Journal of Computer Applications
计算机应用
ISSN 1001-9081
CODEN JYIIDU
http://www.joca

doi:10.11772/j.issn.1001-9081. 2022121887

改进 Deeplab v3 plus 网络的图书书脊分割算法

姬晓飞*，张可心，唐李荣

(沈阳航空航天大学 自动化学院，沈阳 110136)

(*通信作者电子邮箱 jixiaofei7804@126.com)

摘 要: 图书的定位是实现图书馆智能化发展的重要技术之一，而精确的书脊分割算法成为实现该目标的一大难题。针对这一难题，本文提出了改进 Deeplab v3 plus 网络的图书书脊分割算法，旨在解决图书密集排列、图书存在倾斜角度、书脊纹理极相似等情况下的图书分割难点。首先，为了提取图书图像更密集的金字塔特征，将原始 Deeplab v3 plus 网络中的 ASPP 框架用多空洞率、多尺度的 DenseASPP 框架进行替换；其次，针对原始 Deeplab v3 plus 网络对长宽比大的目标对象分割边界不敏感的问题，在 DenseASPP 框架内的支路上加入条形池化(SP)增强书脊的长条形特征。最后，参考 Vision Transformer 模型中的多头自注意机制提出了一种全局信息增强的自注意机制，增强网络获取长距离特征的能力。将提出的算法在开源数据库上进行对比测试，实验表明，与原始 Deeplab v3 plus 网络分割算法相比，所提算法在近竖直书脊数据库上的平均交并比(MIoU)提升了 1.2 个百分点；在倾斜书脊数据库上的平均交并比提升了 4 个百分点，达到了 93.3%。进一步证实了该网络实现了一定倾斜角度的、密集的、长宽比大的书脊目标的精确分割。

关键词: 书脊分割；智能图书馆；Deeplab v3 plus 网络；DenseASPP；自注意机制

中图分类号:TP391.1　　**文献标志码: A**

Book spine segmentation algorithm by improved Deeplab v3 plus network

JI Xiaofei*, ZHANG Kexin, TANG Lirong

(College of Automation, Shenyang Aerospace University, Shenyang Liaoning 110136, China)

Abstract: The positioning of books is one of the critical technologies to realize the intelligent development of libraries, and the accurate spine segmentation algorithm has become a major challenge to achieve this goal. To address this problem, an improved book spine segmentation algorithm based on Deeplab v3 plus network was proposed, aiming to solve the difficulties in book segmentation caused by dense arrangement, skewed angles, and extremely similar spine textures. Firstly, to extract more dense pyramid features from book images, the Atrous Spatial Pyramid Pooling (ASPP) framework in the original Deeplab v3 plus network was replaced with the multi-dilation rate and multi-scale Dense Atrous Spatial Pyramid Pooling (DenseASPP) framework. Secondly, to solve the problem of insensitivity of the original Deeplab v3 plus network to the boundary segmentation of objects with large aspect ratios, strip pooling (SP) was added to the branch in the DenseASPP framework to enhance the long strip features of book spines. Finally, a global information enhancement self-attention mechanism was proposed based on the multi-head self-attention mechanism in the Vision Transformer (ViT) model to enhance the network's ability to obtain long-distance features. The proposed algorithm is tested and compared on an open-source database, and the experiments shows that compared with the original Deeplab v3 plus network segmentation algorithm, the proposed algorithm improves the Mean Intersection over Union (MIoU) by 1.2 percentage points in the nearly vertical spine database and by 4 percentage points in the skewed spine database, achieving 93.3%. This further confirms that the network achieves accurate segmentation of book spine targets with a certain skew angle, dense arrangement, and large aspect ratio.

Keywords: book spine segmentation; intelligent library; Deeplab v3 plus network; DenseASPP (Dense Atrous Spatial Pyramid Pooling); self-attention mechanism

0　引言

随着信息化社会的发展，图书馆作为知识和信息的聚集地受到了信息化浪潮带来的冲击。读者数量与馆藏量的增加

收稿日期: 2022-12-22;　　**修回日期**:2023-03-21;　　**录用日期**: 2023-03-22。

基金项目:辽宁省教育厅重点公关项目（LJKZZ20220033）。

作者简介: 姬晓飞(1978—)，女，辽宁鞍山人，副教授，博士，主要研究方向：视频分析与处理、模式识别；张可心(1996—)，女，辽宁锦州人，硕士研究生，主要研究方向：图像处理、视频分析与处理；唐李荣(2000—)，男，四川南充人，硕士研究生，主要研究方向：图像处理、视频分析与处理。

使得图书馆传统查找图书的方式已经不能满足读者高效获取图书的需求。相比之下，基于图像处理的图书自动定位方法已经成为研究热点。对于在架图书来说，只有书脊部分可以被观察到，而每本图书书脊的分割是对图书实现精确定位的前提。本文致力于解决在架图书书脊图像的分割问题。在此项工作中主要的挑战在于：1)图书摆放数量较多，属于对密集对象的分割。2)由于书籍的薄厚不一致导致书脊具有差别较大的长宽比。3)相同系列书籍的排放，在纹理上具有极高的重复或者相似性，难以区分边界。4)拍摄角度或者图书的倾斜摆放致使图像中的书籍呈现不同的倾斜角度。

