引文格式：杨德志,陈永贵. 基于特征自适应选择网络的遥感图像语义分割[J]. 地理空间信息, 2025, 23(10):1-4. doi:10. 3969/j.issn. 1672-4623.2025.10.001

# 基于特征自适应选择网络的遥感图像语义分割

杨德志[1]，陈永贵[2]
（1. 北京东方至远科技股份有限公司，北京 100191；2. 河南测绘职业学院 河南 郑州 451464）

**摘　要：**语义分割能够完成像素级的区域标注任务，在遥感图像处理领域发挥着重要作用。当前遥感图像分割的主要方法是采用多尺度策略来提高分割网络的性能。然而，高分辨率遥感图像中地物目标的尺度不确定性难以用常规模型进行分割。针对这一问题，设计了自适应特征选择网络用于遥感图像语义分割，其中注意力模块学习每个特征块在不同尺度下的权重贡献。在实验的过程中采用金字塔场景解析网络（PSPNet），DeepLabV3和U-Net与所提出的方法在2个基准（Vaihingen数据集和WHU建筑数据集）上进行比较。实验结果和综合分析验证了该方法在遥感图像语义分割中的有效性和实用性。

**关键词：**自适应特征选择；遥感图像；语义分割；注意力机制
**中图分类号：**P237　　**文献标志码：**B　　**文章编号：**1672-4623（2025）10-0001-04

# Semantic Segmentation for Remote Sensing Images Based on Self-adaptive Feature Selection Network

YANG Dezhi[1], CHEN Yonggui[2]
(1. Beijing Vastitude Technology Co., Ltd., Beijing 100191, China;
2. Henan Collage of Surveying and Mapping, Zhengzhou 451464, China)

**Abstract:** Because semantic segmentation can complete pixel-level regional annotation tasks, it plays an important role in the remote sensing image processing field. The major current remote sense image segmentation method is using multiple scales strategy to improve the performance of segmentation networks. However, the ground objects with uncertain scale in high-resolution aerial images are difficult to be segmented with conventional models. To address this problem, we designed a self-adaptive feature selection network, in which attention module learned weight contributions of each feature blocks in different scales. We compared the pyramid scene parsing network (PSPNet), DeepLabV3, U-Net with the proposed network on two benchmarks (the Vaihingen set and the WHU Building data set). The experimental results and comprehensive analysis validate the efficiency and practicability of the proposed method in semantic segmentation of remote sensing images.

**Key words:** self-adaptive feature selection, remote sensing image, semantic segmentation, attention mechanism

语义分割在遥感领域具有重要意义。遥感图像分析已成为一个活跃的研究课题，并广泛应用于环境监测、城市规划、土地利用管理等领域。许多传统的分割方法使用手工设计的特征提取器以获得分割的结果。然而，该方法难以对高分辨率遥感图像中地物复杂的区域变化有显著的表现。

近年来，随着深度学习的发展，出现了许多可以用作编码骨干的最先进的网络，例如VGG[1]，ResNet[2]和Inception[3]。语义分割作为计算机视觉的一个重要分支，在图像理解中起着极其重要的作用，并已经构造了一系列基于高性能卷积神经网络（CNN）的模型，以在许多视觉任务中实现出色的性能，例如全卷积网络（FCN）[4]，UNet[5]，DeepLab[6]，金字塔场景解析网络（PSPNet）[7]等。当前，许多方法已经通过在高分辨率遥感图像领域中使用这些现有的深度学习模型[8-10]（例如，Hourglass、U-Net、DenseNet等）进行了尝试。最近，Zhang[11]等提出了一种基于恢复策略的尺度感知模块（SAM），旨在使像素在不同尺度下调整其位置，以捕获高分辨率遥感图像的尺度不一致特征。Ya[12]等提出了一种集成网络，它结合了空间金字塔池（ASPP）网络和编码器-解码器结构的优点，其目标不仅是提取多尺度上下文信息，而且同时恢复更清晰的目标边界。为了解决由于使用下采样操作而导致的细节信息丢失的问题，Li[13]等提出了一种最大融合方案，旨在通过有效结合来自深层的语义信息和来自浅层的细节信息，利用数字表面模型（DSM）信息

