·图形图像处理·　　文章编号：1000-3428(2023)01-0242-08　　文献标志码：A　　中图分类号：TP391.41

基于改进自注意力机制的金字塔场景解析网络

郑秋梅[1]，徐林康[1]，王风华[1]，林　超[2]

（1.中国石油大学（华东）计算机科学与技术学院，山东 青岛 266580；2.中国石油大学（华东）信息化建设处，山东 青岛 266580）

摘　要：金字塔场景解析网络存在图像细节信息随着网络深度加深而丢失的问题，导致小目标与物体边缘语义分割效果不佳、像素类别预测不够准确。提出一种基于改进自注意力机制的金字塔场景解析网络方法，将自注意力机制的通道注意力模块与空间注意力模块分别加入到金字塔场景解析网络的主干网络和加强特征提取网络中，使网络中的两个子网络能够分别从通道和空间两个方面提取图像中更重要的特征细节信息。针对现有的图像降维算法无法更好地提高自注意力机制计算效率的问题，在分析“词汇”顺序对自注意力机制计算结果影响的基础上，利用希尔伯特曲线遍历设计新的图像降维算法，并将该算法加入到空间自注意力模块中，以提高其计算能力。仿真实验结果表明，该方法在PASCAL VOC 2012和息肉分割数据集上的精度均有提高，小目标与物体边缘分割更加精细，其中在VOC 2012训练集中平均交并比与平均像素精度分别达到75.48%、85.07%，较基准算法分别提升了0.68、1.35个百分点。

关键词：语义分割；金字塔场景解析网络；自注意力机制；图像降维；希尔伯特曲线

开放科学（资源服务）标志码（OSID）：

中文引用格式：郑秋梅，徐林康，王风华，等.基于改进自注意力机制的金字塔场景解析网络[J].计算机工程，2023，49(1)：242-249.

英文引用格式：ZHENG Q M，XU L K，WANG F H，et al.Pyramid scene parsing network based on improved self-attention mechanism[J].Computer Engineering，2023，49(1)：242-249.

Pyramid Scene Parsing Network Based on Improved Self-Attention Mechanism

ZHENG Qiumei[1]，XU Linkang[1]，WANG Fenghua[1]，LIN Chao[2]

（1.College of Computer Science and Technology，China University of Petroleum（East China），Qingdao，Shandong 266580，China；
2.Information Construction Department，China University of Petroleum（East China），Qingdao，Shandong 266580，China）

【Abstract】In pyramid Scene Parsing Network（PSPNet），image detail information is lost as the network depth deepens，resulting in poor semantic segmentation of small objects and object edges and inaccurate pixel category prediction. To solve this problem，this paper presents a pyramid scene resolution network method based on an improved self-attention mechanism.The channel and spatial attention modules based on the mechanism are added to the main network of the pyramid scene analysis network and the enhanced feature extraction network，respectively，so that the two sub-networks in the network can extract important feature details from the channel and spatial aspects. Moreover，considering that the current image dimensionality reduction algorithm cannot further improve the calculation effect of the self-attention mechanism，a Hilbert Curve（HC）traversal design is proposed based on analyzing the influence of the order of "words" on the calculation results of the self-attention mechanism.A new image dimensionality reduction algorithm is added to the spatial self-attention module to improve its computing power.The simulation results show that the improved method proposed in this paper has improved accuracy on both PASCAL VOC 2012 and polyp segmentation datasets，and the segmentation of small objects and object edges is more refined. Among them，the average intersection ratio in the VOC 2012 training set reaches 75.48%，which is 0.68 percentage points higher than that of the benchmark algorithm，and the average pixel accuracy reaches 85.07%，which is 1.35 percentage points higher than that of the benchmark algorithm.

【Key words】semantic segmentation；Pyramid Scenarios Parse Network（PSPNet）；self-attention mechanism；image dimensionality reduction；Hilbert Curve（HC）

DOI：10. 19678/j. issn. 1000-3428. 0063652

基金项目：国家自然科学基金“基于超声速膨胀的天然气非均质凝结机理”（52074341）；国家自然科学基金“多相流管道泥砂颗粒冲蚀机制研究”（51874340）；中央高校基本科研业务费专项资金（19CX02030A）。

作者简介：郑秋梅（1964—），女，教授，主研方向为图像处理、目标检测；徐林康，硕士研究生；王风华，讲师、博士；林　超，高级工程师、硕士。

收稿日期：2021-12-29　　**修回日期：**2022-03-06　　**E-mail：**fenghuawang@upc.edu.cn

0　概述

语义分割是对目标图像中的所有像素进行分类，并赋予对应的ID，利用颜色标注表示该像素的语义信息在自动驾驶和医疗图像中都得到了应用。

图像分割早期阶段提取的是图像的浅层特征，研究人员利用图像中的像素值梯度或设置阈值提出了Ostu[1]、FCM[2]、分水岭[3]等算法，但这些算法无法精细分割复杂场景的图像，分割图也不具有语义信息。

