

# 深度学习在高分辨率遥感图像语义分割中的算法研究

邓露露, 张长伦, 邢 思

北京建筑大学理学院, 北京

收稿日期: 2022年11月3日; 录用日期: 2022年11月16日; 发布日期: 2022年11月25日

## 摘 要

遥感图像语义分割是为遥感图像分配像素级语义标签的计算机视觉任务。随着传感器技术以及深度学习的发展, 深度学习算法在精度与速度上远超传统算法。其中, 基于深度学习的高分辨率遥感图像语义分割的算法成为众多学者的主要研究方向之一。本文主要针对深度学习在遥感图像语义分割中的相关算法以及网络结构进行介绍。首先介绍语义分割CNN网络, 其次分别从三个方面对高分辨率遥感图像语义分割算法进行阐述: 一是结合多尺度、多阶段、上下文聚合策略, 二是在语义分割之后采用后处理技术, 三是结合注意力机制。随后介绍经典数据集, 最后对未来深度学习在高分辨率遥感图像语义分割中的算法的发展进行总结与展望。

## 关键词

深度学习, 高分辨率遥感图像, 语义分割

# Algorithm Research of Deep Learning in High-Resolution Remote Sensing Image Semantic Segmentation

Lulu Deng, Changlun Zhang, Si Xing

School of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: Nov. 3<sup>rd</sup>, 2022; accepted: Nov. 16<sup>th</sup>, 2022; published: Nov. 25<sup>th</sup>, 2022

## Abstract

Remote sensing image semantic segmentation is a computer vision task to assign pixel level se-

mantic labels to remote sensing images. With the development of sensor technology and deep learning, deep learning algorithm is far superior to traditional algorithms in accuracy and speed. Among them, the algorithm of high-resolution remote sensing image semantic segmentation based on deep learning has become one of the main research directions of many scholars. This paper mainly introduces the related algorithms and network structure of deep learning in remote sensing image semantic segmentation. First, the semantic segmentation CNN network is introduced, and then the semantic segmentation algorithm of high-resolution remote sensing images is described from three directions: first, combining multi-scale, multi-stage, context aggregation strategies, second, using post-processing technology after semantic segmentation, and third, combining attention mechanism. Then we introduce the classical datasets, and finally, summarize and prospect the development of deep learning algorithm in the high-resolution remote sensing images semantic segmentation in the future.

## Keywords

Deep Learning, High-Resolution Remote Sensing Image, Semantic Segmentation

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

拍摄技术和卫星传感器技术的快速发展,为解析高分辨率遥感图像开辟了新的视野[1]。许多基本的计算机视觉任务,如目标检测[2]、图像匹配[3]、语义分割[4]等都与之密切相关。遥感图像语义分割是一项关键且具有挑战性的任务,特别是高分辨率遥感图像的语义分割,在道路提取[5]、城市规划[6]和土地覆盖分类[7]等方面具有越来越重要的应用意义。

传统的遥感图像语义分割方法主要分为两类:一类方法基于区域合并法,另一类是基于机器学习的方法。区域合并法主要依靠手工制作特征,通常包括两个过程:初始图像过分割和最优分割[8]。区域合并方法的有效性在很大程度上取决于建立一个能够反映相邻区域相似性的合并标准。Maxwell 等人[9]和 Ton 等人[10]选择低或中空间分辨率遥感图像的光谱特征作为合并标准。然而,光谱特征在高空间分辨率的图像中比较复杂,这些方法导致较大的类内方差和较小的类间方差,因此仅利用光谱特征进行分割是无效的[11]。Lu 等人[12]和 Yang 等人[13]提取纹理特征用于高空间分辨率图像的分割。然而,这些手工制作的特征是昂贵的,并且需要一个特殊的设计。如何自动提取有区别的特征,特别是在复杂的场景中,也是一个挑战。另一类传统的语义分割方法主要是基于机器学习方法,如支持向量机(SVM) [14],随机森林(RF) [15],条件随机场(CRFs) [16],神经网络(ANNs) [17],包括多层感知(MLP) [18]。这些分类器主要依赖于为像素级土地覆盖分类所提取的特征的质量。然而,这种对手工提取特征的高度依赖限制了这些传统方法的灵活性和适应性[19]。上述两类传统的语义分割方法,由于从图像中所提取的特征都比较浅层,并不适合目前具有高分甚至超高分辨率的遥感图像。

近年来,由于深度学习有强大的特征提取和对象表示能力,因此被广泛应用于图像处理和计算机视觉领域。目前,国内外学者提出了许多基于深度学习的遥感图像语义分割方法。针对超高分辨率遥感图像,语义分割任务主要面临的挑战大致可以分为三类:1) 是提高语义分割的精度;2) 是对非常规数据的分析;3) 是缺乏训练数据集。其中,针对如何提高语义分割精度,获得超高质量的分割结果的研究最为广泛,主要解决策略有:结合多尺度、多阶段、上下文聚合策略;在语义分割之后采用后处理技术;结

合注意力机制。在本文的其余部分中，第二章介绍语义分割任务中经典的 CNN 网络框架，第三章将围绕上述三种策略的深度学习方法进行梳理与讨论，第四章介绍经典的数据集，最后第五章对未来深度学习在高分辨率遥感图像语义分割中的算法的发展进行总结与展望。

## 2. 语义分割经典的 CNN 网络框架

全卷积网络(FCN) [20]是用于图像高质量语义分割的第一个 CNN 架构。基于 FCN 的改进网络也相继被广泛提出，其中大多数网络都采用编码器 - 解码器架构，例如 SegNet [21]和 U-Net [22]等。通常，基于 FCN 的体系结构从输入图像中提取信息，进而生成高级特征图，接着利用高级特征图重建掩模，通过单个或多层上采样过程进行像素级分割。DeepLab [23] [24] [25] [26]使用空洞卷积来放大滤波器的视野。虽然很多学者基于经典网络进行了改进，但是针对具有精细细节的高分辨率遥感图像，特征表示需要更高的分辨能力，运行也需要非常高的 GPU 内存，所以大部分改进依旧无法生成高质量的分割结果。

