

doi:10.6046/zrzyyg.2023150

引用格式:苏腾飞.深度卷积语义分割网络在农田遥感影像分类中的对比研究——以河套灌区为例[J].自然资源遥感,2024,36(4):210-217.(Su T F.A comparative study on semantic segmentation-orientated deep convolutional networks for remote sensing image-based farmland classification:A case study of the Hetao irrigation district[J].Remote Sensing for Natural Resources,2024,36(4):210-217.)

深度卷积语义分割网络在农田遥感影像分类中的对比研究

——以河套灌区为例

苏腾飞

(内蒙古农业大学水利与土木建筑工程学院,呼和浩特 010018)

摘要:在现代化农业生产管理中,不同类型作物的空间分布是重要的农情信息。从卫星遥感影像中识别农田种类,是获取该类信息的基本途径之一。虽然目前用户可选择的遥感影像地物识别算法较为丰富,但进行可靠的农田分类依旧具有一定的挑战性。该文选取了3类具有代表性的深度卷积语义分割模型,包括 UNet, ResUNet 和 SegNext, 对比其在河套灌区高分二号遥感影像上的作物分类性能。在3类算法的框架内,共实现了9种具有不同复杂度的模型,以分析各个网络结构在农田遥感影像作物分类中的性能差异,从而为后续的相关模型研究提供一些优化思路与实验基础。实验结果说明,具有6层网络结构的 UNet 取得了最高的总精度(88.74%),而6层 SegNext 的精度最差(84.33%);具有最高模型复杂度的是 ResUNet,但对于研究数据集,这类算法的过拟合现象最为严重;在计算效率方面,ResUNet 也显著低于另外2类算法。

关键词:深度卷积;语义分割;农田分类;河套灌区

中图法分类号: TP 79 **文献标志码:** A **文章编号:** 2097-034X(2024)04-0210-08

0 引言

粮食安全是国家长治久安的重要保障,也是经济发展的基础之一。要确保粮食安全,及时获取农田空间分布信息是非常必要的,因为该信息在作物长势观测、农业水利调配、粮食收成预测等方面,都具有非常重要的实际应用价值。农田地区的遥感影像分类,已成为现代化粮食安全监测的基础手段^[1]。然而,要在遥感影像中精准区分不同种类的作物依旧存在较大挑战,这主要是由于2方面的因素:“同谱异物”现象广泛存在于农田遥感影像中,即不同作物具有相近的遥感光谱特征,导致分类算法的混淆;“同物异谱”现象也经常出现在农田遥感影像中,这主要是由于气候变化、土壤条件、耕作活动存在局部差异,致使同类作物处于不同的生长阶段,从而给分类任务带来困难。

为解决以上难题,一个直接且重要的途径是优

化机器学习分类算法的作物识别性能。目前已出现非常丰富的农田遥感影像分类算法研究,大致可以分为2类:基于传统机器学习的算法和基于深度学习的算法。虽然前者的研究更早,但近年来其相关工作也较为频繁。You 等^[2]利用随机森林(random forest, RF)算法与 Sentinel-1/2 的时序数据,进行了黑龙江主粮作物的识别研究;Hao 等^[3]利用美国农田分布数据对 RF 算法进行训练,用以区分 Landsat8 与 Sentinel-2 时序影像中的农田类别;梁继等^[4]选择了支持向量机(support vector machine, SVM)算法来分析高分六号卫星影像中的2个红边波段对农作物分类精度的效果;Gao 等^[5]提出了一种作物生长曲线对齐方法,并将其与对象级 SVM 算法的核函数进行融合,并基于此开展了 Sentinel-1 时序影像农田分类的实验;马战林等^[6]融合了 Sentinel-1 与 Sentinel-2 的时序数据,并利用超像素的 RF 算法,对河南省的大蒜农田进行了识别研究。