项目来源：河南省软科学研究计划资助项目（232400410141）。
第一作者简介：杨德志（1992—），硕士，工程师，从事摄影测量与遥感工作，E-mail：zhi0114@tom.com。

进一步改善语义分割结果；Cao[14]等提出了DSM融合（DSMF）分支结构模型。

金字塔池化层（PPL）[15]与DeepLab[16]在遥感图像的特征表达方面展现出色的学习能力。PPL能够通过不同尺度提取多层次的语义信息，并融合全局上下文；而DeepLab则依靠多尺度的扩张卷积（ASPP）扩展感受野。尽管如此，这2种方法都存在一个共性问题：通常需要根据经验手动调整多尺度或扩张卷积的参数。针对这一不足，本文提出了一种自适应特征选择（AFS）模块。此前，Li[15]等提出了选择性内核（SK）单元，通过融合不同卷积核大小的多分支结构，并借助softmax注意力机制，实现了不同感受野的自适应融合。受到SK单元的启发，本文在PPL和DeepLab结构中引入了动态选择机制，使得神经元能够根据不同尺度或感受野，自主从多种特征图中选择最优的特征信息。

## 1 研究方法

### 1.1 特征自适应选择

为了让模型能够自动筛选有效的特征信息，本文设计了一个AFS模块。该模块可以将不同感受野下获得的多个特征图块通过注意力机制进行自适应融合。如图1所示，假设输入的特征图块为表示输入的通道数，分别代表空间尺寸，具有不同感受野的特征图块数量。该模块首先进行逐元素求和运算。

$$\tilde{X}=\sum_{i=1}^{s}X_i \tag{1}$$

然后，使用全局平均池化来捕获全局特征信息，以生成通道统计量 $S$，且 $S\in\Re^{c\times 1}$。

$$S=\mathrm{GP}\left(\tilde{X}\right) \tag{2}$$

一般情况下，特征向量首先输入全连接层（FC），随后经过批量归一化（BN）操作和修正线性单元（ReLU）激活函数，从而实现特征的自适应筛选。其具体计算方式如下：

$$\hat{S}=\sigma\left(B\left(\mathrm{FC}(S)\right)\right) \tag{3}$$

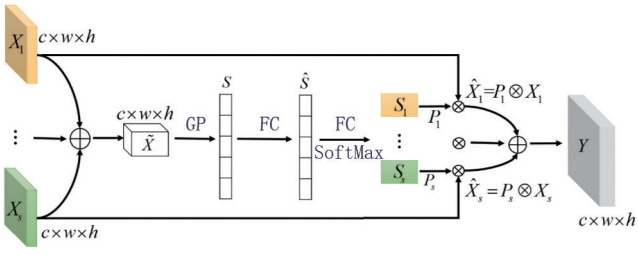



图1 自适应特征选择模块

为进一步提升效率，本模块首先在第一个全连接层（FC）中对输出维度进行压缩，随后利用第二个FC层（不包含BN和ReLU）对维度进行调整，以对应不同数量的特征图块。为分析不同感受野特征图的作用，引入了跨通道的软性注意机制，通过SoftMax操作自适应地生成各自的权重。

$$P_i=\sum_{i}^{s}S_i \tag{4}$$

式中，$P_i$为第$i$个特征图块的注意力权重。在接下来的步骤中，使用矩阵乘法变换来获得新的特征图块$\hat{X}$，其具有不同感受野上的注意力权重。其计算公式为：

$$\hat{X}=\left\{X_i\cdot P_i,i\in[1,s]\right\} \tag{5}$$

通过上述方式，神经元可以通过学习注意力权重来实现自动特征选择。最后，通过逐元素求和完成特征选择。

$$Y=\sum_{i=1}^{s}\hat{X}_i \tag{6}$$

借助AFS模块进行特征筛选，注意力机制能够学习不同感受野下各特征图块的权重分配。最终输出的特征图融合了多分支特征，有效减少了冗余信息，达到了自适应特征选择的目的。挤压与激励网络（SENet）采用门控机制，根据通道的重要性自适应地重新调整特征通道的响应。而相较之下，本文提出的方法则是利用注意力机制，关注不同分支的权重分布。