随着语义分割在许多实时、低延迟的应用程序(如自动驾驶)中变得越来越重要，高效或快速的分割模型得到了越来越多的关注。ENet [27]使用了具有早期降采样的非对称编解码器结构，以减少浮点操作。ICNet [28]在适当的标签引导下，从多分辨率分支进行级联特征图，以及模型压缩。

然而，当高分辨率遥感图像的应用越来越广泛时，经典的语义分割深度神经网络框架无法得到高质量的分割结果，因此，越来越多的针对高分辨率遥感图像的语义分割网络被提出，并且不断被改进，以提高分割的精度。

## 3. 历史发展与现状

为了减少卷积中空间细节的损失，从而提高像素级分割的精度，主要有三种策略：一是结合多尺度特征、多阶段网络或者上下文聚合，二是使用后处理技术增强分割结果，三是结合注意力机制。在本节的其余部分中，我们将梳理与每个策略相关的方法，并解释其关键思想。

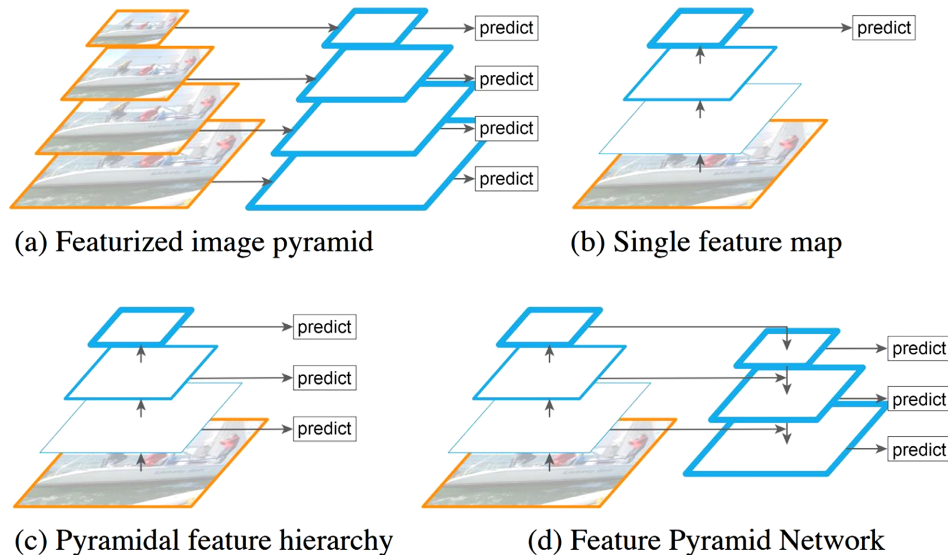
### 3.1. 多尺度、多阶段、上下文聚合

#### 3.1.1. 多尺度策略

结合多尺度的特征信息可以帮助网络聚合不同的视野，并为每个像素提供更多的上下文信息[29] [30]。识别不同尺度的特征是计算机视觉中的一个基本挑战，常见的特征金字塔网络结构如图 1 所示 [31]。图 1(a)是对图像做一个特征金字塔，在每一个图像上做 CNN，进行预测，但这种结构的计算与内存开销太大。与 Featurized image pyramid 不同，图 1(b)对图像做一个 CNN，相当于一个特征金字塔，在最高层做一个预测，使用 Single feature map 结构的网络包括 R-CNN、SPPNet、Fast R-CNN、Faster R-CNN、YOLOv1 以及 R-FCN 系列。但这存在感受野问题，越抽象感受野越大，导致出现小特征漏检的问题。图 1(c)在每一层特征做预测，使用 Pyramidal feature hierarchy 结构的网络有 SSD。该结构的问题在于底层特征图语义信息薄弱，虽然与 Single feature map 结构相比能够更好地检测出小物体，但语义信息少的小物体容易被错分。图 1(d)特征金字塔网络(FPN)相当于先进行传统的自上而下的特征卷积(图 1(d)左侧)，然后试图融合左侧特征图的相邻特征图，这样 FPN 对不同尺度的特征图进行上采样，将语义信息多的高层特征和语义信息少但位置信息多的低层特征进行了融合。

除了上述的四种特征金字塔网络结构以外，还有其他经典网络架构结合了多尺度策略。ICNet [28]使用了一种级联体系结构来处理不同的降采样输入的特征图。RefineNet [32]引入了一种多路径细化块，通过上采样低分辨率特征来结合多尺度特征。DeepLab [24]为了减少降采样造成的细节损失，空洞空间金字塔通过使用在多个采样率滤波器探测之前的卷积层来描述不同尺度上的对象。PSPNet [33]也达到了同样的效果，它结合了不同的比例特征图来扩大感受野。HRNet [34]提出了另一种规模融合模式，在每个阶

段后添加一个具有更大感受野的新分支。Zhao 等人[35]提出了一个多尺度卷积神经网络(MCNN)来学习具有空间关系的深度特征。MCNN 从图像中构建了一个金字塔结构,在不同的尺度上呈现空间特征。将高级空间特征与光谱特征连接起来,形成一个用于训练逻辑回归的数据集,最终结果以多数票产生。Cheng 等人[36]扩展了 SegNet,通过构建边缘检测网络和分割网络,提出了一种具有边缘感知能力的卷积网络。利用分割网络提取不同尺度下的语义特征,用于边缘检测网络的训练,进而利用边缘检测网络中的边缘图对网络进行微调。类似地,Marmanis 等人[37]通过添加边界检测扩展了 SegNet 编码-解码器结构,这使得对象边界在模型中以像素级轮廓似然的形式明确。



