文章编号:1671-4598(2019)07-0231-05 DOI:10.16526/j.cnki.11—4762/tp.2019.07.050 中图分类号:TP242.6 文献标识码:A

深度学习语义分割方法在遥感影像分割中的性能分析

王俊强[1,2], 李建胜[1], 丁 波[2], 蔡 富[1]
(1. 信息工程大学, 郑州 450000; 2. 中国人民解放军78123部队, 成都 610000)

摘要: 针对如何应用深度学习语义分割方法实现遥感影像高性能分割的问题, 选择了当前流行的SegNet、PSPnet以及Deeplabv3+三种基于深度学习语义分割算法, 利用南方某区域无人机高分辨率遥感影像中4类要素分割为实验, 以总体精度、平均精度及平均交并比(MIoU)作为精度衡量指标, 全面对比分析了3种算法的精度; 结果表明, 在迁移学习支持下, 3种算法总体精度可提升2至5个百分点; 通过对PSPNet算法运用不同骨干网络, 验证了不同结构网络对精度的贡献, 优选出复杂度低的骨干网络; 采用集成学习的思路, 利用投票法对多算法模型进行结果融合可提升总体精度1%左右; 3种算法对植被及水体的分割效果均要优于建筑物及道路, 其中Deeplabv3+算法精度最高, 总体精度达到89.3%, MIoU达到80.4%, 可实现要素的鲁棒分割。

关键词: 遥感影像; 深度学习; 语义分割; 总体精度; 迁移学习

Performance Analysis of Semantic Segmentation Method Based on Deep Learning in Remote Sensing Image Segmentation

Wang Junqiang[1,2], Li Jiansheng[1], Ding Bo[2], Cai Fu[1]
(1. Information Engineering University, Zhengzhou 450000, China; 2. Unit 78123 Troops, Chengdu 610000, China)

Abstract: Towards how to use deep learning semantic segmentation method to realize high performance segmentation of remote sensing image, three popular semantics segmentation algorithms based on deep learning are selected, which are SegNet, PSPnet and Deeplabv3+. Based on the classification of four types of elements in high-resolution remote sensing images of a UAV in the south, this paper takes the overall accuracy, average accuracy and MIoU as the accuracy measurement indicators, and comprehensively analyzes the performance of the three algorithms. The experimental results show that the overall accuracy of the three algorithms can be improved by 2 to 5 percentage points with the support of migration learning. By replacing the PSPNet algorithm backbone network, the contribution of different structural networks to accuracy is verified, and a backbone network with low complexity is optimized. Using the idea of integrated learning, the result fusion of multi-algorithm model based on voting method can improve the overall accuracy by about 1%. The three algorithms have better segmentation effects on vegetation and water body than buildings and roads. Among them, Deeplabv3+ algorithm has the highest accuracy, the overall accuracy reaches 89.3%, and MIoU reaches 80.4%, which can achieve the robust segmentation of the elements.

Keywords: remote sensing images; deep learning; semantic segmentation; overrall accuracy; transfer learning

0 引言

遥感影像分割作为遥感影像解译的重要分支之一, 是将图像分割为若干对象区域, 每个区域内的像素之间具有较好的相似性, 同时保证对象区域之间有较大的异质性[1]。智能的遥感影像分割可实现典型要素自动提取, 如快速提取道路网数据, 能够为导航图提供数据支持。传统分割方式一般使用随机森林[2]或者纹理基元森林方法[3]来构建用于语义分割的分类器, 这类方法需要人工制作复杂特征, 鲁棒性差, 难以满足大范围自动化作业需求。近年来, 深度学习在多种高级计算机视觉任务中取得成功, 特别是监督学习下的卷积神经网络在图像分类、目标检测方面成功鼓舞着研究人员探索此类网络对于像素级标记, 如语义分割方面的能力。2014年Jonathan Long等提出的全卷积神经网络(Fully Convolutional Network, FCN)[4], 是深度学习应用于图像语义分割的开山之作, 将传统卷积神经网络(Convolutional Neural Network, CNN)中的全连接层转化成卷积层, 编码部分通过卷积和池化操作获取特征图, 解码部分通过反卷积上采样恢复原图尺度, 实现像素级分割。然而, 相比于CNN下采样阶段的结构规整, FCN上采样时的结构相对凌乱。因此, 2015年Vijay Badrinarayanan提出SegNet算法[5], 采用了几乎和下采样对称的上采样结构, 分割精度及效率均得到提升。针对现有模型由于没有引入足够的上下文信息及不同感受野下的全局信息而存在分割出现错误的情景, Zhao H提出了使用全局场景下的类别信息的PSPNet算法[6], 另外还提出了引入辅助损失的深度残差网络(ResNet)[7]优化方法。Deeplab系列(v1, v2, v3, v3+)是由Liang-Chieh Chen等[8-10]提出的, 通过不断优化, 最近的Deeplabv3+[10]在引用多孔空间金字塔池化