2015年，LONG[4]等提出的全卷积网络（Fully Convolutional Network，FCN）通过全卷积化将最后的全连接层删除，满足任意分辨率的输入，并结合跳跃结构保证网络分割的健壮性和准确率。文献[5]提出了金字塔场景解析网络（Pyramid Scenarios Parse Network，PSPNet），通过引入金字塔池化模块，融合多尺度上下文信息，弥补了FCN缺乏对图像全局特征和上下文信息利用的问题。而此类基于特征融合的网络[6-7]在将卷积神经网络（Convolutional Neural Nework，CNN）作为主干网络进行特征提取时，针对特征图中所有的通道和空间都默认具有相同的权重，但并非所有的通道都值得关注，这也导致PSPNet得到的分割图表现出对小目标与物体边缘语义分割效果不佳，因此可在主干网络中添加注意力模块[8-15]来计算不同的权重以提高主干网络提取特征的能力。

2018年，HU[16]等提出SE注意力模块，它从通道维度来计算全局信息并取得特征通道的权重，但忽略了特征图空间位置的权重信息。文献[17]针对城市场景数据集中像素的空间排布规律提出了层次对齐网络，通过类似于全连接的方式取得高度驱动的注意力图，从而实现分割精度的提高，但是这种网络只是针对特定的数据集，并不具有泛化性。文献[18]提出了卷积块注意模块（Convolutional Block Attention Module，CBAM），它对特征图的通道、空间进行注意力的计算，将其添加到基础网络中可使网络更注意目标物体的本身以得到更好的精度。该模块虽然关注了通道和空间的权重信息，但这种采用全连接的方式并未考虑两种维度的内在相关性。对于网络中的特征图，无论空间或是通道信息之间都具有很强的相关性，网络通过计算这种相关性能更好的识别目标并进行分割，2021年，YUAN[19]等提出了对象上下文网络，它是对通道和空间两个维度采用自注意力机制来捕获内在相关性，自适应地结合局部与全局特征。但该网络并未考虑到使用自注意力机制时对特征图进行降维后，像素排列的顺序会对结果产生影响。而当前对二维特征图进行降维的方式会使具有相似语义的相邻像素在降维后导致距离过大，这样并不利于自注意力机制的计算。

GILBERT[20]提出一种空间填充曲线，可用一维曲线去包含整个二维空间，称为希尔伯特曲线（Hilbert Curve，HC）。文献[21]利用HC对二维点云数据进行遍历，这种利用HC曲线进行遍历的方法在计算机多个领域得到了广泛应用，但在图像领域较少应用。

本文提出一种基于希尔伯特曲线遍历算法和自注意力机制的金字塔场景解析网络HA-PSPNet。将通道、空间注意力模块添加自注意力机制并分别放入PSPNet的两个子网络中，加强两个子网络提取特征的能力，同时利用HC遍历算法改变图像降维方式，在不增加网络参数量的情况下提高自注意力计算能力，以使网络分割小目标和物体边缘更加精确。

1　相关理论

1.1　金字塔场景解析网络

针对FCN缺乏对全局信息和上下文信息关注的问题，ZHAO提出了金字塔池化模块[5]。该模块将主干特征提取网络输出的特征图$X\in\mathbb{R}^{C\times H\times W}$进行平均池化，生成分辨率为1×1、2×2、3×3、6×6的特征图。通过卷积再聚合这些具有不同程度上下文信息的特征图，以提高全局信息获取能力，将聚合后的特征图进行上采样，再和特征图$X\in\mathbb{R}^{C\times H\times W}$进行通道维度的拼接，通过像素分类最后输出预测图。金字塔池化模块结构如图1所示。

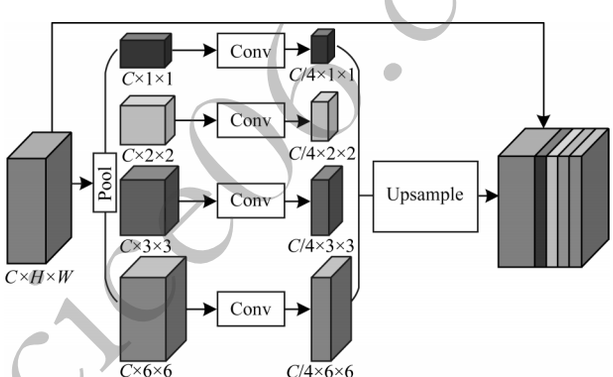


图1　金字塔池化模块结构

Fig.1　Structure of pyramid pooling module

为了使网络具有采集全局上下文信息的能力，该模块把不同区域的上下文信息进行聚合，但PSPNet仍存在对图像中目标的边缘轮廓和一些细小目标进行分割时不够精细的问题，这主要是由于PSPNet的主干网络进行特征提取时，对特征图中所有通道都默认具有相同的权重，然而并不是所有的通道都值得关注，且在加强网络中经过平均池化后的特征图具有丰富的语义信息，只进行普通的卷积并不能完全利用这些深层次信息。