**Figure 1.** Pyramid networks with different feature  
**图 1.** 不同的特征金字塔网络

由于地理对象在高分辨率遥感图像中具有不同尺度和各种分布的特征,它们在获得语义间连续性和边界外精度方面仍面临挑战。虽然上述网络架构都结合了多尺度策略,但是这些模型无法很好地进行处理高分辨率遥感图像语义分割。受迁移学习的启发,Chen 等人[38]提出了一个高效语义分割框架 SMAF-Net,该框架将多尺度对抗特征共享为 U 形语义分割模型。具体而言,利用训练良好的生成对抗网络获得的多尺度对抗特征表示,掌握像素相关性,进一步提高多尺度地理对象的边界精度。Ma 等人[39]提出一种具有空洞卷积的多尺度跳跃连接网络,以解决多模态、多尺度高分辨率遥感影像的分割问题。首先,该网络在编码器中应用了空洞卷积来扩大卷积核的感受。其次,基于 U-Net 网络,该网络通过重新设计跳过连接并结合每个 U-Net 层中的多尺度特征,合并了不同尺度的光和深特征。最后,采用逐个像素分类的方法,得到遥感影像的语义分割结果。

### 3.1.2. 多阶段网络

除了结合多尺度策略,另一种处理高分辨率图像的方法是使用多阶段网络,图像在几个阶段或子网络之后进行分割。Xia 等人[40]提出了分层自动缩放网,这是一种在滑动窗口通过大图像时缩放视场的策略。对于超高分辨率图像,Takahama 等人[41]通过在分割前预测整个补丁是否包含前景像素,解决了背景和前景之间的不平衡问题。为了同时处理大型人工和精细结构的对象,Liu 等人[42]提出了一种自级联网络,使用空洞卷积在编码器的最后一层进行多尺度表示。除了包含更大范围的上下文信息外,多尺度表示还集成了上下文的层次依赖关系。许多扩张速率用于提取缩放特征,其结果通过跳跃连接和由粗到

细的方式实现细化目标对象。

### 3.1.3. 上下文聚合

对于高分辨率遥感图像，一般的语义分割网络无法直接处理。一种有效的预处理方法是将整个图像切分成若干个小块，接着对每个图像块进行语义分割，最后将所有图像块的分割图进行拼接还原到原图大小。然而，该方法在对每个图像块进行语义分割存在着感受野小导致精度低的问题，因此将全局上下文传播到局部是处理高分辨率图像语义分割的一种有效的方法。ParseNet [43]将全局上下文汇集到局部视场中，以获得更多信息。BiSeNet [44]还包含了一个全局池化分支，将全局上下文在最后阶段添加到特征图中。虽然这些方法是有效的，但它们需要大量的 GPU 内存。Tokunaga 等人[45]提出了一种超高分辨率图像语义分割的方法，利用独立的多尺度网络和自适应高分辨率权重，将网络的输出与相应的训练权重相结合，得到最终的输出，但该方法存在网络分支之间没有信息共享的缺点。

随着超高分辨率图像语义分割的要求越来越高，对算法精度和效率的要求也就越来越高。目前针对高分辨率遥感图像语义分割的方法主要分为两类：一是对一个超高分辨率的图像进行降采样，二是将其裁剪成小块进行单独处理。但是无论哪种方式，局部细节或全局上下文信息的丢失都导致了分割精度不高。针对这个问题，GLNet 网络[46]提出了一种高内存率的方式，有效地保存了全局和局部信息。GLNet 由一个全局分支和一个局部分支组成，分别将降采样的整个图像及其裁剪后的局部斑块作为各自的输入。在分割时，GLNet 深度融合了来自两个分支的特征映射，从放大的局部斑块中捕获高分辨率的精细结构，并从降采样的输入中捕获上下文依赖关系。与[45]不同的是，GLNet 包含两个具有共享信息的子网络，其中全局分支利用降采样提取全局上下文，局部分支通过斑块和相应的全局特征来提高高分辨率图像的细节。GLNet 网络使得分割性能明显提升，这体现了上下文信息对提高分割性能的重要性。GLNet 网络的大体框架如图 2 所示。

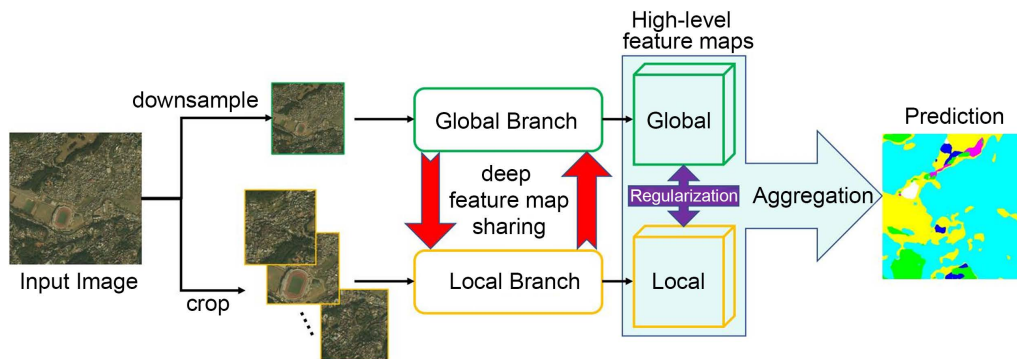


Figure 2. GLNet network framework  
图 2. GLNet 网络框架

但是，由于全局分支和局部分支的特别组合，很难扩展到两个以上的规模，而且 GLNet 网络的特征共享方案并没有将局部特征与全局特征关联起来，因此不能很好地利用它们的相关性，这使得它们的模型过于复杂。为了充分利用超高分辨率地理空间图像中的丰富信息，FCtL 网络[47]提出了一种具有局部感知上下文方案的超高分辨率地理空间图像分割模型。该模型使用了高分辨率图像分割管道，将超高分辨率图像分割成规则的斑块进行局部分割，然后将局部结果合并为高分辨率语义掩码。特别地，该网络引入了一种新的基于局部感知上下文的分割模型来处理局部斑块，其中局部斑块与其各种上下文之间的相关性被联合和互补地用来处理变化较大的语义区域。此外，该网络还提出了一个上下文语义细化网络，将局部分割结果与其上下文语义关联起来，因此在生成最终高分辨率掩模的过程中，具有减少边界伪影