以上研究主要以 RF 或 SVM 为基础开展。尽

收稿日期:2023-05-23;修订日期:2023-09-07

基金项目:国家自然科学基金项目“对象级深度学习的河套灌区农田遥感影像分类算法研究”(编号:62361050)和内蒙古自治区高等学校科学技术研究项目“对象级主动学习的河套灌区遥感作物分类算法研究”(编号:NJZY22495)共同资助。

作者简介:苏腾飞(1987-),男,硕士,实验师,主要从事农业遥感影像分析与深度学习算法研究。Email:stf1987@126.com。

管这类算法具有较好的分类精度与较低的使用难度,其性能依然在很大程度上受输入特征种类的影响,这使得用户需要在算法训练之前开展特征工程(feature engineering, FE)^[1]。FE的效果取决于用户的经验与技术。在很多场景下,用户趋向于通过FE得到数目庞大的特征集合,以尽可能对数据进行完整表达。然而,过多的特征量易导致过拟合现象,反而降低了传统机器学习算法的性能。近年来兴起的深度学习技术,避免了FE的步骤,进而克服了相关问题。

深度学习技术利用多层次的神经网络来构建数据分析模型。网络结构的多样化决定了模型的特征提取与目标预测能力。卷积神经网络(convolutional neural network, CNN)因其优秀的影像数据处理能力,已成为深度学习在遥感影像分析领域的主流模型。Zhong等^[7]将二维CNN与条件随机场结合,区分了高光谱影像中的农田作物类型;Liu等^[8]比较了3种CNN在农田遥感影像分类中的性能,实验结果显示二维CNN既高效又准确;Gallo等^[9]利用三维CNN对意大利伦巴第平原的当季作物进行了分类研究;Li等^[10]收集了河套灌区的Sentinel-2的时序数据,对比了机器学习算法与CNN深度学习模型(包括UNet和DeepLabv3)的农田分类精度;许晴等^[11]讨论了CNN模型在弱训练样本集合情况下的适用性,并在辽宁的玉米和水稻农田分类中取得了较高精度。

以上研究中,遥感影像的农田分类被构建为基于CNN的语义分割问题。目前,UNet是利用CNN进行语义分割的主流结构框架,而该模型的架构对其性能的影响是非常显著的。鉴于此,有必要开展主流UNet架构的对比分析,这对该类算法的优化与应用都具有较好的实际意义。

在UNet^[12]的基础上,ResUNet模型考虑了多尺度残差连接结构^[13],SegNext模型结合考虑了多尺度特征提取与卷积自注意力机制^[14],二者都是近年来CNN研究中里程碑式的算法,在其架构设计中都考虑了深度学习领域的关键技术与思想。因此,本文实现了这3类模型,并利用河套灌区高分辨率遥感影像,以基础UNet为基准,对比各模型的作物区分性能,以期为后续基于深度网络的农田遥感影像分类研究提供实验参考与技术支持。

1 基于深度卷积网络的语义分割

1.1 UNet 语义分割网络架构

UNet最早被用于医学影像的语义分割,它属于

经典的编码器-解码器架构。编码器主要负责特征的提取与抽象,而解码器对所提取的抽象特征进行逐层还原,以提升空间细节的预测精度。图1展示了一个具有4层结构的UNet语义分割模型,图中 \oplus 表示张量的相加。由图可见,随着编码器的层数加深,特征的数目逐层翻倍,但其空间维度也随之缩小。经过浅层编码器模块的特征具有更丰富的空间信息,但其在不同类别的区分度方面相对较低;相反,深层模块的特征具有较高的类别区分能力,但也丢失了较多局部细节。鉴于语义分割是一个注重空间细节还原的任务,UNet将编码器与解码器的输出进行逐层连接(即图1的跳跃连接),以提升语义分割精度。