收稿日期:2019-01-22; **修回日期**:2019-02-18。

作者简介:王俊强(1990-),男,江西万安人,硕士研究生,主要从事计算机视觉、导航制导与控制方向的研究。

(ASPP) 网络模块，利用解码编码的形式，扩展了感受野，获取更多的上下文信息，能够实现图像鲁棒分割。总而言之，这些基于深度学习的语义分割算法不断优化，性能得到提升，但这些算法都基于公开自然场景数据集上评价分析，当前针对于高分辨率遥感影像分割分析较少，因此，研究分析这些算法在遥感影像中的分割性能，对于选择合适算法用于遥感图像语义分割具有参考价值。

本文将基于无人机遥感影像，通过定性对比试验和定量评价分析典型的 3 种语义分割算法 SegNet、PSPNet、Deeplabv3+的分割性能。本文首先介绍了这 3 种语义分割算法基本原理，其次阐述了优化策略，最后通过实验进行全面分析评价。

1 算法原理

1.1 SegNet 算法

SegNet 类似于全卷积网络的解码编码形式，但编码和解码使用的技术不一致，其网络结构如图 1 所示。该网络结构是一种对称结构，编码部分使用的是 VGGNet 网络[11]的前 13 层卷积网络，通过卷积提取高维特征，并通过池化使图片变小，解码部分通过上采样与反卷积操作使特征图像变大，恢复至原输入图像大小，最后通 Softmax 层，输出每个像素点不同分类的最大值。由于最大池化和子采样的叠加，会导致边界细节损失增大，因此 SegNet 在编码特征图过程中储存了最大池化标记位置，并且在上采样过程恢复最大池化位置。

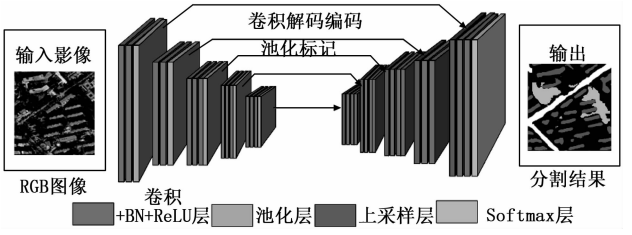

图 1 SegNet 算法原理

1.2 PSPNet 算法

针对大部分全卷积网络的模型都缺少合适的策略去利用全局场景下的类别信息，PSPNet 算法引入金字塔池化模块，其原理示意图如图 2 所示。输入图像通过特征提取网络得到原来图像尺寸 1/8 的特征图像，特征图像送入金字塔池化模块，金字塔池化分为 4 种不同尺度，池化之后可得到不同尺寸的特征图，对每个金字塔层级特征图进行 1 * 1 卷积降维操作，然后直接对低维的特征图进行上采样，得到原图尺寸。最后，不同层的特征图与原特征图融合连接后经过卷积输出结果。

1.3 Deeplabv3+算法

Deeplabv3+算法采用类似于 FCN 的编码器－解码器的方式，其原理示意图如图 3 所示。

输入图像利用特征提取网络生成比原图缩小 16 倍特征图。该算法以 Xception 网络作为骨架网络，其网络结构由一系列深度可分离卷积、类似 ResNet 中的残差连接和一些