基于传统图像处理的方法主要是依靠人工提取特征送入分类器来实现，如颜色、纹理、尺度不变特征变换等特征与支持向量机(Support Vector Machine，SVM)的配合使用。对于密集排列图书的分割，最大困难在于边缘部分的分割。文献[1-2]是直接通过霍夫直线检测或者 LSD(Line Segment Detection)线段检测对书脊两侧直线进行提取。崔晨等[3]提出了一种基于文本检测的书脊区域粗选方法，利用相似字符提取候选书脊图像的方向梯度直方图特征输入到 SVM 中进行判断。Nevetha 等[4]则提出一种带有若干启发式规则的线段检测器来获取书脊边缘。这些传统方法受限于手工提取特征的单一性，容易受到密集排列书脊高纹理区域的相似性和边界模糊性的影响，产生错误的分割线，鲁棒性很差。

近年来，卷积神经网络(Convolutional Neural Networks, CNN)已经帮助计算机视觉系统在包括图像分类[5]、目标检测[6]和语义分割[7]在内的广泛应用中取得了更好的表现。分割的准确性由局部特征(颜色和强度)和全局特征(纹理和背景)决定。在不同的 CNN 变体中，被广泛认可的对称编码器-解码器体系结构命名法 U-Net[8]显示了突出的细分潜力。它主要由一系列连续的卷积层和下采样层组成，通过收缩路径捕获上下文语义信息，然后在解码器中，使用来自编码器的横向连接，对粗粒度深特征和细粒度浅特征映射进行上采样，来生成精确的分割映射。遵循这一技术路线，为了进一步提高分割性能，随后出现了很多 U-Net 的变体，如 U-Net++[9]和 Res-UNet[10]。这种体系结构的一个重要缺点是感受野大小存在限制，这使得深度模型无法捕获足够的上下文信息，导致在边界等复杂区域分割失败。为了缓解这个问题 Chen 等[11]提出了 Deeplab 网络，引入了一种使用上采样滤波器的新型卷积操作，即膨胀卷积，来扩大滤波器的视野，以吸收更大的上下文，而不增加计算量。其次，该网络为了能够捕捉更精细的细节，采用条件随机场来细化分割结果。在此基础上，为了提取目标的多尺度特征，研究者又提出了 Deeplabv2[12]，该网络使用空洞金字塔池化(Atrous Spatial Pyramid Pooling, ASPP)模块来实现对多尺度对象的分割。ASPP 模块通过探测具有不同采样率的多个膨胀卷积的特征映射来获取多尺度的信息表示。随后，DeepLabv3[13]设计了一个带有膨胀卷积的编码器-解码器架构，以获得更清晰的对象边界，利用深度可分离卷积来提高计算效率。最终，Chen 等[14]提出了 Deeplab v3 plus 网络模型，通过添加一个简单而有效的解码器模块来扩展 Deeplabv3，以提高分割性能。Deeplab 系列网络经过一系列优化，成为目前语义分割领域的主流网络之一。此类网络模型经过不断的改进，得到了令人满意的分割效果，但其不足仍然突出，由于局部性和权值共享的归纳偏差[15]，它们不可避免地在学习远程依赖性和空间相关性方面存在约束，导致复杂结构的次优分割。

新颖的架构 ViT (Vision Transformer)[16]由于其优雅的设计和注意机制的存在，在计算机视觉时代引发了讨论。与 CNN 不同的是，ViT 网络具备了学习长距离特征和全局信息的能力，这也使得 ViT 网络在图像分割任务上表现突出。尽管 ViT 是捕捉全局上下文信息和长距离信息的一个很好的设计选择，但在捕捉低级像素信息方面很弱，这对于精确的分割任务无疑是最大的难点。因此，为了避免其高内存需求，Swin-transformer[17] 提出了一种具有非重叠窗口的局部计算的分层 ViT。面对高效的 CNN 和强大的 ViT 之间的困境，这两个领域之间的交叉出现了，例如 Trans-UNet[18]、TransDeepLab[19]等，此类方法使用 Transformer 来重构一个经典的 CNN 网络，但这无疑会增加模型的复杂性。有研究[20]证明，ViT 网络的优越性表现一部分原因在于多头自注意(Multi-Headed Self-Attention，MHSA)机制的引入，而 MHSA 能够对输入的特征进行全局建模。

综合考虑 CNN 和 ViT 的网络优势，本文提出了改进 Deeplab v3 plus 网络的图书书脊分割算法，此模型兼具了 CNN 出色的低级像素处理能力又结合了 ViT 中对全局信息建模的能力，在书籍分割中表现出了优异的效果。其主要贡献在于：1)针对分布密集的目标使用 DenseASPP(Dense Atrous Spatial Pyramid Pooling)结构取代 ASPP 网络，在密集目标分割任务上有更好的效果。2)引入了条纹池化(Strip Pooling，SP)模块，保留书脊的长条形特征。3)参考 ViT 中的 MHSA 结构搭建自注意机制，并将其应用到 CNN 网络中，使得该网络可以增强特征的上下文信息。