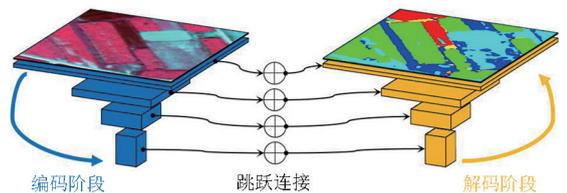


图1 4层UNet模型的基本结构

Fig.1 A backbone of UNet structure with 4 layers

一般而言,UNet的编码器由一系列卷积与池化操作构成,其中前者负责特征的丰富化,后者则对特征进行压缩与抽象化。在本文实现的UNet中,每层编码器先进行前馈卷积、批规范化和ReLU激活函数,其输出再由最大池化操作进行空间维度的压缩。以第 L 层编码器为例,其输入特征的维度是 $(S \times 2^{L-1}, H \times 2^{-L+1}, W \times 2^{-L+1})$,其输出为 $(S \times 2^L, H \times 2^{-L}, W \times 2^{-L})$,其中 S 是特征的丰富化常数,大多相关研究将其设定为32, H 与 W 分别表示原始输入影像的高度和宽度。在第一层编码器之前,本文利用 1×1 卷积操作将原始影像的波段维度拓展为 S 。

UNet的解码器主要包含2个操作:张量的合并与空间维度拓展。前者首先将上一层解码器与对应层编码器的输出进行张量相加,再通过卷积操作进行进一步的特征提取;后者利用双线性插值算法对所提取的特征进行空间维度的拓展,每层解码器的空间拓展系数为2,使拓展后的特征较拓展前具有两倍的高度和宽度。

在农田遥感影像分类实验中,根据编码器与解码器的层数,本文共实现了3种基本的UNet语义分割模型,它们分别是具有4,5,6个编码阶段和解码阶段的UNet模型,下文分别将其简称为UNet_L4, UNet_L5和UNet_L6。

1.2 ResUNet 网络结构

对比基础的UNet,ResUNet主要包括3点不同:

在每一层的编码器与解码器中,基本的卷积操作之后具有一个多尺度残差连接模块,以进行多个尺度的特征提取;在编码器与解码器的最后一层,添加了PSPPooling^[15]来开展多尺度的特征压缩;对于每一层的跳跃连接,使用了合并模块,即先对2个输入张量在特征维度上进行串联,再利用1×1卷积对串联后的特征进行维度压缩,以确保下一层解码阶段的计算。

在以上3个方面中,最重要的是多尺度残差连接模块,因为它是ResUNet最主要的特征提取环节,其结构如图2所示。其中,⊕表示张量的相加;BRC表示一个序列的操作,包含批规范化、ReLU激活层、膨胀参数为*d*的空洞卷积。空洞卷积能够使较大感受野的卷积保持较少的参数量;*d*决定了空洞卷积的感受野,较小的*d*注重局部信息提取,较大的*d*则关注大范围的空间纹理信息。根据层数的不同和是否利用PSPPooling,ResUNet的提出者定义了3种模型:6个编码层与解码层的ResUNet(下文简称为ResUNet_D6);7个编码层与解码层的ResUNet,且在最后一个的编码阶段用普通卷积替代PSPPooling(下文简称为ResUNet_D7v1);在ResUNet_D7v1的基础上,利用PSPPooling对最后一个编码层的输出进行特征提取(下文简称为ResUNet_D7v2)。ResUNet_D7v1与ResUNet_D7v2可以对比出PSPPooling模块的影响。对于不同层的编码阶段和解码阶段,*d*的定义如表1所示。

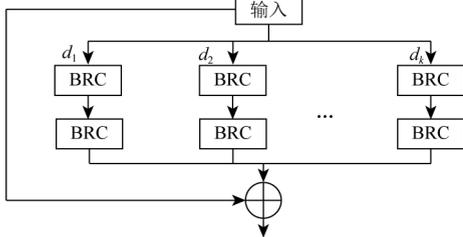


图2 ResUNet中的多尺度残差连接模块

Fig.2 Multi-scale residual connection module in ResUNet

表1 ResUNet中空洞卷积的膨胀参数*d*的定义情况

Tab.1 Definition of the dilation parameter *d* of the atrous convolution in ResUNet