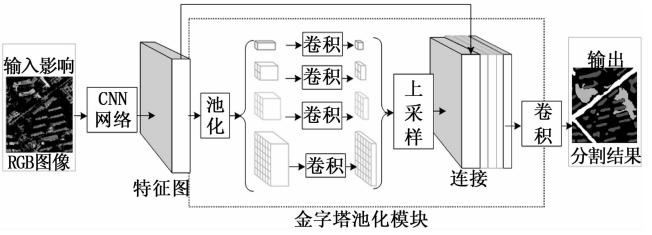

图 2 PSPNet 算法原理

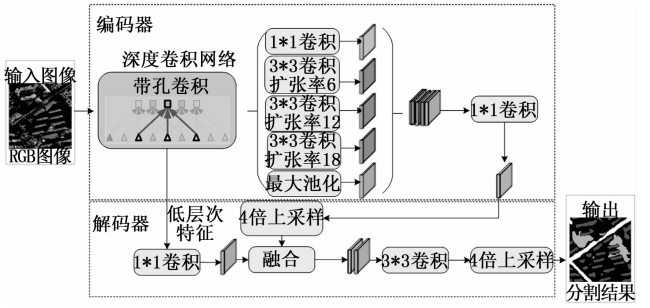

图 3 Deeplabv3+原理示意

其他常规的操作组成，由于 Deeplabv2 版本采用的是 ResNet 作为骨架网络，Deeplabv3+精度较 Deeplabv2 提升，因此本文不再对该算法下以 ResNet 作为骨架网络进行训练。Xception 网络中引入了 ASPP 模块，可以在多尺度上捕获信息，实现鲁棒分割。然后将特征图输入到一个 256 通道的 1 * 1 卷积层中。最后，将卷积后的特征图输入至解码器部分实现恢复至原图像大小的分割结果。该解码器借鉴全卷积网络的跳步连接方式，首先利用 48 通道 1 * 1 卷积对低层次特征图卷积，实现特征图降维，再将其与经 4 倍双线性内插上采样的高层次特征图融合，最后进行 3 * 3 卷积操作后经 4 倍的双线性内插恢复至原图大小，获得分割预测图。

2 优化策略

2.1 迁移学习

受限于硬件环境能力，实际训练中可能不具备分布式 GPU 环境，并且训练样本规模有限，如果随机给定初始化模型参数权值，训练效果不一定良好。

迁移学习是将算法中的骨架网络在经过海量图像分类数据预训练好的权值，迁移至训练任务中，对网络权值进行初始化。通过迁移学习可加速网络训练速度，提高训练精度[12]。但语义分割算法中的骨干网络和图像分类任务预训练的模型不一定完全一致，如 PSPNet 算法中的 ResNet 网络引入空洞卷积策略，而实际预训练 ResNet 模型不具备该参数，导致两者参数无法一一对应，因此，实际只从预训练模型中加载模型中包含的相关参数字段。

2.2 CNN 网络替换

骨干网络是语义分割算法的基础网络，对于提取特征至关重要，网络结构越合理，模型算法效果将更优，也更容易进行训练。为寻求更优的骨干网络，可对算法骨干网络部分进行替换实验，例如，针对 PSPNet 算法，文献 [6]

中特征提取网络采用的是带有空洞卷积的 ResNet 网络，为对比不同骨架网络下的算法性能，本文同时设计采用 DenseNet 网络[13] 作为特征提取网络进行对比实验。DenseNet 是一种具有密集连接的卷积神经网络，其基本思路与 ResNet 一致，但相对于 ResNet 的“短路连接”，其建立的是前面所有层与后面层的密集连接，可实现特征重用，提升效率。在图像分类领域里，同等精度下，DenseNet 的参数量要小于 ResNet。

在 PSPNet 算法基础上，本文利用 ResNet 和 DenseNet 不同网络层数的网络结构作为骨干网络设计模型进行实验，分别为 ResNet－34，ResNet－50、ResNet－101、DenseNet121、DenseNet169 以及 DenseNet201，算法生成模型大小如表 1 所示，其中以 DenseNet121 为骨架网络的 PSPNet 模型参数最少，模型复杂度最低，训练时需要梯度下降的参数更少。本文将在节 3.3 对不同骨干网络进行精度分析。

表 1　模型参数大小

算法	骨架网络	参数量(M)	骨架网络	参数量(M)
PSPNet	ResNet－34	110.4	DenseNet121	76.8
	ResNet－50	214.7	DenseNet169	140.3
	ResNet－101	290.9	DenseNet201	183.1