1 网络结构

本文提出的网络模型如图 1 所示，该网络遵循 Deeplab v3 plus 的原始框架，骨干网络选用 MobileNetV2 网络。首先，将图 1 中左上方的书籍图像输入到 MobileNetV2 网络中进行特征提取，对 MobileNetV2 网络的后三层的特征图进行上采样融合，将融合结果作为浅层特征。其次，将 MobileNetV2 网络的最后一层输出送入 DenseASPP 模块。编码阶段，本文利用 DenseASPP 模块取代了 ASPP 模块，来产生更大的接受域，生成更密集的图像特征。考虑到书脊长宽比较大的情况，在 DenseASPP 模块中引入条纹池化模块来保留长条形的图像特征。最后，DenseASPP 模块产生的特征经过 1×1 卷积操作实现通道压缩，送入自注意模块得到深层特征。译码阶段，对浅层特征层利用 1×1 卷积调整通道数，送入自注意模块，与上面深层特征进行拼接，随后进行两次卷积和一次上采样操作后得到最终的预测结果，如图 1 中右下方图像所示。

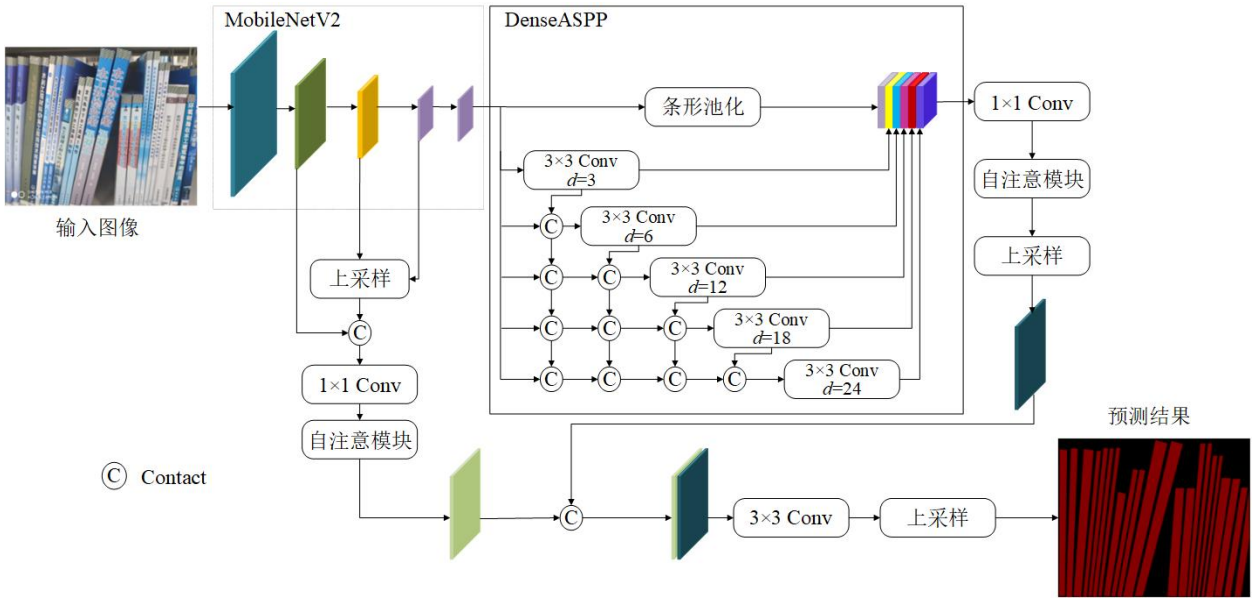


图 1 提出的网络模型

Fig.1 Proposed network model

2 网络细节介绍

2.1 DenseASPP 模块

针对书籍图像这种密集型分割任务，引入了 DenseASPP 模块以生成更密集的特征。其结构如图 1 所示，空洞卷积层以级联方式组织，膨胀率小的层在下部，膨胀率大的层在上部，每一层的膨胀率逐层增加。将每一层的输出与输入的特征图和较低层的所有输出连接起来，并将这些连接起来的特征图送入下一层。DenseASPP 的最终输出是由多空洞率、多尺度的卷积生成的特征图。通过一系列的空洞卷积，较后面层次的神经元获得越来越大的感受野，而不会出现 ASPP[13] 的核退化问题。与原始的 ASPP 相比，DenseASPP 将所有空洞卷积层堆叠在一起，并用紧密的连接将它们连接起来。这种变化主要带来两个好处：