和细化掩模轮廓的能力。FCtL 网络的大体框架如图 3 所示。

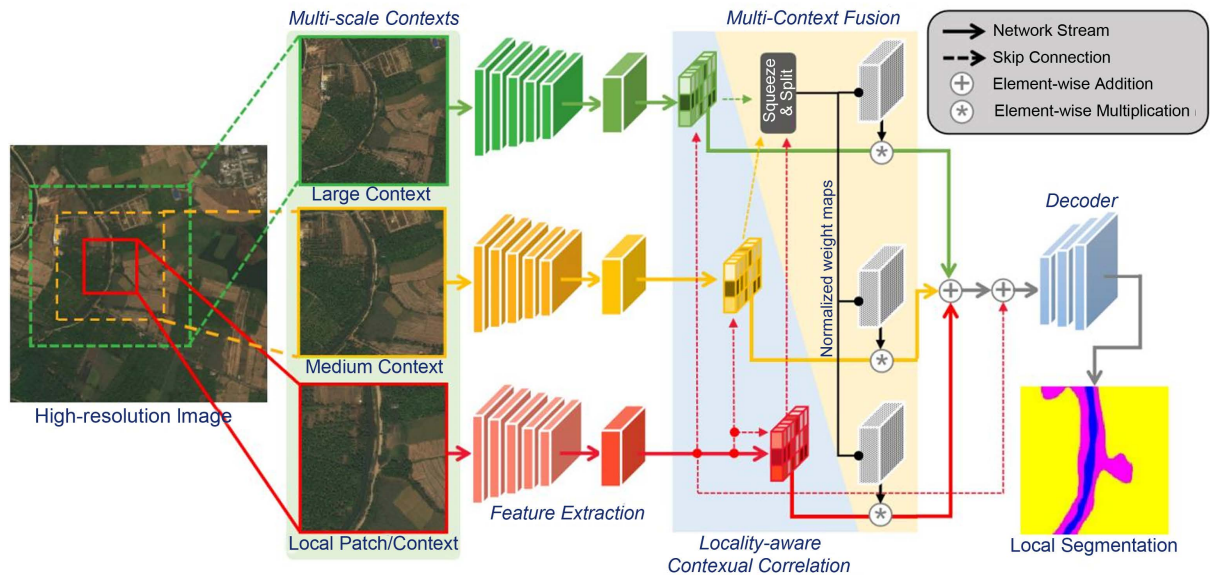


Figure 3. FCtL network framework  
图 3. FCtL 网络框架

除了以上的算法外, Bai 等人[48]提出了一个用于高分辨率遥感图像语义分割的分层上下文聚合网络 (HCANet)。该网络具有类似于 U-Net 的编码器 - 解码器结构, 它设计了两个紧凑的空洞空间金字塔 (CASPP 和 CASPP+)模块。CASPP 模块取代了 U-Net 中的复制和裁剪操作, 以提取 ResNet 多语义特征的多尺度上下文信息。CASPP+模块嵌入在 HCANet 解码器的中间层, 以提供上下文信息的强大聚合路径。在 HCANet 的解码器中, CASPP 模块获得的多尺度上下文信息被逐层分层合并, 用于高分辨率遥感图像的语义分割。

### 3.2. 后处理技术

为了提高语义分割的精度, 除了结合上述策略, 还可以通过引入后处理技术来改进分割输出。一种后处理技术是在深度学习网络产生的分割掩模上使用经典方法, 如条件随机场(CRFs) [16]或引导滤波器 (GF) [49]。然而, 这些方法是缓慢的, 而且改进是渐进式的。进而基于深度学习版本的引导过滤器(DGF) [50]被提出, 该方法可以提高推理速度。另一种进行后处理的方法是使用深度网络, 例如迭代实例分割 (ISS) [51]多次将输入图像和分割图通过细化模块对输出进行细化。该方法基于自反射, 每个细化阶段的输入图像相同。与 ISS 一样, 级联 PSP [52]使用了相同的细化方案, 但在每个细化阶段输入的分辨率是不同的。然而, 这些方法在任何中间阶段的错误预测都会显著影响后续步骤的性能。还有些方法旨在只改进输出的部分, 例如 PointRend [53], 改进在边界处或者预测不确定处的像素分类结果。PointRend 网络的大体框架如图 4 所示。

然而, 一些边界细化方法未能恢复小对象, 而且 PointRend 只使用局部上下文进行细化。此外, 由于 PointRend 的输入是一个深度网络的高级特征, 因此必须针对每个分割主干进行专门的训练。在此基础上, MagNet [54]提出了一个具有任意数量尺度级别的模块化框架。它对于细化粗分割输出简单但有效, 能够保持粗分割输出的整体结构, 同时每个阶段后添加更多的细节。MagNet 有多个处理阶段, 其中每个阶段对应一个放大级别, 一个阶段的输出输入到下一个阶段进行由粗到细的信息传播。每个阶段以比