3　实验分析

3.1　实验平台与数据

实验硬件为联想 P920 工作站，操作系统为 Ubuntu16.04，配置 96G 内存及 NVIDIA TITAN Xp 显卡。编程语言为 Python，深度学习框架为 Pytorch。

实验训练数据为南方某区域 2015 年 5 张不同像素大小及区域的无人机遥感影像及相应标记数据，标记分为 5 类，分别为背景、植被、建筑、水体及道路。将其中 4 张裁切为 300 * 300 大小图片 1900 张，另外 1 张裁切为 300 * 300 大小图片 300 张作为验证集。另有多张不同场景的测试影像数据（不带标记）作为分割效果可视化对比验证。由于训练样本数据量有限，仅利用上述样本训练，会造成过拟合现象，为解决这个问题，本文设计了适应遥感数据特点的数据增强器，该数据增强器相比于常规的通过图像单个变换方式增强，不同之处在于其以概率的形式对多种图像变换方式进行组合操作，再将数据固定到指定尺寸，如图 4 所示。

在该图像增强器的处理下，每次输入网络中的图片能从色彩、亮度、纹理、尺度等方面保持差别。因此，通过数据增强器能够有效丰富样本的数量。

3.2　精度评价指标

传统影像分类方法采用总体精度（overrall accuracy，OA）、平均精度（average accuracy，AA）和 Kappa 系数作为评价指标[15]。本文将继续采用 OA 和 AA 指标，并引入深度学习标准度量 MIoU 作为评价指标。假设图像共有 $k+1$ 个待分割签类别（从 L_0 到 L_k，其中 L_0 为背景类），p_{ii} 表

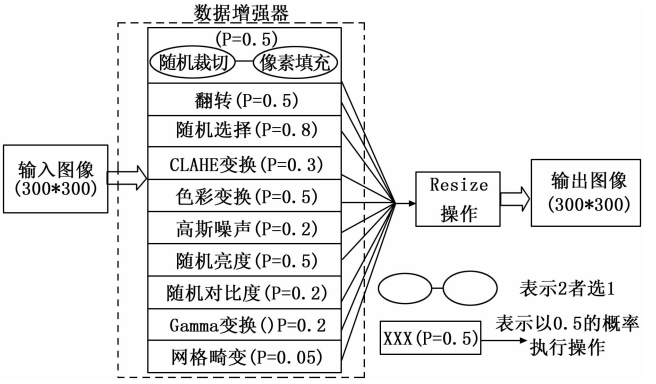

图 4　数据增强器设计

示本属于类 i 但被预测为类 j 的像素数量。即 p_{ii} 表示真正例的数量，而 p_{ij}、p_{ji} 则分别为假正例和假负例。则总体精度可表示为：

$$OA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}} \tag{1}$$

平均精度是总体精度的一种简单提升，计算每个类别被正确分类像素数量的比例后，计算所有类别的平均值，可表示为：

$$AA = \frac{1}{k+1}\sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{2}$$

MIoU 是计算真实值和预测值两个集合的交集和并集之比，在每个类别计算 IoU 后取平均值，可表示为：

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

该指标综合反映了目标的捕获程度（使预测标签与标注尽可能重合）和模型的精确程度使并集尽可能重合）情况。

3.3　迁移学习支持下的性能分析

以 PSPNet 和 SegNet 算法为例，骨干网络分别采用 ResNet－101 和 VGG16 网络，对迁移训练和未迁移训练的总体精度随着训练 epoch（训练集中的全部样本训练一次为 1 个 epoch）变化情况进行对比分析，如图 5 所示，两种算法均训练 150 个 epoch，批处理尺寸为 8，初始学习率为 0.001，每 50 个 epoch 下降为原来 0.1 倍。

从图 5 可知，在 SegNet 及 PSPNet 算法中迁移训练方式结果均要优于随机初始化权值方式，总体精度大概能提升 2～5 个百分点。收敛速度方面，迁移训练方式快于初始化权值方式，尤其 PSPNet 算法表现更明显，PSPNet 学习率为 0.001 的情况下，迁移训练方式训练前 20 个 epoch 精度上升较快，在第 50 个 epoch 降低学习率后，第 60 个 epoch 达到收敛，初始化权值方式明显经过训练更多 epoch 后收敛。因此，利用迁移学习方式是一种有效提升训练效率和精度的方式，尤其是在数据量小或深度学习设备性能有限情况下，迁移学习方式作用更突出。