### 1.2 PSPNet-AFS

为评估AFS模块的有效性，本文选用PSPNet[7]和ResNet50[2]作为编码部分。PSPNet是一种基于金字塔池化（PPL）的场景分割网络，能够为像素级场景解析提供全局上下文信息，并且可以提取多层次的特征，多分辨率特征对于提升网络表现有着关键作用。PSPNet会生成4种不同尺度的特征图。已有研究表明，PPL在多个常用数据集上都具有良好的效果。在实验过程中发现，基础PSPNet与加入AFS模块后的PSPNet性能差异明显。图2a展示了AFS模块嵌入PSPNet的位置。为此，本文提出了一种融合PSPNet和AFS模块的新模型，即PSPNet-AFS。在本研究中，我们采用了大于4个尺度的感受野对遥感图像进行分割，以获得更具代表性的特征。在设置特征图块数量时，本文选取了32作为比例。辅助监督作为深度学习优化的重要手段，为了平衡主损失和辅助损失，在训练时设置辅助损失权重为0.2。

### 1.3 DeepLabV3-AFS

为进一步验证所提模块的泛化能力，本文选用

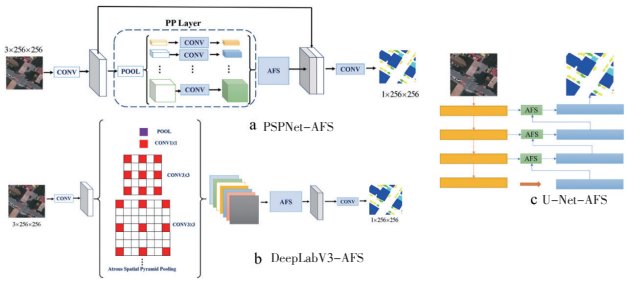



图2　基于AFS的语义分割网络的体系结构

DeepLabV3作为基线，对方法进行补充实验。图2b展示了结合AFS模块的DeepLabV3模型结构。作为高分辨率航空影像领域的热门课题，多尺度分割常通过多种膨胀率来实现。在本研究中，AFS模块能够针对不同空洞率进行自适应选择，从而有效扩大卷积核的接收范围，使多尺度的全局与局部特征信息得以充分融合。为此，模型采用了不同速率来生成丰富且具有代表性的感受野特征图，并通过AFS模块进行自适应融合。具体实验中，分别选取8、16和32为最大膨胀率，步长设置为1、2和4。在训练过程中，辅助分支损失的权重设定为0.2。

### 1.4　U-Net-AFS

图2c给出了U-Net-AFS的网络结构示意。PSPNet和DeepLabV3主要通过多尺度特征映射（大于2个尺度）来捕获图像的语义信息。U-Net则利用skip-connection结构，将上下文信息进行融合，并把浅层的高分辨率特征与解码阶段的上采样特征结合，从而提升特征图的信息丰富度。因此，本文提出的结构能够自适应地在下采样和上采样阶段选择融合上下文信息。

## 2　实验与分析

### 2.1　实验数据集

Vaihingen数据集[15]包含33个高分辨率航空图像（不同大小，平均大小约为2 500×2 000）。每个图像具有3个波段，包括红外、红色和绿色波段（IRRG）。在这些数据集中，有16张地面实况图像可用于训练，其余的用于测试。地物目标包含5个类别，包括不透水表面、建筑物、低矮植被、树木和汽车。请注意，只使用IRRG图像与多光谱信息且在实验中没有DSM信息。在这项工作中，将训练数据分为训练集和验证集进行实验，所有训练数据都用于测试阶段的训练。此外，在WHU建筑数据集[16]上测试了该方法，以进一步评估基于AFS的模型的性能。实验中使用了卫星数据集I（全球城市），其中包含204幅图像（512像素×512像素，分辨率从0.3~2.5 m）。