上一阶段使用更高分辨率解析图像，恢复之前由于降采样步骤而丢失的细节，并通过多级处理阶段逐步细化分割输出。MagNet 网络的大体框架如图 5 所示。

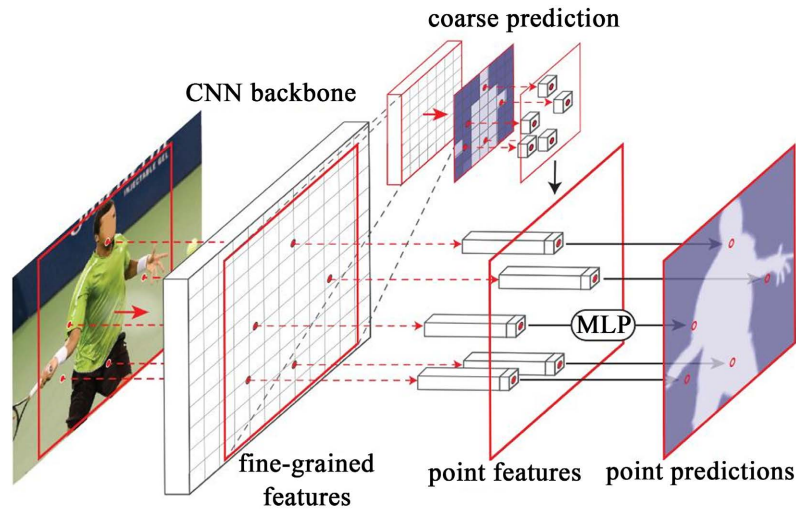


Figure 4. PointRender network framework  
图 4. PointRender 网络框架

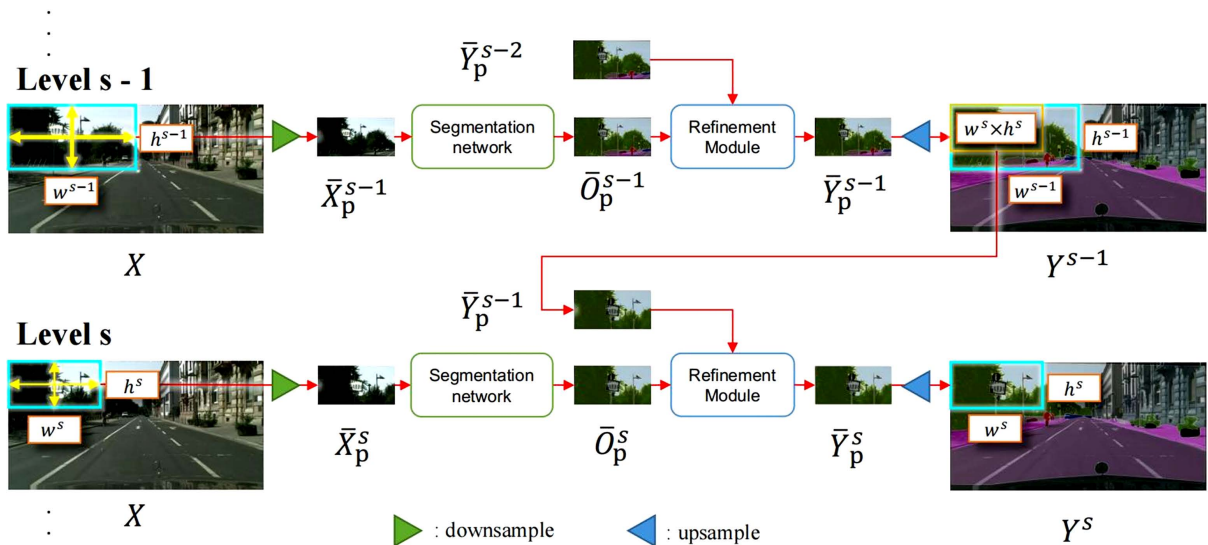


Figure 5. MagNet network framework  
图 5. MagNet 网络框架

### 3.3. 结合注意力机制

近几年，随着注意力机制的发展，更多的学者结合注意力机制来获得多尺度的特征，以提高分割精度。由于图神经网络可以捕获全局上下文表示，利用远程像素的相关性，从而提高语义分割性能。为此，Zi 等人[55]提出一种新型的自建图注意力神经网络。首先，利用 ResNet50 作为特征提取网络的骨干，获取遥感影像的特征图。其次，利用影像特征图构建了像素依赖关系图，设计图注意网络，从而获得遥感影像像素的相关性。第三，通道线性注意力机制获得了图像的通道依赖性，进一步提高了语义分割的预测能力。Lv 等人[56]提出了一种用于语义分割的新型多尺度特征聚合轻量级网络 MFALNet。与标准卷积不同，非对称

深度可分离卷积残差(ADCR)单元用于减少网络的参数,虽然使优化的结构更深,但整体更轻更简单。所提出的网络为编码器-解码器结构,其中在编码器和解码器中均实现多尺度特征聚合。空间自注意块有助于捕获远程上下文信息,并且门控卷积模块进一步用于在解码器中触发高级和低级特征映射时优化特征。

#### 4. 经典数据集

经典的图像语义分割的公共数据集(表 1)有 DeepGlobe、ISIC、Inria Aerial、Cityscapes、CamVid、COCO-Stuff、VOC2012,其中前三个数据集属于超高分辨率。

DeepGlobe 土地覆盖分类数据集是第一个提供关注农村地区的高分辨率亚米卫星图像的公共基准测试。DeepGlobe 提供了七个类的地面真实像素级标签:城市、农业、牧场、森林、水、贫瘠和未知。它包含 1146 张带注释的卫星图像,大小均为  $2448 \times 2448$  像素。与以前的土地覆盖分类数据集相比,DeepGlobe 具有明显更高的分辨率和更具挑战性。

ISIC 数据集收集了大量的皮肤镜图像。它的子集,ISIC 病变边界分割数据集,由 2594 张来自用于皮肤癌筛查的患者样本的图像组成。所有的图像都用地面真实二进制掩模标注,指示原发性皮肤病变的位置。超过 64% 的图像具有超高分辨率:最大的图像有  $6682 \times 4401$  像素。

Inria Aerial 数据集涵盖了不同的城市景观,从密集的大都市区到高山度假胜地。它提供了 180 张  $5000 \times 5000$  像素的图像,每张图像都标注有建筑、非建筑区域的二进制掩码。与 DeepGlobe 不同的是,它按城市划分训练集、测试集,而不是随机的。