层数	<i>d</i>
第一、二层	{1, 3, 15, 31}
第三、四层	{1, 3, 15}
第五、六、七层	{1}

1.3 SegNext网络结构

SegNext开创性地提出了卷积自注意力模块^[14],用轻量化的多尺度卷积操作来替代Vision Transformer中的自注意力机制,以提升图像语义分割的性能。本文实现的SegNext与基本UNet的结

构相似,唯一不同的是在每个编码层的基本卷积之后,利用卷积自注意力模块进行特征提取,该模块的结构由图3展示。图3中⊕和⊗分别表示张量的相加与内积,C(1×*d*)表示卷积核大小等于1×*d*的深度可分离卷积。多分支的1×*d*与*d*×1卷积操作用于进行多尺度特征提取;这种垂直与水平方向的条形卷积核,可在保持较少参数量的前提下得到感受野相当于*d*×*d*的效果;此外,条形卷积核有助于线状地物的特征提取,例如道路边缘、农田田垄、分支渠道等。深度可分离卷积的使用,进一步降低了参数量,从而减少过拟合风险。

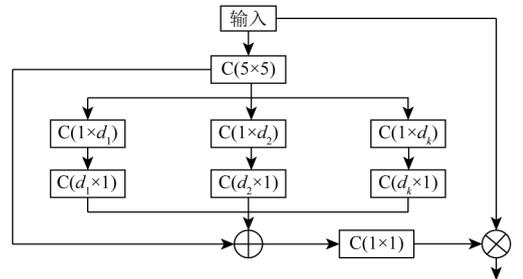


图3 SegNext中的多尺度卷积自注意力模块

Fig.3 Multi-scale convolutional self-attention module in SegNext

本文共构建了3种SegNext网络,分别是具有4,5,6层编码器与解码器的模型,下文分别将其简称为:SegNext_L4, SegNext_L5, SegNext_L6。对于SegNext中的多分支结构,*d*参数的设定遵循了原始SegNext中的定义:{7, 11, 21}。

1.4 损失函数与模型训练

要训练类似UNet的深度CNN模型,目前主要采用的损失函数包括Cross-Entropy函数和Dice系数函数2种。前者在场景分类中的应用更为广泛,而后者在与语义分割相近的任务中得到了越来越多的关注。Diakogiannis等^[13]在ResUNet的训练中提出了一种互补的Dice损失函数,可加快语义分割训练的收敛速度。鉴于此,本文采用该方法来构建损失函数,其基本形式为:

$$T_{Dice}(p, t) = 1 - \frac{1}{C} \times \frac{\sum_k n_k^{-2} \times \sum_i p_{i,k} \times t_{i,k}}{\sum_k n_k^{-2} \times [\sum_i (p_{i,k}^2 + t_{i,k}^2) - \sum_i p_{i,k} \times t_{i,k}]}, \quad (1)$$

$$T_{Dice}^{\sim}(p, t) = [T_{Dice}(p, t) + T_{Dice}(1 - p, 1 - t)] / 2, \quad (2)$$

式中:*p*和*t*分别为预测概率值向量与实际地物真值向量;*p*_{*i,k*}和*t*_{*i,k*}分别为第*i*个像素对于第*k*个类别的预测概率与真值标签,前者可以在[0, 1]范围内连续取值,而后者的取值是0或1;*C*为类别个数;*N*

为输入影像的像素总数; n_k 为第 k 个类别的样本数目,其在式(1)中的作用是降低不同类别样本数目差异对训练的影响。

每个模型的训练过程都包含了 100 个 Epoch,且数据的 batch size 参数设置为 4。本文采用了 Adam 算法进行模型参数优化,其 β_1, β_2 与权重衰减率参数分别设置为 0.9, 0.999, $1E-3$ 。本文采用了 multi-step learning rate 的训练策略,使 learning rate 参数在前 50 与后 50 个 Epoch 的大小分别等于 $1E-4$ 和 $1E-5$,以微调模型在后半阶段的训练效果。本文所采用的模型都是在 PyTorch 框架下实现的,实验平台配置为: Intel i7 12700H CPU @ 2.3 GHz, RTX3060 移动版 GPU (6 GB 显存), 16 GB 内存, Windows11 操作系统。所有模型输入张量的空间维