1)更密集的特征金字塔

密集抽样规模：DenseASPP 是一个有效的架构，可以对不同规模的输入进行采样。其使用紧密的连接来实现不同膨胀率的不同层次的集成。

密集的像素采样：与 ASPP 相比，DenseASPP 在特征金字塔的计算中涉及到更多的像素。ASPP 采用 4 个膨胀率分别为 6、12、18、24 的卷积层构成特征金字塔。与相同接收域的传统卷积层相比，大扩张率的卷积层的像素采样率非常稀疏。在 DenseASPP 中，膨胀率 d 逐层增加，因此，上层的卷积可以利用下层的特征，使像素采样更加密集。

2)更大的接受域

DenseASPP 带来的另一个好处是更大的接受域。膨胀卷积在传统的 ASPP 中是并行工作的，而四个分支在前馈过程中是不共享任何信息的。相反，DenseASPP 中的空洞卷积层通过跳层连接来共享信息。小膨胀率和大膨胀率的层之间是相互依赖的，其中前馈过程不仅会构成一个更密集的特征金字塔，而且会产生一个更大的过滤器感知更大的上下文。

2.2 条纹池化

在 DenseASPP 模块中引入条纹池化模块，如图 2 所示，其核心思想在于，在空间维度上应用了一个长条状的池化卷积核，从而增强捕获长距离信息的能力，旨在保留书脊的长条型特征。其水平、竖直方向的池化计算公式分别为：

$$y_j^w = \frac{1}{H}\sum_{0\le i\le H} x_{i,j} \tag{1}$$

$$y_j^h = \frac{1}{W}\sum_{0\le j\le W} x_{i,j} \tag{2}$$

其中，H,W 分别为特征图的高和宽，$x_{i,j}$ 表示特征图第 i 行 j 列的像素值。

结合图 2，利用式(1)、(2)对输入张量中的某一像素所在行和列的局部特征值进行平均条纹池化得到图示框内两部分，对其分别做一维卷积操作，将得到的结果进行上采样至输入张量大小，然后进行特征融合，经过卷积、sigmoid 环节后与输入张量按像素相乘得到输出张量。在上述过程中，实现了输出张量中的每个位置均与输入张量中的位置建立关系。输出张量中以红框为界的正方形连接到与其具有相同水平或垂直坐标的所有位置，进而实现了长条信息的保留。

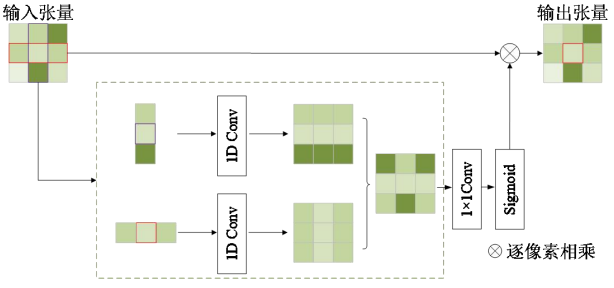


图 2 条形池化过程

Fig.2 strip pooling process

2.3 自注意模块

为了使 CNN 网络更好的提取全局信息，参考 MHSA，搭建了自注意结构，如图 3 所示。将 Deeplab v3 plus 生成的深层、浅层特征分别送入自注意模块，以深层特征为例。特征图 x 的尺寸为 $[c,w,h]$，经过 1×1 Conv 环节调整通道数后，将 $\boldsymbol{k,v,q}$ 与模块进行展平降维操作，对 $\boldsymbol{q}$ 转置后与 $\boldsymbol{k}$ 相乘，得到大小为 $[w*h,w*h]$ 特征图，经过 softmax 模块后与 v 相乘后得到特征大小为 $[c/n,w*h]$ 经过 1×1 Conv 后与 $\boldsymbol{x}$ 相加得到最终特征图。该网络的输入是特征，旨在对特征进行全局建模。

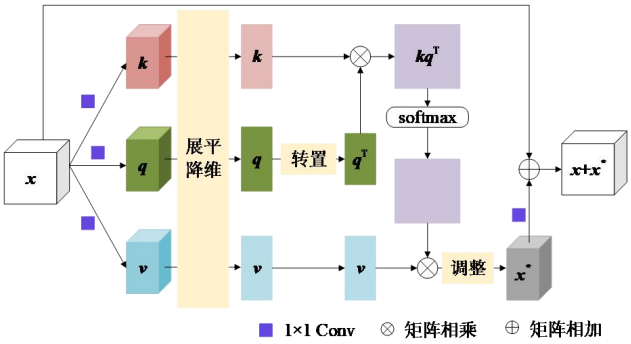


图 3 自注意机制

Fig.3 Self-attention mechanism

3 实验及结果分析

3.1 数据集

本文测试的书脊数据库均来自于文献[21]提供的开源数据库。一共 661 张书籍图像，包含了不同拍摄的角度、不同的书籍倾斜角度等。为了方便进行后续算法的测试，本文将 661 张书籍图像数据按照倾斜与否分为两类，其中倾斜图像共有 283 张，近竖直角度(倾斜角度小于 5°)为 378 张。各按照 1:3 的比例取出测试集与训练集。