1.2　CBAM与自注意力机制

CBAM主要包括通道注意力模块和空间注意力模块。通道注意力模块的作用是为特征图通道分配不同的权重，给具有重要特征的通道更高的权值。该模块会对输入的特征图进行基于空间的平均与最大池化，再将两种特征图分别进行两次全连接后相加，经过Sigmoid进行激活操作得到通道注意力特征图，其公式如下：

$$\omega=\sigma(F_{\mathrm{C}}(\theta(F_{\mathrm{C}}(\mathrm{AvgPool}(X_{\mathrm{in}}))))+F_{\mathrm{C}}(\theta(F_{\mathrm{C}}(\mathrm{MaxPool}(X_{\mathrm{in}})))))\tag{1}$$

$$X_{\mathrm{out}}=X_{\mathrm{in}}\times\omega\tag{2}$$

其中：$X_{\mathrm{in}}\in\mathbb{R}^{B\times C\times H\times W}$表示在主干网络中输出的特征图；$\omega\in\mathbb{R}^{B\times C\times 1\times 1}$表示各个通道的权重；$F_{\mathrm{C}}$表示全连接操作；$\theta$为全连接后的ReLU操作；$\sigma$是Sigmoid操作。经过Sigmoid操作后得到了每一个特征通道的权重信息，再和X_{in}进行相乘最终得到$X_{\mathrm{out}}\in\mathbb{R}^{B\times C\times H\times W}$。

空间注意力模块是将上面模块输出的特征图作为输入，将该特征图进行基于通道的全局最大与平均池化，得到两个形状为$B\times 1\times H\times W$的特征图，然后对这两个特征图进行拼接，经过卷积降维到1个通道，利用Sigmoid函数生成空间注意力特征图，最后与该模块的输入做乘法，得到最终生成的特征。空间注意力机制公式如下：

$$\omega=\sigma\left(\mathrm{Conv}([\mathrm{AvgPool}(X_{\mathrm{in}});\ \mathrm{MaxPool}(X_{\mathrm{in}})])\right)\tag{3}$$

$$X_{\mathrm{out}}=\omega\times X_{\mathrm{in}}\tag{4}$$

上述两种注意力机制采用类似于全连接的方式，在该方式下计算的权重需要目标图像参与，再对权重值进行计算，忽略了源特征图通道和空间两种维度的内在相关性，而采用自注意力机制可以解决这一问题。它将目标像素与剩余像素进行协方差计算，把像素看成随机变量，参与的目标像素只是所有像素值的加权和，得到的权值信息都与所有像素有关。

2018年，WANG[22]等在CNN中提出用self-attention来实现像素级预测的全局参考。

自注意力机制的协方差公式如下：

$$\mathrm{Cov}(X,Y)=\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})\tag{5}$$

$$\bar{X}=\frac{1}{N}\sum_{i=1}^{N}X_i\tag{6}$$

$$\bar{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i\tag{7}$$

把自注意力机制引入到CBAM模块中，会使改进后的注意力模块更好地关注于两种维度的自相关性，得到特征图更好的权重信息。

但将自注意力机制放入网络[23]时，它对“词”向量的位置很敏感，不同位置向量所代表的含义也是不同的。例如，“我喜欢她”和“她喜欢我”，虽然两句话所用的字都一样，但是因为字的顺序不一样，所表达的意思也不一样。目前，大多数的降维工作都只是将后一行的像素拼接到前一行，使得特征图上下相似语义像素的信息距离太远，不利于自注意力的计算。

1.3 希尔伯特曲线

从数学的角度看，每一阶伪希尔伯特曲线是一个映射H_n，它的起、终点是正方形的左下角和右下角：

$$H_n(0)=(0,0),H_n(1)=(1,0)\tag{8}$$

通过适当调整，使每$\frac{1}{4}$的小区间映射到4个区域内，即把$\left[0,\frac{1}{4}\right]$映射到左下角$\left[0,\frac{1}{2}\right]\times\left[0,\frac{1}{2}\right]$，把$\left[\frac{1}{4},\frac{2}{4}\right]$映射到$\left[0,\frac{1}{2}\right]\times\left[\frac{1}{2},1\right]$，以此类推，且这个性质在迭代中一直保持。而真正的希尔伯特曲线，就是这一系列伪希尔伯特曲线的极限，如式(9)所示：

$$H(x)=\lim_{n\to\infty}H_n(x)\tag{9}$$

通过将希尔伯特曲线在特征图上的遍历来代替传统的降维方式有一种优点，即能够保证二维平面在被降维到一维向量时四周的像素距离并未拉大，使一维向量能够自然过渡。

2 基于PSPNet网络的改进

2.1 改进的整体网络结构HA-PSPNet

本文将改进的通道注意力模块添加到主干网络的ResNet[24]中，通过对每个通道进行权重计算来提高主干网络的特征图提取能力，再把改进的空间注意力模块添加进PSPNet的加强网络中，并将基于希尔伯特曲线设计的遍历算法加入具有丰富语义特征的空间注意力模块中，使得语义信息高度相似的像素排在一起，提高自注意力的计算能力，具体网络结构如图2所示。