度是 256×256 。

2 数据集与实验设置

2.1 数据集

本文采用的高分辨率遥感影像由 WorldView-2 获取,空间分辨率为 0.5 m,包含近红外、红光、绿光、蓝光 4 个波段。影像成像时间为 2017 年 8 月 5 日,中心经纬度为 $N41.010^\circ, E108.155^\circ$,尺寸为 $4\ 608$ 像素 \times $3\ 072$ 像素。

本文的研究区域位于内蒙古巴彦淖尔市的五原县,该县涵盖了河套灌区中部腹地,且盛产玉米、葵花、小麦等经济作物。图 4 为遥感影像数据与样本数据。

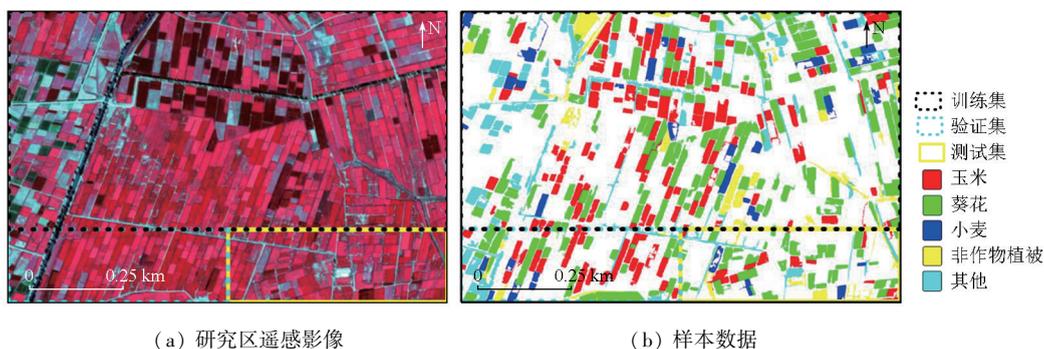


图 4 研究区域与遥感影像数据集

Fig.4 Study area and the remote sensing image dataset

本文的样本数据包括 5 个地物类别:玉米、葵花、小麦、非作物植被、其他。本文的样本收集由 3 个步骤完成:

1) 实地调查。通过实地踏勘,在研究区获取少量典型地物的位置信息(经纬度)。

2) 目视解译。根据经纬度信息,将步骤 1) 得到的样本在遥感影像中进行匹配,为目视解译提供典型样例。

3) 样本扩充。在步骤 2) 的基础上,在全景影像中进行样本标注,这一步首先利用了多尺度分割技术^[16]来自动化地获取农田地块,使斑块成为样本选择的操作单元,该做法可显著提升农田地块级样本的标注效率。

要完成算法的精度对比,需要将所有数据划分为训练集、验证集、测试集,图 4 显示了 3 个数据集的空间范围。验证集、测试集分别用于算法训练的超参数调试与不同算法的精度对比,其中精度的计算标准采用的是总精度(overall accuracy, OA)和 Kappa 系数(Kappa coefficient, KC)。

2.2 实验设置

为了充分对比 3 类深度卷积网络在农田地区语

义分割上的性能,本文设计了 3 方面的实验:算法精度对比分析、训练与过拟合风险分析以用运算效率分析。实验 1 主要对比各个算法在精度上的差异,包括 OA 和 KC,以及描述特定类别区分效果的生产者精度(producer accuracy, PA)与用户精度(user accuracy, UA);实验 2 主要考察不同算法在训练过程中的表现,通过对比训练集与测试集上的精度变化,来分析过拟合现象的程度;实验 3 对不同网络的运算效率进行了对比,所采用的标准是算法参数量与每秒十亿级浮点操作量(giga-floating point operations, GFLOPs),2 个标准的规律均为:值越大,表示模型复杂度越高,对应其计算效率越低。