**Table 1.** Classic dataset

**表 1.** 经典数据集

数据集	最大尺寸	平均尺寸	4K UHR 图像的占比	图像个数
DeepGlobe	6M pixels ( $2448 \times 2448$ )	(统一尺寸)	100%	803
ISIC	30M pixels ( $6748 \times 4499$ )	9M pixels	64.1%	2594
Inria Aerial	25M pixels ( $5000 \times 5000$ )	(统一尺寸)	100%	180
Cityscapes	2M pixels ( $2048 \times 1024$ )	(统一尺寸)	0	25000
CamVid	0.7M pixels ( $960 \times 720$ )	(统一尺寸)	0	101
COCO-Stuff	0.4M pixels ( $640 \times 640$ )	0.3M pixels	0	123287
VOC2012	0.25M pixels ( $500 \times 500$ )	0.2M pixels	0	2913

#### 5. 研究展望

近年来,图像语义分割发展迅速,相关算法也层出不穷,但是高分辨率遥感图像语义分割依旧面临着诸多挑战。

##### 1) 整合领域知识

作为数据驱动的方法,深度学习模型在标记的训练数据不足时往往表现不佳,这需要将可以整合到学习过程中的先验领域知识,对训练构成约束或指导训练;从而得到可靠的派生模型。然而,遥感领域的领域知识存在笨拙的合并策略(例如,后处理)。

##### 2) 对多种模式的利用

为了使深度学习模型在解释遥感图像方面取得进展,它们需要能够解释和推理多模态信息。目前,丰富的遥感传感器和观测技术(如主动和被动)产生了来自不同模式的数据。然而,研究中要么未考虑多模



态数据，或者融合策略未能完全挖掘跨模式的互补特征。

### 3) 遥感图像类别的不平衡

在高分辨率遥感图像分类任务中，很难弄清楚哪些类别是最重要的；有些类别可能只占所有类别的一小部分，但它们仍然发挥着重要的作用。类别代表性不足的准确性对评价模型的整体性能很重要。扭曲的数据问题是一个普遍存在的挑战，并且在从传统的机器学习到深度学习的所有学习范式中都无处不在。

## 基金项目

国家自然科学基金(No. 62072024)；北京建筑大学北京未来城市设计高精尖创新中心资助项目(UDC2017033322, UDC2019033324)；北京建筑大学市属高校基本科研业务费专项资金资助(No. X20084, ZF17061)；北京建筑大学研究生创新项目(PG2022144)；北京高等教育本科教学改革创新项目(201910016004)。

## 参考文献

- [1] Maggiori, E., Tarabalka, Y., Charpiat, G. and Alliez, P. (2017) Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, **55**, 645-657. <https://doi.org/10.1109/TGRS.2016.2612821>
- [2] Cheng, G., Zhou, P. and Han, J. (2016) Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 7405-7415. <https://doi.org/10.1109/TGRS.2016.2601622>
- [3] Zhu, H., Jiao, L., Ma, W., Liu, F. and Zhao, W. (2019) A Novel Neural Network for Remote Sensing Image Matching. *IEEE Transactions on Neural Networks and Learning Systems*, **30**, 2853-2865. <https://doi.org/10.1109/TNNLS.2018.2888757>
- [4] Zhu, H., Ma, W., Li, L., Jiao, L., Yang, S. and Hou, B. (2020) A Dual-Branch Attention Fusion Deep Network for Multiresolution Remote-Sensing Image Classification. *Information Fusion*, **58**, 116-131. <https://doi.org/10.1016/j.inffus.2019.12.013>
- [5] Maboudi, M., Amini, J., Malihi, S. and Hahn, M. (2018) Integrating Fuzzy Object Based Image Analysis and Ant Colony Optimization for Road Extraction from Remotely Sensed Images. *ISPRS Journal of Photogrammetry and Remote Sensing*, **138**, 151-163. <https://doi.org/10.1016/j.isprsjprs.2017.11.014>
- [6] Zhang, Q. and Seto, K.C. (2011) Mapping Urbanization Dynamics at Regional and Global Scales Using Multi-Temporal DMSP/OLS Nighttime Light Data. *Remote Sensing of Environment*, **115**, 2320-2329. <https://doi.org/10.1016/j.rse.2011.04.032>
- [7] Marcos, D., Volpi, M., Kellenberger, B. and Tuia, D. (2018) Land Cover Mapping at Very High Resolution with Rotation Equivariant CNNs: Towards Small Yet Accurate Models. *ISPRS Journal of Photogrammetry and Remote Sensing*, **145**, 96-107. <https://doi.org/10.1016/j.isprsjprs.2018.01.021>
- [8] Li, A., Jiao, L., Zhu, H., Li, L. and Liu, F. (2022) Multitask Semantic Boundary Awareness Network for Remote Sensing Image Segmentation. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1-14. <https://doi.org/10.1109/TGRS.2021.3050885>
- [9] Maxwell, S.K., Schmidt, G.L. and Storey, J.C. (2007) A Multi-Scale Segmentation Approach to Filling Gaps in Landsat ETM+ SLC-Off Images. *International Journal of Remote Sensing*, **28**, 5339-5356. <https://doi.org/10.1080/01431160601034902>
- [10] Ton, J., Sticklen, J. and Jain, A.K. (1991) Knowledge-Based Segmentation of Landsat Images. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 222-232. <https://doi.org/10.1109/36.73663>
- [11] Liu, D., Han, L., Ning, X. and Zhu, Y. (2018) A Segmentation Method for High Spatial Resolution Remote Sensing Images Based on the Fusion of Multifeatures. *IEEE Geoscience and Remote Sensing Letters*, **15**, 1274-1278. <https://doi.org/10.1109/LGRS.2018.2829807>
- [12] Lu, L., Wang, C. and Yin, X. (2019) Incorporating Texture into SLIC Super-Pixels Method for High Spatial Resolution Remote Sensing Image Segmentation. 2019 8th International Conference on Agro-Geoinformatics, Istanbul, 16-19 July 2019, 1-5. <https://doi.org/10.1109/Agro-Geoinformatics.2019.8820692>
- [13] Yang, P., Hou, Z., Liu, X. and Shi, Z. (2016) Texture Feature Extraction of Mountain Economic Forest Using High Spatial Resolution Remote Sensing Images. *IEEE International Geoscience and Remote Sensing Symposium*, Beijing,