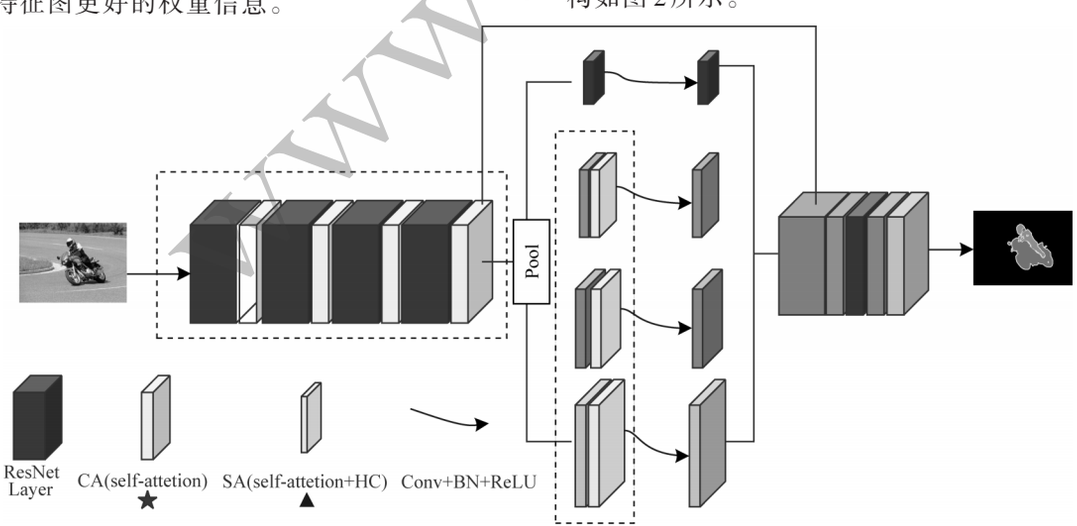


图2 HA-PSPNet网络结构

Fig.2 Structure of HA-PSPNet network

在图2中，标有星号和三角形方块是本文基于自注意力的通道注意力（Channel Attention，CA）和空间注意力（Spatial Attention，SA）模块。其中空间注意力模块中添加了希尔伯特曲线遍历算法，虚线框是本文对于PSPNet改进的位置。

2.2　基于通道注意力机制的改进

由于特征图中的每个通道是某一特征的响应，而这些特征之间本身存在一定的联系，若采用简单的全连接获取通道权值容易忽略这种特征间的依赖关系。同时，CNN在计算全局信息来捕获长范围特征依赖时需要多层网络的累积，导致学习效率低下。自注意力机制的非局部操作可以直接计算两个语义像素之间的关系捕获长范围的依赖关系，且相较于堆叠的卷积网络也更快速。

因此，将通道注意力模块的中间两次全连接变换成自注意力机制，利用自注意力机制从非局部的角度来捕获通道维度上的特征依赖关系，为每个通道分配权重，其模块结构如图3所示。

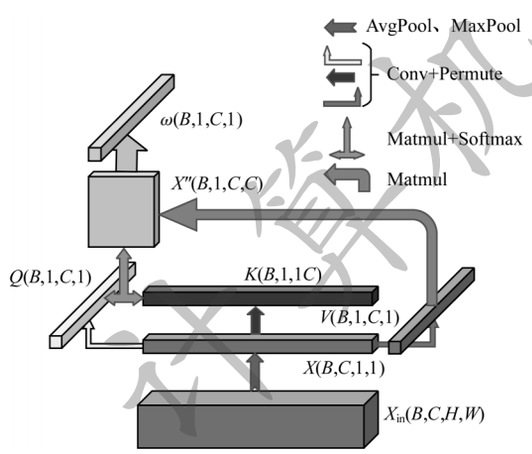


图3　非局部通道注意力模块

Fig.3　Non-local channel attention module

该模块是先将平均池化和最大池化后的特征图进行维度的交换，随后通过三次1×1的卷积变成3个卷积Q（Query）、K（Key）、V（Value），Query和Key分支进行矩阵乘法得到一个形如$B\times1\times C\times C$的关系矩阵，接着对该矩阵使用Softmax函数得到每个通道与其他通道的关系的特征图，最后将Value的分支与特征图相乘。此时得到的结果中每个值都与其他位置的值有关联，代表该通道与其他通道都进行了计算，故有了全局上下文的信息。最后将各个通道的权值图ω与X_{in}相乘，得到所需的特征图X_{out}。自注意力公式如下：

$$Q,K,V=\gamma(\text{Conv}(\text{Pool}(X_{\text{in}})))\tag{10}$$

$$X_{\text{out}}=\sigma(\varepsilon(Q\otimes K)\otimes V)\times X_{\text{in}}\tag{11}$$

其中：γ表示对卷积后的特征图的维度变换；ε表示Softmax函数。

通过在主干网络中添加改进后的通道注意力模块，提高了主干网络特征的提取能力，使得平均交并比和平均像素准确率两个指标都有所提升。

2.3　基于空间注意力机制的改进

为了得到特征图中每个像素的权重值，将原模块进行下采样后再上采样还原至原分辨率。通过全连接的方式利用源特征图与目标图的关系来找到权重，而这忽略了特征图像素内部间的关系。因此，本文先将二维的特征图降维到一维向量，再利用自注意力机制来得到特征图中像素上下文的依赖关系，如图4所示。