### 2.2　Vaihingen数据集比较

将所提出的算法与几种主流分割网络进行对比，以验证其有效性。表1中汇总了在Vai-hingen测试集上的实验结果。为了保证对比的公平性，所有模型均采用相同的训练策略。实验显示，基于AFS的模型整体优于基线网络。尤其是Deep-LabV3-AFS（32, 4）在mIoU上相比原始DeepLabV3提升了约1.09%，且参数数量更少。值得注意的是，对于小目标“Car”的识别准确率提升了4.3%，这说明AFS模块能够自适应地从多种感受野中选取有效特征。此外，PSPNet-AFS同样取得了性能提升，其中PSPNet-AFS（32, 4）的mIoU、*F*1和*OA*分别提高了0.44%、0.27%和0.24%，进一步验证了方法的有效性与可扩展性。我们还对比了不同尺度特征图数量的影响，如采用8个尺度的PSPNet与PSPNet-AFS（8, 1）相比，所提出方法的mIoU提升了0.4%，*F*1和*OA*分别提升了0.37%和0.28%。这进一步证明了AFS模块无需依赖更多尺度即可有效提升遥感影像的分割性能。

### 2.3　WHU Building数据集比较

为了更好地展示基于AFS的模型的性能，测试所提出的方法在卫星数据集I（全球城市）数据集，所有模型都使用预训练的ResNet18作为编码骨干。表2描述了模型中使用的不同注意力机制的比较。PSPNet-AFS通过使用挤压和激励（SE）模块[16]比PSPNet在相同尺度下获得了更好的结果。虽然所提出的模型并没有显著提高分值，但在AFS策略下，分割模型能够以更少的参数获得更好的性能和更强的泛化能力。

### 2.4　可视化分析

以上进行了大量的实验来评估所提出的方法的性能，我们的模型优于基线和现有的网络具有更好的性能，但参数较少。为了更好地描述所提出方法的有效性，利用从DeepLabV3-AFS，PSPNet-AFS，U-Net-AFS及其基线获得的预测结果来呈现Vaihingen测试集上的可视化结果。如图3所示，可以看到Deep-LabV3-AFS，PSPNet-AFS和U-Net-AFS显著提高了预测的结果。在图3的第一个示例中，可以注意到，传统的PSPNet和DeepLabV3无法有效地分割红色框中标记的阴影区域。此外，还注意到测试集中有一些错误标记的数据，例如图3中的sample-2，AFS模块实现了更好的结果。在实验过程中，由于训练集存在误标记数据的问题，使得“低植被”和“树木”的分类比较困难，而

本文提出的方法仍能较好地实现分割任务。

表1 Vaihingen数据集5个类别的实验结果/%

| Networks | mIoU | *F*1 | *OA* |
|---|---|---|---|
| SANet-resnet101[11] | 71.29 | 82.67 | 86.47 |
| HSN+OA_WBP[8] | – | 83.41 | 85.39 |
| SegNet[8] | – | 81.40 | 84.07 |
| DenseU-Net[9] | – | 84.95 | 85.63 |
| FPL[10] | – | 78.56 | 83.69 |
| DeepLabV3-AFS(32,4) | 75.70 | 85.96 | 87.39 |
| PSPNet-AFS(32,4) | 75.85 | 86.07 | 87.55 |