3 实验结果

3.1 算法精度对比

图 5 展示了各个算法在测试集上的精度变化。根据图 5(a)和(b),精度最高的模型为 UNet_6L,其 OA 和 KC 分别是 88.74% 和 0.836 0;精度最差的是 SegNext_L6,OA 和 KC 分别为 84.33% 和 0.769 3。对于 UNet 与 ResUNet,模型复杂度的增加均使其精度

提升;然而,对于 SegNext,虽然其 5 层的精度优于 4 层,但是 6 层的精度却出现了明显下降。此外,对于相同层数的 UNet 与 SegNext,前者的精度均好于后者。这主要是因为 SegNext 在其卷积自注意力模块

中采用的是线状卷积核,而在农田地区里,线状地物较少,而主要地物(农田地块)大多为矩形块状,从而导致其编码器在特征提取能力方面的不适性,进而影响精度。

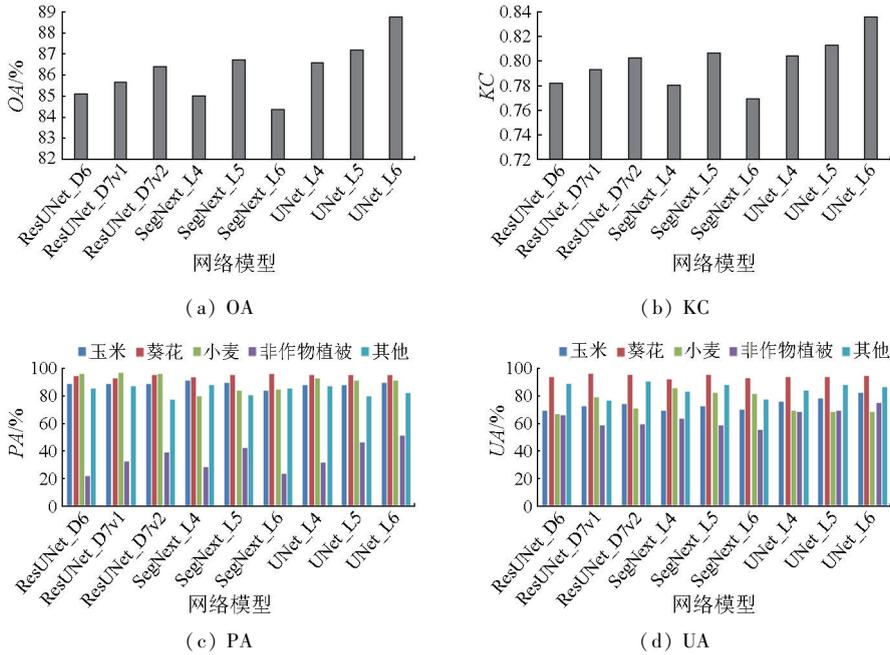


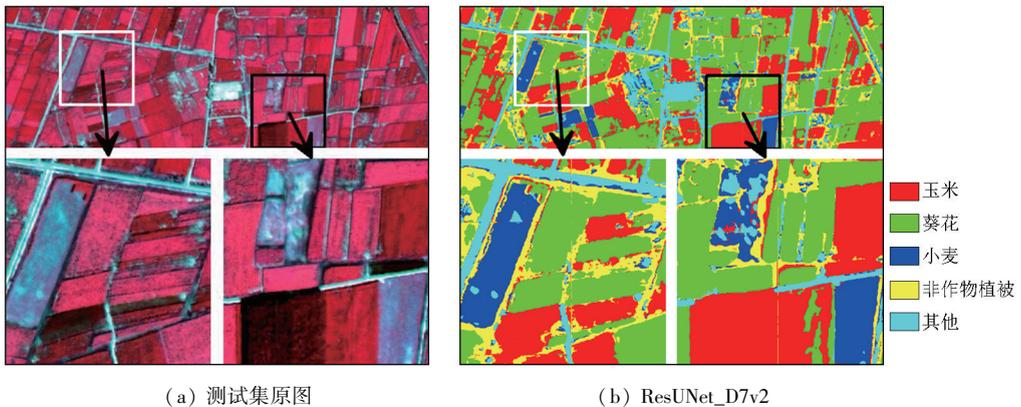
图 5 各个深度卷积语义分割网络算法的精度对比

Fig.5 Accuracy comparison for different deep convolutional semantic segmentation networks