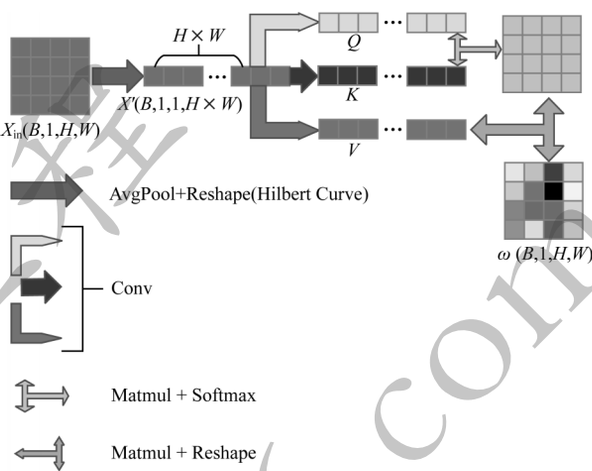


图4　非局部空间注意力模块

Fig.4　Non-local spatial attention modules

对于输入的特征图，本文只采用了平均池化，随后对池化后的特征图进行维度变换，通过三次卷积，分别生成Q、K、V。Query和Key进行矩阵乘法后使用Softmax函数生成注意力图，之后和Value矩阵相乘得到各个像素权重值的向量，最终将向量维度转换成权重图ω，并和输入的特征图X_{in}相乘。

在1×1的特征图中不存在和其他像素的关联，故本文只对2×2、4×4、8×8这3种特征图进行空间注意力的转化。经过空间注意力计算后，特征图有了更好的表达能力，进一步提高了图像的分割效果。

2.4　希尔伯特曲线的遍历

本文采用希尔伯特曲线而不采用最简单的遍历方式的原因在于：经过多次下采样和池化后的特征图中的像素具有丰富的语义信息，原来语义紧密相关的像素在被降维后，因距离过大而使得用自注意力机制计算出的相关性也产生了变化。而在采用希尔伯特曲线进行遍历后，周围像素没有产生很大的距离，这样可以使自注意力机制发挥更好的作用。

本文针对空间注意力中的遍历算法进行替换，利用希尔伯特曲线遍历后的一维向量进行自注意力机制得到权重值向量，最后通过反希尔伯特曲线遍历升维得到权重图。

图5所示分别是一~三阶的希尔伯特曲线在特征图中的遍历，可以看出，在二维的特征图中相邻的像素在变成一维后是接近的，尽可能地保持原空间

中相邻点的相关性，便于自注意力机制的计算，从而得到更加精确的权重信息。

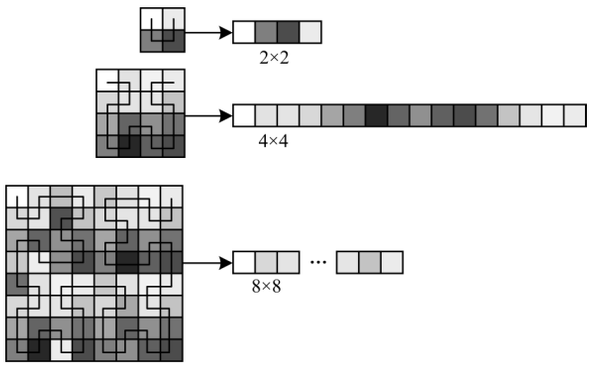


图5 特征图的希尔伯特曲线遍历

Fig.5 Hilbert curve traversal of feature graph

本文通过对通道注意力模块、空间注意力模块的改进，将它们添加进PSPNet网络中不同的位置，使得原网络中的主干网络和加强网络的特征提取能力都有了明显提升，进而提高了整个网络的分割精度。

3 实验与结果分析

3.1 PASCAL VOC 2012数据集

本次实验采用的是PASCAL VOC 2012数据集，该数据集针对语义分割部分有20个分类，外加1个背景类。

在训练网络时，由于训练集每张图像的大小都不一样，因此本次实验将图像和标签的大小设置为473×473。同时，在对图像和标签进行大小调整时，分别对它们采用双三次（BiCubic）插值和最邻近（NEAREST）插值，随机对图像和对应标签进行左翻转。BiCubic函数如下：

$$W_{(x)}=\begin{cases}(a+2)|x|^3-(a+3)|x|^2+1, & |x|\leqslant 1\\ a|x|^3-5a|x|^2+8a|x|-4a, & 1<|x|<2\\ 0, & \text{其他}\end{cases} \tag{12}$$

其中：a取−0.5。

3.2 评价指标

本文采用语义分割的两种指标平均像素准确率（Mean Pixel Accuracy，MPA）、平均交并比（Mean Intersection over Union，MIoU）来评价实验效果。