3.2 实验设计

在实现细节上，该网络是基于 PyTorch 库实现的，并在单个 Nvidia RTX 3060 GPU 上进行训练，处理器为 12th Gen Intel(R) Core(TM) i5-12400F，批次大小为 4(资源限制)，初始学习率为 0.05，使用随机梯度下降法(Stochastic Gradient Descent，SGD)作为优化方法。采用 Dice 损失和交叉熵损失作为目标函数，采用 L2 范数进行模型正则化。使用旋转和翻转技术作为数据增强方法，使得训练集多样化。分割模型训练分为两个部分，首先不考虑正负样本的平衡关系进行全网络训练，训练的损失如图 4 所示，当训练到损失值基本不能下降后，即 1800 次左右。将正负样本损失比重设置为 1:8，启用 focal loss 继续训练。本文采用平均交并比(Mean Intersection over Union，MIoU)指标对分割效果在测试集上进行评价。

$$V=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{3}$$

其中，$k+1$ 为类数，包含一个背景，p_{ii} 为真正例(预测标签与真实标签相同，均为书脊区域)，p_{ij} 为假负例(预测结果为非书脊区域，真实标签为书脊区域)，p_{ji} 为假正例(预测结果为书脊区域，真实标签为非书脊区域)，i 表示真实标签，j 表示预测标签。V 表示预测区域与手工标记区域的平均交并比。

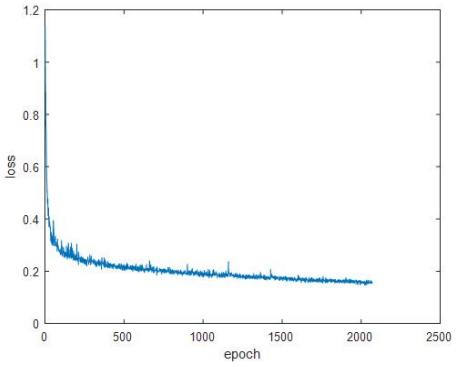


图 4 损失-训练次数关系图

Fig.4 Loss-Epoch Graph

3.3 评估结果

3.3.1 改进算法的有效性验证

为了验证改进算法的有效性，对以上改进操作逐一进行实验测试。实验基于 Deeplab v3 plus 的原始网络展开，骨架网络选用 MobileNetV2，学习率为 0.0025，使用相同的线性衰减率，训练次数为 3000，且不启用 focal loss 训练，对全测试集(包含近竖直测试集与倾斜数据集)进行统计。

1)DenseASPP 有效验证

Deeplab v3 plus 网络框架中分别使用 DenseASPP 模块与 ASPP 模块得到分割结果分别为 91.2%，89.3%。使用 DenseASPP 模块替换 ASPP 模块后，该网络分割的准确度提高了 0.9%，证明了其优势。

为了减轻模型的复杂度，本文选用大小为 3 的卷积核且不同膨胀率构成空洞卷积层，不同层之间进行级联，考虑到 DenseASPP 模块的网络层数对分割效果的影响，开展以下实验，实验结果见表 1。

表 1 DenseASPP 模块的网络层数对分割效果的影响

Tab.1 Influence of the number of network layers of DenseASPP module on the accuracy

实验	网络层数	MIoU/%
实验一	3	79.4
实验二	4	88.5
实验三	5	91.2
实验四	6	89.9

从表 1 的结果来看，当网络层数较低或者较高时，对准确率存在一定的影响，网络层数较低时，细节信息较少，特征不明显，因此准确率不高；但当网络层数较高时，会出现过拟合的现象导致准确率降低。

2)自注意模块验证

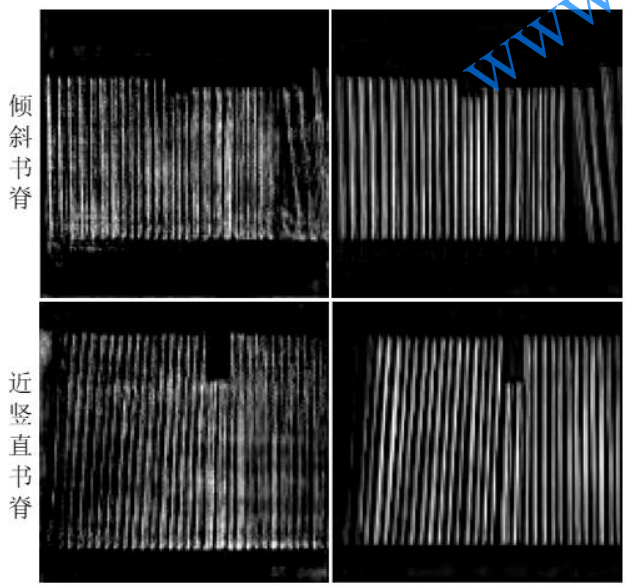
实验分别在 Xception 和 MobileNetV2 两种骨架网络上进行测试，保留 Deeplab v3 plus 网络原始框架(DenseASPP 模块代替 ASPP 模块)，只增加自注意模块，其结果如表 2 所示。