3.4　不同骨干网络下的性能分析

在迁移学习的基础上，对表 1 设计的不同骨干网络模

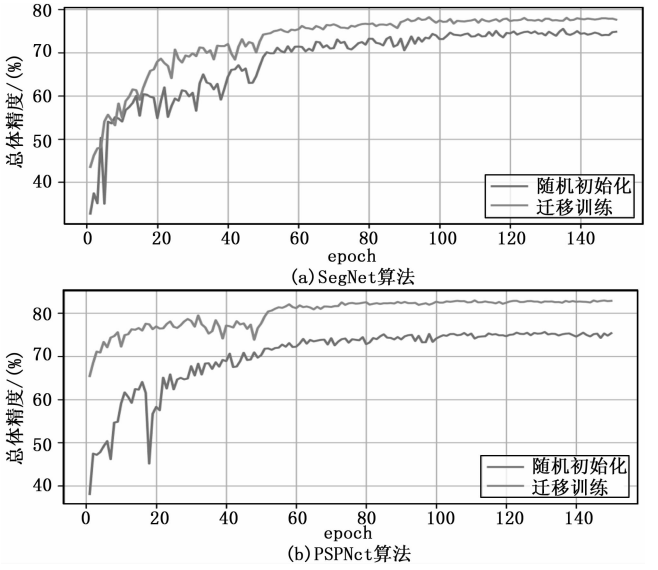

图 5 迁移学习支持下的总体精度对比

型进行训练，训练参数设置同上，训练总体精度随着 epoch 变化情况如图 6 所示。结合表 1 及图 6 可知，以 DenseNet121 为基础的 PSPNet 算法参数量最少，约为以 ResNet－34 为基础的 0.7 倍，但最终精度略高于 ResNet－34。ResNet 和 DenseNet 相互对应的几种级别网络中，DenseNet 精度总体与 ResNet 相当，表明 DenseNet 这种密集连接型的卷积神经网络更加优化，能够保证精度的前提下降低模型复杂度，从而降低训练难度。

3.5 验证集精度分析

通过以上实验，验证了迁移学习方式对于训练效果的提升。本文将使用迁移学习对以上 3 种算法进行训练，其中 PSPNet 算法采用 ResNet－50、ResNet－101、DenseNet169 以及 DenseNet201 网络，SegNet 算法采用 VGG 网络，Deeplabv3＋采用 Xception 网络，3 种算法均训练 150 个 epoch，批处理大小为 8，初始学习率为 0.001，每 50 个 epoch 下降为原来 0.1，统计分析总体精度、平均精度以及 MIoU 如表 3 所示。从总体精度和平均精度来看，Deeplabv3＋算法达到最高的总体精度 89.3％及平均精度 88.9％，相对于 SegNet 和 PSPNet 算法提升幅度较大，PSPNet 算法中不同骨架网络精度略有不同，但总体变化幅度不大，以 ResNet－101 为骨架网络精度最高。从 MIoU 来看，Deeplabv3＋算法达到最高 80.4％，较 SegNet 和 PSPNet

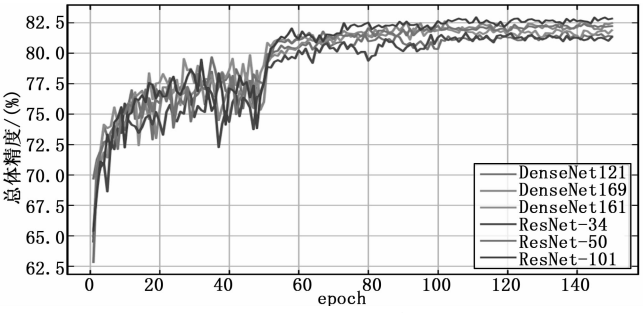

图 6 迁移学习支持下的总体精度对比

算法提升较大。从分类别精度（AA 和 IoU）来看，所有类别 Deeplabv3＋算法均达到最高精度，SegNet 算法所有类别精度均不如 PSPNet 算法，道路及建筑物分割由于复杂程度高于水体及植被，其精度相对于水体及植被更低。为尝试集成学习方法对结果的影响，以 PSPNet 算法与 SegNet 算法为基础，利用简单的多数投票法的方式对各模型结果进行融合[16]，获得最终结果，从表 2 可知，融合后的 MIoU 值相对于 PSPNet 可以提升 1 个百分点。