1）MPA

像素准确率（Pixel Accuracy，PA）表示分类正确的像素数与总像素数的比例。表1所示的是网络对于像素预测值与其真实值相比所包含的4种结果，其中，T_{TP}表示被模型预测为正类的正样本，T_{TN}表示被模型预测为负类的负样本，F_{FP}表示被模型预测为正类的负样本，F_{FN}表示被模型预测为负类的正样本。

$$A_{\text{Accuracy}}=\frac{T_{TP}+T_{TN}}{T_{TP}+T_{TN}+F_{FP}+F_{FN}} \tag{13}$$

$$P_{PA}=\frac{\sum_{i=0}^{20}p_{ii}}{\sum_{i=0}^{20}\sum_{j=0}^{20}p_{ij}} \tag{14}$$

$$M_{MPA}=\frac{1}{21}\sum_{i=0}^{20}\frac{p_{ii}}{\sum_{j=0}^{20}p_{ij}} \tag{15}$$

其中：p_{ii}表示将第i类分割成第i类的像素数量（分类正确的像素数量）；p_{ij}表示将第i类分割成第j类的像素数量（图像中的像素总数）。

表1 像素预测的4种结果

Table 1 Pixels predict four results

真实值	预测值	
	Positive	Negative
Positive	T_{TP}	F_{FN}
Negative	F_{FP}	T_{TN}

2）MIoU

交并比（Intersection over Union，IoU）表示的是对图像中某一种类别预测结果和真实值的交集与并集的比例，公式如下：

$$I_{IoU}=\frac{T_{TP}}{T_{TP}+F_{FP}+F_{FN}} \tag{16}$$

$$M_{MIoU}=\frac{1}{21}\sum_{i=0}^{20}\frac{p_{ii}}{\sum_{j=0}^{20}p_{ij}+\sum_{j=0}^{20}p_{ji}-p_{ij}} \tag{17}$$

其中：M_{MIoU}是对每种类别计算出的交并比求和后的平均值。

3.3 实验参数

模型构建、数据集的验证与测试均在pytorch框架下进行，网络的训练是用单块16 GB的Tesla P100，而网络的评估是在NVIDIA 1650的显卡下进行的。实验的Batch值为12，迭代次数Epoch为60。优化器采用的是Adam优化算法，初始学习率为0.000 1。该算法实现简单且计算高效，对内存的需求较少，适合应用于大规模的数据场景。设置交叉熵损失函数Cross Entropy Loss（CE）与Dice_loss之和为损失函数，其公式如下：

$$C_{\text{entropy}}^{\text{Cross}}=-\sum_{k=1}^{N}(p_k\times\log_a q_k) \tag{18}$$

$$D_{\text{loss}}^{\text{Dice}}=1-2|X\cap Y|/(|X|+|Y|) \tag{19}$$

$$T_{\text{loss}}^{\text{Toal}}=C_{\text{entropy}}^{\text{Cross}}+D_{\text{loss}}^{\text{Dice}} \tag{20}$$

其中：p表示真实值，为one-hot形式；q表示预测值；X表示Ground Truth分割图像；Y表示网络预测出的分割图像；$T_{\text{loss}}^{\text{Toal}}$表示本次实验的损失函数为两种损失函数之和。

3.4 实验评估

3.4.1 自注意力模块性能评估

本节实验将通道注意力模块、空间注意力模块以及改进后的模块依次放入网络中后对网络在验证

集上分割效果进行对比、分析。

实验将通道注意力分别放入以ResNet为主干网络的4个模块后，此时特征图中的底层特征依然保留，利用通道注意力的加权保留特征图中的一些边缘和关键点的信息；将空间注意力放入到加强网络中用来计算特征图中高层特征的内在相关性，对这些高层特征进行加权计算，同时也为了后面在模块中添加遍历算法。如表2所示，当初始的通道注意力模块、空间注意力模块加入到网络后，网络的评价指标都得到提升，说明注意力模块提高了网络的特征提取能力。当实验将自注意力机制分别加入两种模块后，网络分割指标逐步得到提升，尤其是MPA提升了1%以上，说明改进后的两种自注意力模块对于基础网络的分割性能具有促进作用，其中，※表示添加自注意力模块，√表示添加，×表示不添加，粗体表示最优结果。

表2　自注意力模块在验证集上的性能

Table 2　Performance of the self-attention module on validation set

主干网络	通道注意力	空间注意力	HC	MIoU/%	MPA/%
ResNet50	×	×	×	74.80	83.72
ResNet50	√	√	×	75.33	84.33
ResNet50	√	※	×	75.60	84.13
ResNet50	※	※	×	**75.74**	**84.83**

3.4.2　希尔伯特曲线模块性能评估

本节将希尔伯特曲线加入到改进的模块中，并对网络在验证集上的表现进行分析。同时，把网络在测试集的实验结果放入官网的服务器中进行评价，并与经典网络进行对比。为了证明希尔伯特曲线具有提高自注意力机制计算能力且不增加计算量的优点，增加一个息肉数据集来进行验证。在表3中加入HC后，MPA达到了最高值，比基础网络提升了1.35个百分点，而MIoU比基础网络也有0.68个百分点的提升，但还是低于HA*-PSPNet（表示该网络未添加希尔伯特曲线）。从表4可以看出，HA*-PSPNet只在少数几个类别上IoU的提升幅度较大，导致MIoU较高，而HA-PSPNet对于数据集的绝大多数类别都有提升，且1/2类别的IoU达到最高值，因此HA-PSPNet在整体表现上更优，其中，粗体表示最优结果。