表 2 引入自注意模块前后对比实验结果

Tab.2 Comparison of experimental results before and after the introduction of self-attention module

网络骨架	自注意	MIoU/%
Xception	是	92.7
	否	92.2
MobileNetV2	是	93.8
	否	93.1

图 5 分别展示了经过 MobileNetV2 骨架特征提取后，自注意力模块使用前后，对书脊上下文特征的影响。相比较图 5(a)，5(b)得到的书脊特征更为清晰。最终，依据表 2 和图 5(a)(b)的结果，不论采用哪种的特征提取网络骨架，在引入自注意模块后，准确率均有所上升，证明了自注意机制可以关联全局信息，在分割任务中发挥重要的作用。



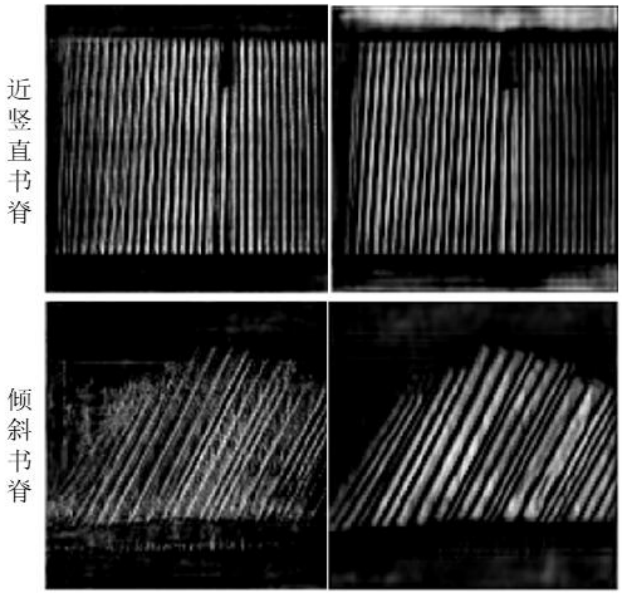
(a)加入自注意力模块前　(b)加入自注意力模块后

图 5 特征可视化对比

Fig.5 Feature visualization comparison

3)条纹池化模块有效验证

利用Deeplab v3 plus网络原始框架(DenseASPP模块代替ASPP模块)，比较有无条纹池化模块在书脊分割上的差异，以证明条纹池化模块的应用价值。引入条纹池化模块前后，深层特征和浅层特征融合得到的特征可视化结果如图6所示。



(a)加入条纹池化模块前　(b)加入条纹池化模块后

图 6 特征可视化对比

Fig.6 Feature visualization comparison

从结果来看，相比较图6(a)，图6(b)在加入条纹池化模块后，使书脊的长条特征得到了增强，但与此同时受环境的影响，如书架横栏等也会被条纹池化进行特征增强，因此可能会出现一些无关特征。在总体框架中可利用自注意模块对无关特征进行抑制，这也证明了自注意模块的重要性。

3.3.2 不同算法对比结果

在进行不同网络分割算法的比较时，将书脊库划分为近竖直书脊数据与倾斜书脊数据，其中训练集采用倾斜、近竖直混合数据进行训练。为了考察书脊的倾斜给各类算法带来的影响，分别在近竖直与倾斜两测试库上给出对比结果。不同网络分割算法的对比测试结果见表3。表3中*代表相应文献提供开源代码和默认参数在本文数据集上进行重新训练得到的测试结果。

1)近竖直书脊测试结果

表3上半部分显示了提出算法与其他经典分割算法在近竖直书脊数据集中的测试结果。从结果来看，本文算法在近竖直方向上的书脊分割任务有较好的表现。其中Mask R-CNN(Mask Region-based Convolutional Neural Networks)网络使用了区域生成网络(Region Proposal Network，RPN)，该网络只能生成规模、尺寸不同的矩形框，但由于书籍的密集性导致此类方法的分割效果不佳。Deeplab v3 plus网络并未对单个目标设计全卷积特征提取网络，这使得该算法在对长宽比例差异大的对象进行检测和分割时，效果较差，这个缺点在目标密集分布的情况下显得格外明显。本文经过对Deeplab v3 plus网络的改进，虽然在一定程度上增加了模型的复杂度，但同时大大地增强了分割算法对书脊特征的表征能力，其在竖直数据集上的测试结果也证明了该算法对于书脊分割的优势。

2)倾斜书脊测试结果

表3下半部分显示了提出算法与其他经典分割算法在倾斜书脊数据集中的测试结果。由此可以看出，Mask R-CNN在倾斜书脊方面的应用效果比较差。而文献[22]采用Mask R-CNN与旋转特征提取方法(Rotation Feature Extraction，RFE)结合的算法，使用旋转区域生成网络(Rotation Region Proposal Network，RRPN)来替换RPN网络，除了大小、比例外，引入一个角度参数对Mask R-CNN网络进行优化。该方法可有效地避免RPN网络带来的角度适应性问题，取得了优于本文提出算法的检测精度，但其大大地增加了学习参数的数量，提高了模型的复杂度，在近竖直的书脊库中表现效果较差。