综合以上可知，Deeplabv3＋算法精度远高于其他两种算法，是深度学习语义分割方法运用于遥感影像高精度分割的不错选择。在各种模型算法精度较差情况下，通过多模型投票融合可提升分割精度。PSPNet 算法精度虽不如 Deeplabv3＋算法，但其优势是较为模块化的结构，便于更换 CNN 网路，降低模型复杂程度。

图 7 分别采用以上 3 种算法及 SegNet 与 PSPNet 投票融合的方式对验证集图片进行分割可视化（黑色为背景，绿色为植被，蓝色为河流，红色为建筑物，白色为道路），从图 7 可知，Deeplabv3＋算法分割效果最优，边界信息较其他方法更完整，SegNet 算法分割图像较粗糙，要素信息完整性不如其他两种算法。通过 SegNet 与 PSPNet 算法融合后，分割效果较单个算法更优。

3.6 测试集上结果对比

测试数据上的分割结果如图 8 所示，输入图像选择 3 张 1500＊1500 像素大小的无人机影像，Deeplabv3＋分割完整性及边缘信息均要优于 PSPNet 算法及 SegNet 算法，尤其是道路的连通性，建筑物的边界效果，说明 Deeplabv3＋算法能够实现目标的鲁棒分割。

4 结论

基于深度学习的语义分割算法基本结构均是采用解码和

表 2 不同算法精度统计情况 (％)

算法	骨架网络	总体精度	类别像素精度					平均	IoU					MIoU
			背景	植被	建筑	水体	道路		背景	植被	建筑	水体	道路	
SegNet	VGG16	76.9	75.1	83.6	72.3	88.0	68.9	77.6	57.7	70.4	60.6	73.1	54.6	63.3
PSPNet	ResNet－50	81.9	81.0	85.4	79.4	91.1	73.3	82.0	64.5	75.5	69.2	77.0	62.6	69.8
	ResNet－101	82.1	82.7	83.2	81.1	90.1	70.4	81.5	65.1	74.3	71.1	77.6	61.1	69.9
	DenseNet169	81.9	82.3	84.0	80.1	88.1	70.2	81.1	64.8	74.5	70.0	77.0	62.5	69.8
	DenseNet201	81.5	84.2	81.1	79.8	88.4	70.8	80.9	64.5	72.5	69.6	79.1	61.9	69.5
Deeplabv3＋	Xception	89.3	87.0	91.2	89.6	93.1	81.2	88.9	75.3	86.6	81.1	86.2	72.8	80.4
SegNet 与 PSPNet 融合		82.8	84.8	84.3	80.7	90.5	69.8	82.0	66.5	75.2	71.5	78.9	63.0	71.0