- 10-15 July 2016, 3156-3159. <https://doi.org/10.1109/IGARSS.2016.7729816>
- [14] Fu, Y., *et al.* (2017) An Improved Combination of Spectral and Spatial Features for Vegetation Classification in Hyperspectral Images. *Remote Sensing*, **9**, Article No. 261. <https://doi.org/10.3390/rs9030261>
- [15] Tatsumi, K., Yamashiki, Y., Canales Torres, M.A. and Taïpe, C.L.R. (2015) Crop Classification of Upland Fields Using Random Forest of Time-Series Landsat 7 ETM+ Data. *Computers and Electronics in Agriculture*, **115**, 171-179. <https://doi.org/10.1109/TGRS.2007.907109>
- [16] Zhong, P. and Wang, R. (2007) A Multiple Conditional Random Fields Ensemble Model for Urban Area Detection in Remote Sensing Optical Images. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 3978-3988. <https://doi.org/10.1109/TGRS.2007.907109>
- [17] Adede, C., Oboko, R., Wagacha, P.W. and Atzberger, C. (2019) A Mixed Model Approach to Vegetation Condition Prediction Using Artificial Neural Networks (ANN): Case of Kenya's Operational Drought Monitoring. *Remote Sensing*, **11**, Article No. 1099. <https://doi.org/10.3390/rs11091099>
- [18] Zhang, C., *et al.* (2018) A Hybrid MLP-CNN Classifier for Very Fine Resolution Remotely Sensed Image Classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, **140**, 133-144. <https://doi.org/10.1016/j.isprsjprs.2017.07.014>
- [19] Wang, L., Li, R., Duan, C., Zhang, C., Meng, X. and Fang, S. (2021) A Novel Transformer Based Semantic Segmentation Scheme for Fine-Resolution Remote Sensing Images. ArXiv: 2104.12137. <http://arxiv.org/abs/2104.12137>
- [20] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition*. Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [21] Badrinarayanan, V., Kendall, A. and Cipolla, R. (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 2481-2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
- [22] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W. and Frangi, A., Eds., *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*, Springer, Cham, 234-241. [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
- [23] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2014) Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. ArXiv: 1412.7062.
- [24] Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2018) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [25] Yu, F. and Koltun, V. (2015) Multi-Scale Context Aggregation by Dilated Convolutions. ArXiv: 1511.07122.
- [26] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H. (2018). Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, Cham, 801-818. <https://doi.org/10.1109/CVPR.2016.396>
- [27] Paszke, A., Chaurasia, A., Kim, S. and Culurciello, E. (2016) Enet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation. ArXiv: 1606.02147.
- [28] Zhao, H., Qi, X., Shen, X., Shi, J. and Jia, J. (2018) ICNet for Real-Time Semantic Segmentation on High-Resolution Images. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, Cham, 418-434. [https://doi.org/10.1007/978-3-030-01219-9\\_25](https://doi.org/10.1007/978-3-030-01219-9_25)
- [29] Chen, L.C., Yang, Y., Wang, J., Xu, W. and Yuille, A.L. (2016) Attention to Scale: Scale-Aware Semantic Image Segmentation. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, 27-30 June 2016, 3640-3649. <https://doi.org/10.1109/CVPR.2016.396>
- [30] Hou, L., Vicente, T.F.Y., Hoai, M. and Samaras, D. (2021) Large Scale Shadow Annotation and Detection Using Lazy Annotation and Stacked CNNs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 1337-1351. <https://doi.org/10.1109/TPAMI.2019.2948011>
- [31] Kirillov, A., Girshick, R., He, K. and Dollár, P. (2019) Panoptic Feature Pyramid Networks. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, 15-20 June 2019, 6392-6401. <https://doi.org/10.1109/CVPR.2019.00656>
- [32] Lin, G., Milan, A., Shen, C. and Reid, I. (2017) RefineNet: Multi-Path refinement Networks for High-Resolution Semantic Segmentation. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, 21-26 July 2017, 5168-5177. <https://doi.org/10.1109/CVPR.2017.549>
- [33] Zhao, H., Shi, J., Qi, X., Wang, X. and Jia, J. (2017) Pyramid Scene Parsing Network. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6230-6239.