表3　希尔伯特曲线在验证集上的性能

Table 3　Performance of the HC on validation set

主干网络	通道注意力	空间注意力	HC	MIoU/%	MPA/%
ResNet50	×	×	×	74.80	83.72
ResNet50	※	※	×	**75.74**	84.83
ResNet50	※	※	√	75.48	**85.07**

表4　不同类别在验证集上的分割精度

Table 4　Segmentation accuracy of different classes on verification set　%

类别	PSPNet	HA*-PSPNet (本文方法)	HA-PSPNet (本文方法)
Aero	85.41	86.68	**88.28**
Bike	**42.04**	41.81	41.64
Bird	83.91	**86.33**	84.76
Boat	**72.85**	72.35	70.14
Bottle	**78.46**	76.86	76.90
Bus	93.87	93.56	**94.03**
Car	86.30	86.74	**87.38**
Cat	90.18	90.77	**91.32**
Chair	41.25	39.71	**42.59**
Cow	84.56	84.44	**85.54**
Table	49.78	**55.04**	51.44
Dog	**84.90**	84.06	83.24
Horse	**81.11**	80.66	80.30
Motorbike	83.34	**83.79**	83.45
Person	82.33	83.09	**83.26**
Plant	57.13	62.55	**63.64**
Sheep	83.71	86.01	**86.88**
Sofa	41.16	**46.19**	42.70
Train	85.31	86.80	**87.33**
Tv	**69.79**	69.20	66.58
MIoU	74.80	**75.74(+0.94)**	75.48
MPA	83.72	84.83	**85.07(+1.35)**

从表5可以看出，本文网络在测试集上对于基础网络依然有着一定的提高，证明了本文网络具有泛化性。

表5　不同算法在测试集上的性能

Table 5　Performance of different algorithms on test set　%

方法	MIoU
FCN[5]	62.20
Ye[15]	71.50
DPN[25]	74.10
PSPNet[5]	74.55
HA-PSPNet(本文方法)	75.2(+0.65)
HA*-PSPNet(本文方法)	75.27(+0.72)

为了体现HA-PSPNet的优越性，本文利用kaggle竞赛中的结肠镜息肉分割和肿瘤表征数据集（BKAI-IGH NeoPolyp）进行再次验证。该数据集图像中所有息肉分为肿瘤性或非肿瘤性类别，外加一个背景类。本文在读取数据集时，只将输入图像的大小设置为512×512像素，未作其他的图像预处理。

从表6可以看出，希尔伯特曲线可以提高自注意力机制的计算能力，且由于希尔伯特曲线只是针对图像降维算法的改变，并未设计网络模型的改动，故从表7可以看出，对于整个网络的参数量并未提升，因此无需更多的计算量。

表6 息肉分割数据集验证结果

Table 6 Verification results of polyp segmentation data set %

方法	MIoU
PSPNet	70.18
HA*-PSPNet（本文方法）	69.54
HA-PSPNet（本文方法）	71.36(+1.18)

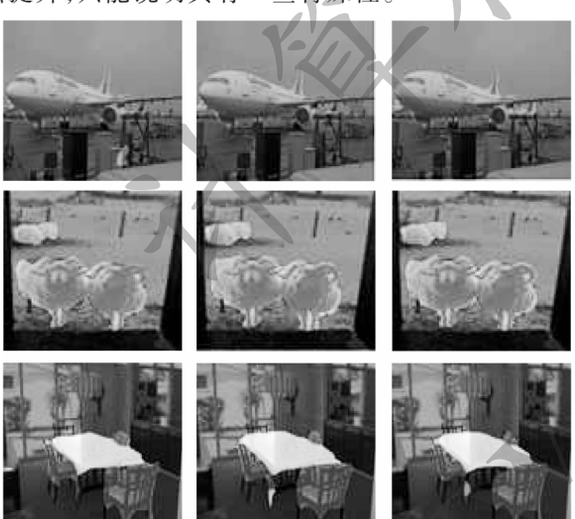
表7 两种方法的参数量

Table 7 Parameters of two methods

方法	参数量/10^6
HA*-PSPNet(本文方法)	86.55
HA-PSPNet(本文方法)	86.55

3.4.3 结果分析

图6所示为三种方法在数据集上的分割效果，其中，HA-PSPNet在一些小目标和物体边缘上处理的更好，比如HA-PSPNet对于靠里面的飞机能够很好地识别出来，对椅子、桌子的腿部和绵羊边缘处分割效果比基础网络更为明显、精细。无论是HA*-PSPNet还是HA-PSPNet，都对原网络有所提升，但从两种数据集的整体表现来看，HA-PSPNet比HA*-PSPNet更好，而HA*-PSPNet在一些类别上分割大幅提升，只能说明具有一些特殊性。



(a)PSPNet (b)HA*-PSPNet（本文方法） (c)HA-PSPNet（本文方法）

图6 三种方法在数据集上的分割效果

Fig.6 The segmentation effect of three methods on data set

从图6可以看出，当两种改进后的自注意力模块分别放入主干网络和加强网络后，使得网络能够识别到图片中的小目标，且对于物体的边缘分割更加精细，同时在不增加网络参数量的情况下通过希尔伯特曲线遍历算法代替传统的遍历算法来提高自注意力模块的计算能力是可行的，表明本文对于PSPNet的改进是有效的。