图 5(c) 和 (d) 展现了各个类别的精度变化。较为明显的是,对于非作物植被,所有算法在 PA 与 UA 方面都得到了较低的分值。UNet_L6 在非作物植被的 PA 与 UA 表现最佳,其值分别是 52.60%与 76.70%;最差的非作物植被 PA 由 ResUNet_D6 得到,为 22.73%,而该类别最低的 UA 是 SegNext_L6 的 56.77%。以上结果说明,9 种算法的主要错误均来源于非作物植被,换言之,该类地物均在不同程度上被各个算法错分为其他类别的地物。值得一提的是,对于所有算法,葵花的 UA 都明显高于其他类别,特别是玉米和小麦。根据以上结果可以推论:大多非农田植被主要被错分为玉米或小麦。该结论与

对影像的目视是相符的,因为大部分非作物植被包括田间林地、路旁树木,以及荒地杂草,前两者在光谱上与玉米相近,而后者的光谱与小麦地块类似。

3 类深度卷积语义分割网络中对测试集的最佳分类结果见图 6,即 ResUNet_D7v2, SegNext_L5 和 UNet_L6 的最优分类结果。为了方便对结果开展分析,选择了 2 处具有代表性的子区域,分别对应图 4 中的白色与黑色方框。通过对白色方框区域的观察,可发现 3 种算法对玉米、葵花、小麦的区分效果较好;对于该区域西侧的小麦地块,只有 ResUNet_D7v2 能够完整地将其提取出来,另外 2 种算法受田内光谱变化的影响,将部分小麦田错分为其他类别。



(a) 测试集原图 (b) ResUNet_D7v2

图 6-1 3 类算法对测试集的最优分类结果

Fig.6-1 The optimal classification results for the 3 types of algorithms

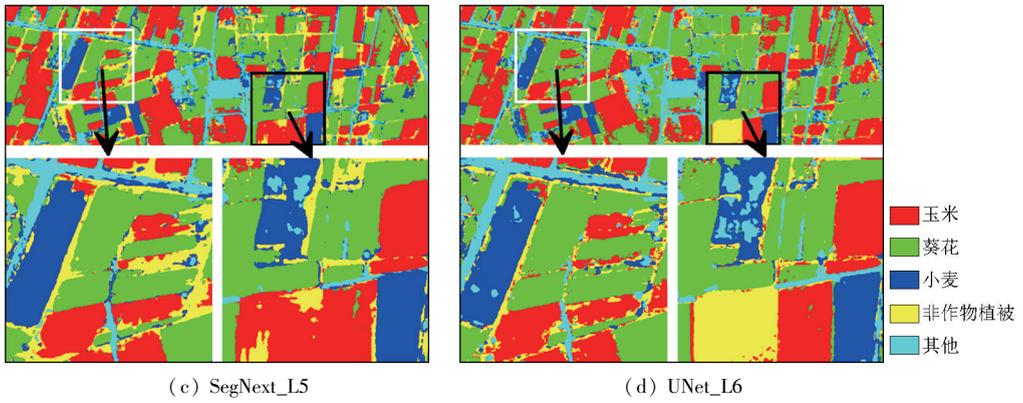


图 6-2 3 类算法对测试集的最优分类结果
 Fig.6-2 The optimal classification results for the 3 types of algorithms

在黑色方框区域西南部,存在一片较大的林地,只有 UNet_6L 将其正确识别,另 2 种算法都将其划分为玉米。以上分析说明,本文所采用的 3 类深度卷积语义分割网络,在农田遥感影像的地物区分中能够取得较好的效果,但对一些光谱相似的地物发生了

