: 1-7[2021-05-10]. <https://doi.org/10.13190/j.jbupt.2020-204>.
- BIAN Jilong, WANG Houbo, LI Jinfeng. Stereo matching method based on multi-scale attention network[J/OL]. Journal of Beijing University of Posts and Telecommunications: 1-7[2021-05-10]. <https://doi.org/10.13190/j.jbupt.2020-204>.
- [33] CHENG Q, LI H, WU Q, et al. Learn to pay attention via switchable attention for image recognition[C]//IEEE Conference on Multimedia Information Processing and Retrieval(MIPR), 2020: 291-296.
- [34] ZENG Y, GUO X, WANG H, et al. Efficient dual attention module for real-time visual tracking[C]//IEEE Visual Communications and Image Processing(VVIP), 2019: 1-4.
- [35] FANG P, ZHOU J, ROY S K, et al. Attention in attention networks for person retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
- [36] YANG Z, ZHANG T, YANG J. Research on classification algorithms for attention mechanism[C]//19th International Symposium on Distributed Computing and Applications for Business Engineering and Science(DCABES), 2020: 194-197.
- [37] LI C. Attention Unet++: a nested attention-aware U-Net for liver CT image segmentation[C]//IEEE International Conference on Image Processing(ICIP), 2020: 345-349.
- [38] HE A, LI T, LI N, et al. CABNet: category attention block for imbalanced diabetic retinopathy grading[J]. IEEE Transactions on Medical Imaging, 2021: 143-153.
- [39] ZHU M, JIAO L, LIU F, et al. Residual spectral-spatial attention network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021: 449-462.
- [40] XING X, YUAN Y, MENG M Q H. Zoom in lesions for better diagnosis: attention guided deformation network for WCE image classification[J]. IEEE Transactions on Medical Imaging, 2020: 4047-4059.
- [41] PENG B, YU Z, LEI J, et al. Attention-guided fusion network of point cloud and multiple views for 3D shape recognition[C]//IEEE International Conference on Visual Communications and Image Processing(VVIP), 2020: 185-188.
- [42] MEI R, WANG H, MEN A. Attention-enhanced and more balanced R-CNN for object detection[C]//IEEE International Conference on Image Processing(ICIP), 2020: 2136-2140.
- [43] SUN L, SHAO W, ZHANG D, et al. Anatomical attention guided deep networks for ROI segmentation of brain MR images[J]. IEEE Transactions on Medical Imaging, 2020: 2000-2012.
- [44] YI Z, SUN Z, FENG J, et al. 3D residual networks with channel-spatial attention module for action recognition[C]//Chinese Automation Congress(CAC), 2020: 5171-5174.
- [45] GAO Y, GONG H, DING X, et al. Image recognition based on mixed attention mechanism in smart home appliances[C]//IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference(IAEAC), 2021: 1501-1505.