4 结束语

本文通过对PSPNet网络进行改进，提出一种HA-PSPNet网络。该网络将带有自注意力机制的CA、SA模块放入到PSPNet网络中，增强了网络的特征提取能力，使得整个网络对细节的处理更加精细，并利用希尔伯特曲线对空间注意力模块的像素进行遍历，在没有增加计算量的前提下，提高了自注意力模块的计算能力。实验结果表明，HA-PSPNet网络可明显提高像素精度。但本文只对希尔伯特曲线遍历算法进行初步研究，目前只运用在前三阶的曲线遍历，后续将其运用在更高阶的特征图上，从而能够得到更广泛的应用。

参考文献

[1] 范朝冬，张英杰，欧阳红林，等. 基于改进斜分Otsu法的回转窑火焰图像分割[J]. 自动化学报，2014，40(11)：2480-2489.
FAN C D，ZHANG Y J，OUYANG H L，et al. Improved Otsu method based on histogram oblique segmentation for segmentation of rotary kiln flame image [J]. Acta Automatica Sinica，2014，40(11)：2480-2489.（in Chinese）

[2] BEZDEK J C，EHRLICH R，FULL W. FCM：the fuzzy c-means clustering algorithm[J]. Computers & Geosciences，1984，10(2/3)：191-203.

[3] 丛培盛，孙建忠. 分水岭算法分割显微图像中重叠细胞[J]. 中国图象图形学报，2006，11(12)：1781-1783，1890.
CONG P S，SUN J Z. Application of watershed algorithm for segmenting overlapping cells in microscopic image[J]. Journal of Image and Graphics，2006，11(12)：1781-1783，1890.（in Chinese）

[4] SHELHAMER E，LONG J，DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2017，39(4)：640-651.

[5] ZHAO H S，SHI J P，QI X J，et al. Pyramid scene parsing network [C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C.，USA：IEEE Press，2017：6230-6239.

[6] ZHOU Z W，SIDDIQUEE M M R，TAJBAKHSH N，et al. UNet：redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging，2020，39(6)：1856-1867.

[7] CHENG H K，CHUNG J，TAI Y W，et al. CascadePSP：toward class-agnostic and very high-resolution segmentation via global and local refinement [C]// Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.，USA：IEEE Press，2020：8887-8896.

[8] CHEN W L，ZHU X G，SUN R Q，et al. Tensor low-rank reconstruction for semantic segmentation [EB/OL]. [2021-11-10]. https://arxiv preprint arxiv：2008. 00490.

[9] WANG Q L，WU B G，ZHU P F，et al. ECA-net：efficient channel attention for deep convolutional neural networks[C]// Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C.，USA：IEEE Press，2020：11531-11539.

[10] HUANG Y, JIA W, HE X, et al. CAA: channelized axial attention for semantic segmentation[EB/OL]. [2021-11-10]. https://arxiv preprint arxiv:2101. 07434.

[11] HUANG Z L, WANG X G, WEI Y C, et al. CCNet: criss-cross attention for semantic segmentation[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 603-612.

[12] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 3141-3149.

[13] LIN H Z, CHENG X, WU X Y, et al. CAT: cross attention in vision transformer[EB/OL]. [2021-11-10]. https://arxiv preprint arxiv:2106. 05786.

[14] VASWANI A. Attention is all you need[EB/OL]. [2021-11-10]. https://arxiv preprint arxiv:1706. 03762.

[15] 叶剑锋，徐轲，熊峻峰，等. 基于注意力机制和辅助任务的语义分割算法[J]. 计算机工程，2021，47(9)：203-209，216.
YE J F, XU K, XIONG J F, et al. Semantic segmentation algorithm based on attention mechanism and auxiliary task[J]. Computer Engineering, 2021, 47(9): 203-209, 216. (in Chinese)

[16] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[17] CHOI S, KIM J T, CHOO J. Cars can't fly up in the sky: improving urban-scene segmentation via height-driven attention networks[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 9370-9380.

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

[19] YUAN Y H, HUANG L, GUO J Y, et al. OCNet: object context for semantic segmentation[J]. International Journal of Computer Vision, 2021, 129(8): 2375-2398.

[20] GILBERT W J. A cube-filling Hilbert curve[J]. The Mathematical Intelligencer, 1984, 6(3): 78-79.

[21] SU T Y, WANG W, LÜ Z H, et al. Rapid Delaunay triangulation for randomly distributed point cloud data using adaptive Hilbert curve[J]. Computers & Graphics, 2016, 54: 65-74.

[22] WANG X L, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 7794-7803.

[23] DOSOVITSKIY A. An image is worth 16×16 words: transformers for image recognition at scale [C]// Proceedings of IEEE ICLR'21. Washington D. C., USA: IEEE Press, 2021: 325-337.

[24] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.

[25] LIU Z W, LI X X, LUO P, et al. Semantic image segmentation via deep parsing network[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 1377-1385.

编辑 索书志