- <https://doi.org/10.1109/CVPR.2017.660>
- [34] Wang, J., Sun, K., Cheng, T., Jiang, B., Deng, C., Zhao, Y., Liu, D., Mu, Y., Tan, M., Wang, X., Liu, W. and Xiao, B. (2020) Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 3349-3364. <https://doi.org/10.1109/TPAMI.2020.2983686>
- [35] Zhao, W. and Du, S. (2016) Learning Multiscale and Deep Representations for Classifying Remotely Sensed Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, **113**, 155-165. <https://doi.org/10.1016/j.isprsjprs.2016.01.004>
- [36] Cheng, D., Meng, G., Xiang, S. and Pan, C. (2017) FusionNet: Edge Aware Deep Convolutional Networks for Semantic Segmentation of Remote Sensing Harbor Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10**, 5769-5783. <https://doi.org/10.1109/JSTARS.2017.2747599>
- [37] Marmanis, D., Schindler, K., Wegner, J.D., Galliani, S., Datcu, M. and Stilla, U. (2018) Classification with an Edge: Improving Semantic Image Segmentation with Boundary Detection. *ISPRS Journal of Photogrammetry and Remote Sensing*, **135**, 158-172. <https://doi.org/10.1016/j.isprsjprs.2017.11.009>
- [38] Chen, J., Zhu, J., Sun, G., Li, J. and Deng, M. (2021) SMAF-Net: Sharing Multiscale Adversarial Feature for High-Resolution Remote Sensing Imagery Semantic Segmentation. *IEEE Geoscience and Remote Sensing Letters*, **18**, 1921-1925. <https://doi.org/10.1109/LGRS.2020.3011151>
- [39] Ma, B. and Chang, C.-Y. (2022) Semantic Segmentation of High-Resolution Remote Sensing Images Using Multiscale Skip Connection Network. *IEEE Sensors Journal*, **22**, 3745-3755. <https://doi.org/10.1109/JSEN.2021.3139629>
- [40] Xia, F., Wang, P., Chen, L.-C. and Yuille, A.L. (2016) Zoom Better to See Clearer: Human and Object Parsing with Hierarchical Auto-Zoom Net. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016*, Springer, Cham, 648-663. [https://doi.org/10.1007/978-3-319-46454-1\\_39](https://doi.org/10.1007/978-3-319-46454-1_39)
- [41] Takahama, S., Kurose, Y., Mukuta, Y., Abe, H., Fukayama, M., Yoshizawa, A., Kitagawa, M. and Harada, T. (2019) Multi-Stage Pathological Image Classification Using Semantic Segmentation. 2019 *IEEE/CVF International Conference on Computer Vision*, Seoul, 27 October-2 November 2019, 10701-10710. <https://doi.org/10.1109/ICCV.2019.01080>
- [42] Liu, Y., Fan, B., Wang, L., Bai, J., Xiang, S. and Pan, C. (2018) Semantic Labeling in Very High Resolution Images via a Self-Cascaded Convolutional Neural Network. *ISPRS Journal of Photogrammetry and Remote Sensing*, **145**, 78-95. <https://doi.org/10.1016/j.isprsjprs.2017.12.007>
- [43] Liu, W., Rabinovich, A. and Berg, A.C. (2015) ParseNet: Looking Wider to See Better. ArXiv: 1506.04579.
- [44] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G. and Sang, N. (2018) BiseNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Computer Vision—ECCV 2018*, Springer, Cham, 334-349. [https://doi.org/10.1007/978-3-030-01261-8\\_20](https://doi.org/10.1007/978-3-030-01261-8_20)
- [45] Tokunaga, H., Teramoto, Y., Yoshizawa, A. and Bise, R. (2019) Adaptive Weighting Multi-Field-of-View CNN for Semantic Segmentation in Pathology. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 12589-12598. <https://doi.org/10.1109/CVPR.2019.01288>
- [46] Chen, W., Jiang, Z., Wang, Z., Cui, K. and Qian, X. (2019) Collaborative Global-Local Networks for Memory-Efficient Segmentation of Ultra-High Resolution Images. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 15-20 June 2019, 8916-8925. <https://doi.org/10.1109/CVPR.2019.00913>
- [47] Li, Q., Yang, W., Liu, W., Yu, Y. and He, S. (2021) From Contexts to Locality: Ultra-High Resolution Image Segmentation via Locality-Aware Contextual Correlation. 2021 *IEEE/CVF International Conference on Computer Vision*, Montreal, 10-17 October 2021, 7232-7241. <https://doi.org/10.1109/ICCV48922.2021.00716>
- [48] Bai, H., Cheng, J., Huang, X., Liu, S. and Deng, C. (2022) HCANet: A Hierarchical Context Aggregation Network for Semantic Segmentation of High-Resolution Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, **19**, 1-5. <https://doi.org/10.1109/LGRS.2021.3063799>
- [49] He, K., Sun, J. and Tang, X. (2010) Guided Image Filtering. In: Daniilidis, K., Maragos, P. and Paragios, N., Eds., *Computer Vision—ECCV 2010*, Springer, Berlin, 1-14.
- [50] Wu, H., Zheng, S., Zhang, J. and Huang, K. (2018) Fast End-to-End Trainable Guided Filter. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 1838-1847. <https://doi.org/10.1109/CVPR.2018.00197>
- [51] Li, K., Hariharan, B. and Malik, J. (2016) Iterative Instance Segmentation. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 3659-3667. <https://doi.org/10.1109/CVPR.2016.398>
- [52] Cheng, H.K., Chung, J., Tai, Y.-W. and Tang, C.-K. (2020) CascadePSP: Toward Class-Agnostic and Very Highresolution Segmentation via Global and Local Refinement. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 8887-8896. <https://doi.org/10.1109/CVPR42600.2020.00891>
- [53] Kirillov, A., Wu, Y., He, K. and Girshick, R. (2020) PointRend: Image Segmentation as Rendering. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 13-19 June 2020, 9796-9805.

- 
- <https://doi.org/10.1109/CVPR42600.2020.00982>
- [54] Huynh, C., Tran, A.T., Luu, K. and Hoai, M. (2021) Progressive Semantic Segmentation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, 20-25 June 2021, 16750-16759. <https://doi.org/10.1109/CVPR46437.2021.01648>
- [55] Zi, W., Xiong, W., Chen, H., Li, J. and Jing, N. (2021) SGA-Net: Self-Constructing Graph Attention Neural Network for Semantic Segmentation of Remote Sensing Images. *Remote Sensing*, **13**, Article No. 4201. <https://doi.org/10.3390/rs13214201>
- [56] Lv, L., Guo, Y., Bao, T., Fu, C., Huo, H. and Fang, T. (2021) MFALNet: A Multiscale Feature Aggregation Lightweight Network for Semantic Segmentation of High-Resolution Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters*, **18**, 2172-2176. <https://doi.org/10.1109/LGRS.2020.3012705>