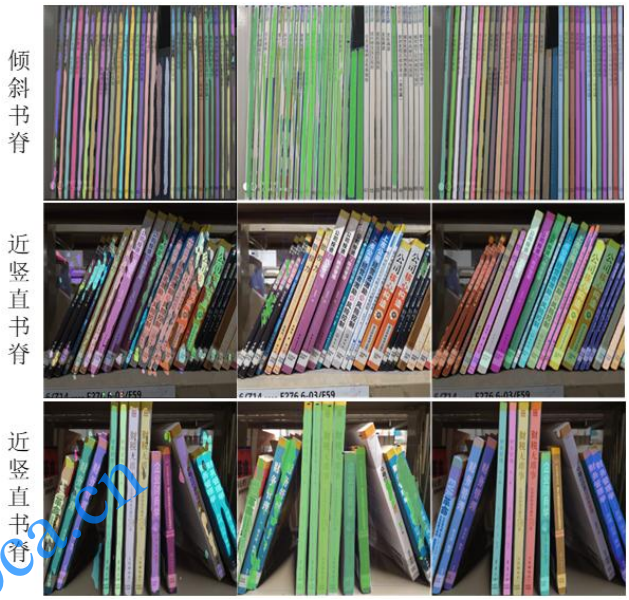
表 3 不同网络分割算法在开源数据库的测试结果

Tab.3 Test results of different network segmentation algorithms in open source databases

数据库	算法	批次	骨架网络	MIoU/%
近竖直书脊数据库	Mask R-CNN分割算法*	2	Resnet50	87.5
	文献[22]分割算法*	2	Resnet50	85.3
	Deeplab v3 plus分割算法*	4	MobileNet V2	92.3
	本文分割算法	4	MobileNet V2	94.1
倾斜书脊数据库	Mask R-CNN分割算法*	2	Resnet50	81.3
	文献[22] 分割算法*	2	Resnet50	93.5
	Deeplab v3 plus分割算法*	4	MobileNet V2	89.2
	本文分割算法	4	MobileNet V2	93.3

综上所述，结合表3的测试结果表明：本文所提出的分割算法在书脊分割上有较好的表现。相比Deeplab v3 plus网络，在相同的特征提取网络和相同训练次数下，本文算法的分割效果更好。在相同操作系统下，相比Mask R-CNN系列，训练了更少的参数，但其性能可以大幅提高。在相同数据集下，文献[21]测试了不同分割算法下的分割效果。其中FCN(Fully Convolutional Networks)模型结构包括FCN32s、FCN16s等结构，32s即从32倍下采样的特征图恢复至输入大小，16s则是从16倍下采样恢复至输入大小。理论上，该数字越小，网络使用的反卷积层进行上采样的操作越多，其对应模型结构更加复杂，理论上分割的效果更为精细。具体的测试结果为：FCN16s、FCN32s、SegNet 、U-Net、Deeplab v3的分割效果(采用MIoU指标)分别为0.8160、0.8193、0.8660、0.8750、0.9186。其中Deeplab v3表现效果最佳，进一步证明了其他分割算法对长条形特征目标的适用性较差，突出了Deeplab系列网络的优越性。

图7揭示了不同算法的分割效果。Deeplab v3 plus网络的分割效果如图7(a)所示，其在密集的目标中表现效果较好，但存在边界分割不清的问题。如图7(b)所示，Mask R-CNN在近似直立的目标上表现一般，且遭遇倾斜目标时容易被相邻目标干扰，甚至出现大量漏检现象。本文算法分割效果如图7(c)所示，该分割算法对密集、具有一定倾斜的目标分割效果较为稳定，尤其对于相邻目标的掩膜预测有更高的隔离性，不会出现其他算法中相邻目标相互影响的情况，有效地提高了分割的正确率。



(a) Deeplab v3 plus (b) Mask R-CNN (c)本文算法

图 7 分割效果

Fig.7 Split effect

4 结语

本文提出了一种基于Deeplab v3 plus的改进网络用于分割密集排列且带有一定倾斜角度的书脊图像。本文提出了一个即插即用的增强全局信息的自注意模块；用DenseASPP模块替换Aspp模块来提取更密集、更广范围的书脊特征；在DenseASPP的支路上插入条纹池化模块，增强书脊的长条特性。大量实验结果证明，本文提出的算法可以增强原网络对密集、大长宽比、倾斜目标的分割效果，相比其他算法有较大的优势。同时该算法也可以扩展到航拍的规则目标分割、密集目标分割等场景。

参考文献

[1] TABASSUM N, CHOWDHURY S, HOSSEN M K, et al. An approach to recognize book title from multi-cell bookshelf images [C]// Proceedings of the 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR). Piscataway: IEEE, 2017:1-6.