表2 卫星数据集I（全球城市）数据集不同比例的结果/%

| Networks | mIoU | *F*1 | *OA* |
|---|---|---|---|
| U-Net | 75.00 | 85.22 | 89.03 |
| U-Net-AFS | 75.65 | 85.66 | 89.44 |
| PSPNet | 72.07 | 83.08 | 87.82 |
| PSPNet(8,1) | 72.26 | 83.25 | 87.82 |
| PSPNet(16,2) | 71.98 | 83.04 | 87.61 |
| PSPNet-AFS | 72.17 | 83.15 | 87.88 |
| PSPNet-AFS(8,1) | 72.37 | 83.34 | 87.72 |
| PSPNet-AFS(16,2) | 72.25 | 83.28 | 87.50 |
| DeepLabV3 | 74.75 | 85.03 | 88.98 |
| DeepLavV3(32,4) | 74.59 | 84.94 | 88.79 |
| DeepLabV3-AFS | 75.23 | 85.41 | 89.04 |
| DeepLabV3-AFS(32,4) | 75.05 | 85.27 | 88.96 |

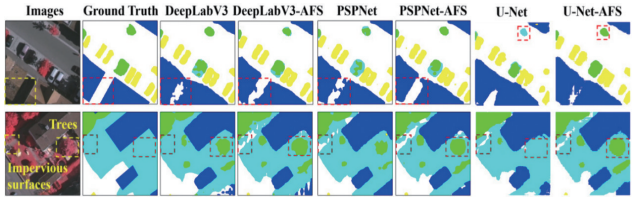



图3 Vaihingen测试集的可视化结果

注：水面：白色；建筑物：深蓝色；植被：浅蓝色；树木：绿色；汽车：黄色。

## 3 结 语

本文采用AFS方法来有效地学习不同感受野上通道之间的权重贡献。通过多个尺度使用由这些特征图中的信息指导的Softmax注意力，产生不同大小的通道之间的有效感受野。使用PSPNet，DeepLabV3和U-Net作为基线进行了广泛的实验来估计所提出的方法。在Vaihingen和WHU建筑数据集上的对比实验结果表明了该方法的有效性，同时对比研究也表明了该方法的稳健性。在未来，将使用更多的基线与建议的模块，以实现更高甚至更快的分割性能。

## 参考文献

[1] Shen X, Wang D, Mao K. Inundation Extent Mapping by Synthetic Aperture Radar:A Review [J]. Remote Sensing, 2019,11(7):879

[2] Zhi Y L, Shi W, Benediktsson J A. A Modified Mean Filter for Improving the Classification Performance of Very High-resolution Remote Sensing Imagery [J]. International Journal of Remote Sensing,2018,39(3):770-785

[3] Chen K, Zou Z, Shi Z. Building Extraction from Remote Sensing Images with Sparse Token Transformers [J].Remote Sensing,2021,13(21):4 441-4462

[4] Zheng C, Wang L. Semantic Segmentation of Remote Sensing Imagery Using Object-based Markov Random Field Model with Regional Penalties [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,2015,8(5):1 924-1 935

[5] Chen D, Shang S, Wu C. Shadow-based Building Detection and Segmentation in High-resolution Remote Sensing Image[J]. Journal of Multimedia,2014(1):181-188

[6] Raziq A, Xu A, Li Y. Automatic Extraction of Urban Road Centerlines from High-resolution Satellite Imagery Using Automatic Thresholding and Morphological Operation Method[J]. Journal of Geographic Information System, 2016,8(4):517-525

[7] Pang Y, Sun M, Jiang X, et al. Convolution in Con-volution for Network in Network[J]. IEEE Transactions on Neural Networks and Learning Systems,2018,29(5):1 587-1 597

[8] Shelhamer E, Long L, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transac-tionson Pattern Analysis and Machine Intelligence,2017,39(4): 640-651

[9] Badrinarayanan V, Kendall A, Cipolla R. SegNet:A Deep Convolutional Encoder-decoder Architecture for Image Segmentation[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2017,39(12):2 481-2 495

[10] Wang X, Wang E, Zhao K, et al.FSL-UNet:Full-scale Linked Unet with Spatial-spectral Joint Perceptual Attention for Hyperspectral and Multispectral Image Fusion[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022,60(1):1-14

[11] Zhang Z, Wang X, and Jung C. DCSR:Dilated Convolu-tions for Single Image Super-resolution[J]. IEEE Trans-

（下转第38页）