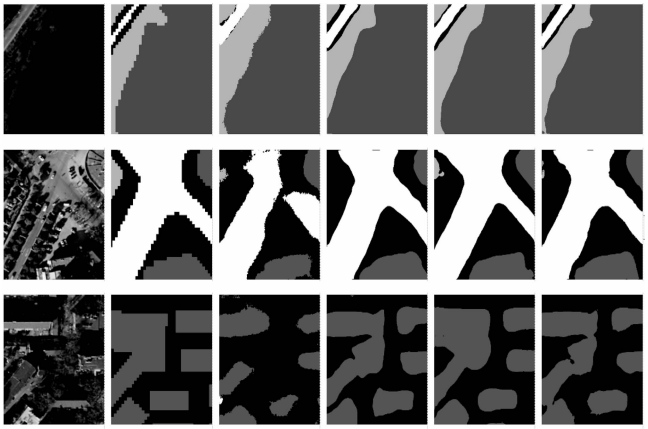

图 7　不同算法在验证集上的预测效果对比

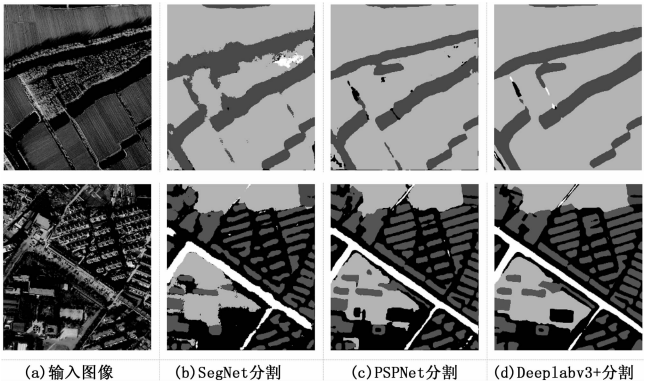

图 8　测试图片分割效果对比

编码形式，但通过在算法结构中引入不同模块，可达到不同分割效果。本文对比分析了代表性的 3 种 3 种深度学习语义分割方法 SegNet、PSPNet、Deeplabv3＋的性能。通过无人机影像数据分割试验分析，可得到以下结论：利用迁移学习方式训练，可提升训练精度和加快训练进度，提升总体精度 2—5 个百分点。编码器部分不同骨架网络可达到不同精度效果，选择一种结构最优且精度较高的骨架网络(如 DenseNet)，可保证精度的前提下，降低模型复杂度。通过投票集成方式对不同模型或者同模型不同尺度下的预测结果进行融合，可提升训练精度。Deeplabv3＋算法较其两种算法精度更优，能够实现目标的鲁棒分割，将其应用于遥感影像高精度解译是较好的选择。

本文分析的几种典型的算法，对于将深度学习方法用于遥感影像分割方面具有一定参考价值，但由于当前深度学习发展迅速，可能会涌现出更多好的算法，如在网络中引入全局上下文信息模块的 EncNet[17]，同样能够达到较好效果，后续可对这些算法作进一步研究。

参考文献：

[1] 刘建华，毛政元. 高空间分辨率遥感影像分割方法研究综述[J]. 遥感信息，2009 (6)：95－101.

[2] 冯文卿，眭海刚，涂继辉，等. 高分辨率遥感影像的随机森林变化检测方法 [J]. 测绘学报，2017 (11)：1880－1890.

[3] 张　蓬，赵书斌，彭思龙. 基于纹理基元的图象分割 [J]. 中国图象图形学报，2003，8 (8)：50－55.

[4] Shelhamer E，Long J，Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2014，PP (99)：1－1.

[5] Badrinarayanan V，Kendall A，Cipolla R. SegNet：A deep convolutional encoder－decoder architecture for scene segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence，2015，PP (99)：1－1.

[6] Zhao H，Shi J，Qi X，et al. Pyramid scene parsing network [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Hawaii：IEEE，2017：2881－2890.

[7] He K，Zhang X，Ren S，et al. Deep residual learning for image recognition [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas：IEEE，2016：770－778.

[8] Chen L C，Papandreou G，Kokkinos I，et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [J]. Computer Science，2014 (4)：357－361.

[9] 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 & Machine Intelligence，2016，PP (99)：834－848.

[10] Chen L C，Zhu Y，Papandreou G，et al. Encoder－decoder with atrous separable convolution for semantic image segmentation [A]. European Conference on Computer Vision [C]. Munich：2018：801－818.

[11] Simonyan K，Zisserman A. Very deep convolutional networks for large－scale image recognition [A]. 3rd International Conference on Learning Representations (ICLR) [C]. Hilton San Diego：Computer Science，2015：1150－1210.

[12] Pan S J，Yang Q. A survey on transfer learning [J]. IEEE Transactions on knowledge and data engineering，2010，22 (10)：1345－1359.

[13] Huang G，Liu Z，Van Der Maaten L，et al. Densely connected convolutional networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii：IEEE，2017：2261－2269.

[14] Chollet F. Xception：Deep learning with depthwise separable convolutions [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Hawaii：IEEE，2017：1251－1258.

[15] 吴　波，林珊珊，周桂军. 面向对象的高分辨率遥感影像分割分类评价指标 [J]. 地球信息科学学报，2013，15 (4)：567－573.

[16] 樊利恒，吕俊伟，邓江生. 基于分类器集成的高光谱遥感图像分类方法 [J]. 光学学报，2014 (9)：91－101.

[17] Zhang H，Dana K，Shi J，et al. Context encoding for semantic segmentation [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Salt Lake City：IEEE，2018：7151－7160.