[2] 康洪雷,牛连强,冯庸,等.基于视觉的错序在架图书检测系统 [J].软件工程,2018,21(04):18-22. (KANG H L, NIU Z Q, FENG Y, et al. A vision-based detection system for out-of-order books on shelves [J]. Software Engineering, 2018, 21(04):18-22.)

[3] 崔晨,任明武.一种基于文本检测的书脊定位方法 [J].计算机与数字工程,2020,48(01):178-182+251. (CUI C, REN M W. A method of

spine location based on text detection [J]. Computer and Digital Engineering, 2020,48(01):178-182+251.)
[4] NEVETHA M P, BARSKAR A. Automatic book spine extraction and recognition for library inventory management [C]// Proceedings of the 3rd International Symposium on Women in Computing and Informatics. New York: ACM, 2015:44-48.
[5] Uçkun F A, Özer H, Nurbaş E, et al. Direction finding using convolutional neural networks and convolutional recurrent neural networks [C]// Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU), Piscataway: IEEE, 2020:1-4.
[6] CAI W, HU D. QRS complex detection using novel deep learning neural networks [EB/OL]. (2020-05-25) [2022-12-18]. https://ieeexplore.ieee.org/document/9099511
[7] SAXENA N, K B N, RAMAN B. Semantic segmentation of multispectral images using res-seg-net model [C]// Proceedings of the 2020 IEEE 14th International Conference on Semantic Computing (ICSC), Piscataway: IEEE, 2020:154-157.
[8] ZHANG Z, LIU Q, WANG Y. Road extraction by deep residual U-Net [J]. IEEE Geoscience and Remote Sensing Letters, 2018, 15(5):749-753
[9] ZHOU Z, SIDDIQUE E M M R, TAJBAKHSH N, et al. UNet++: A nested U-Net architecture for medical image segmentation [EB/OL]. (2018-07-18) [2022-12-18]. https://arxiv.org/pdf/1807.10165.pdf
[10] CAO K, ZHANG X. An improved Res-UNet model for tree species classification using airborne high-resolution images [J]. Remote Sensing, 2020, 12(7):1128.
[11] CHEN L C, PAPANDREOU G, KOKKINOS I. Semantic image segmentation with deep convolutional nets and fully connected CRFs [EB/OL]. (2014-12-22) [2022-12-18]. https://arxiv.org/ pdf/1412.7062.pdf
[12] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848
[13] CHEN L C, PAPANDREOU G, SCHROFF F, et al. Rethinking atrous convolution for semantic image segmentation [EB/OL]. (2017-6-5) [2022-12-18]. https://arxiv.org/pdf/1706.05587.pdf
[14] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [EB/OL]. (2018-08-22) [2022-12-18]. https://arxiv.org/pdf/1802.02611.pdf
[15] XIE Y, ZHANG J, SHEN C, et al. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation [C]// Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention, Cham: Springer, 2021: 171–180
[16] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16X16 words: Transformers for image recognition at scale [EB/OL]. (2020-10-22) [2022-12-18]. https://arxiv.org/pdf/2010.11929v2.pdf
[17] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [EB/OL]. (2021-08-17) [2022-12-18], https://arxiv.org/pdf/2103.14030v2.pdf
[18] CHEN J, LU Y, YU Q, et al. Transunet: Transformers make strong encoders for medical image segmentation [EB/OL]. (2021-2-8) [2022-12-18]. https://arxiv.org/pdf/2102.04306v1.pdf
[19] AZAD R, HEIDARI M, SHARIATNIA M, et al. Transformer based Deeplab v3+ for medical image segmentation [EB/OL]. (2022-08-01) [2022-12-18]. https://arxiv.org/pdf/2208.00713.pdf
[20] SRINIVAS A, LIN T Y, PARMAR N, et al. Bottleneck transformers for visual recognition[C]// Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC: IEEE Computer Society, 2021: 16514-16524.
[21] 曾文雯,杨阳,钟小品. 一种用于在架图书书脊语义分割的山字形网络 [J].图像与信号处理, 2020, 9(4):218-225. (ZENG W W, YANG Y, ZHONG X P. A hierarchical network for semantic segmentation of backbone of books on shelf [J]. Image and Signal Processing, 2020, 9(4):218-225.)
[22] 曾文雯,杨阳,钟小品. 基于改进 Mask R-CNN 的在架图书书脊图像实例分割方法 [J].计算机应用研究, 2021,38(11):3456-3459+3505. (ZENG W W, YANG Y, ZHONG X P. Example segmentation method of spine image of on-shelf books based on improved Mask R-CNN [J]. Computer Application Research, 2021, 38(11):3456-3459+3505.)

This work is supported by Key public relations project of Liaoning Provincial Department of Education (LJKZZ20220033).

JI Xiaofei, born in 1978, Ph. D., associate professor. Her research interests include video analysis and processing, pattern recognition.
ZHANG Kexin, born in 1996, M. S. candidate. Her research interests include image processing, video analysis and processing.
TANG Lirong, born in 2000, M. S. candidate. His research interests include image processing, video analysis and processing.