错分现象。

3.2 训练与过拟合风险分析

为了分析 9 种算法在训练过程中的性能变化,计算了所有模型对于训练集和测试集的 OA,其结果由图 7 展示。

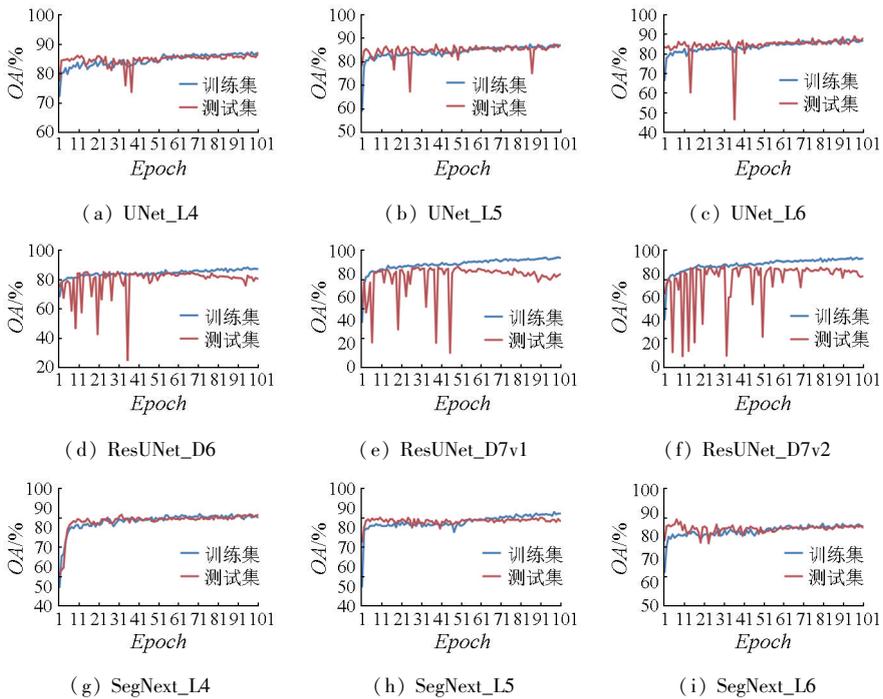


图 7 9 种深度卷积语义分割网络在训练过程中的总精度变化

Fig.7 The variation of overall accuracy for the 9 deep convolutional semantic segmentation networks during training

对于训练集的 OA,所有模型均表现出了相似的规律:随着 Epoch 增加,OA 先迅速上升(当 Epoch < 10 时),此后 OA 缓慢提高或趋于平稳。对于测试集的 OA,3 类算法较为不同,尤其是 ResUNet,其 3 种模型都在 Epoch < 50 时出现了非常大的波动;而随着 Epoch > 70,其 OA 甚至出现了小幅下降,但此时其训练集的 OA 依旧处于平稳上升的态势,可见 ResUNet 对本文数据集出现了明显的过拟合现象。

虽然 SegNext 与 UNet 的训练集和测试集的 OA 较 ResUNet 更为一致,但 3 种 SegNext 模型的测试集

OA 的波动更小。尽管 SegNext_6L 在训练初始几轮的 Epoch 出现了明显抖动,但在后半段的训练过程中,其训练集与测试集 OA 的相符程度很高。对于 3 种 UNet,其测试集 OA 都出现了一定程度的波动,但都显著少于 ResUNet。此外,当 Epoch > 70 时,相较于另外 2 类模型,UNet 的测试集 OA 的上升趋势更为明显,特别是 UNet_6L,其测试集 OA 在 Epoch > 90 时是所有算法中最高的。

以上分析说明,对于本文采用的农田地区高分辨率遥感影像的分类问题,ResUNet 的过拟合风险

最高,虽然 SegNext 的表现最为平稳,但其最优精度低于 UNet。

3.3 运算效率分析

深度学习网络模型的运算效率主要由 2 个方面体现:参数量与计算量,其中后者的主流量化标准是 GFLOPs。GFLOPs 可用于反映不同 GPU 的算力性能,也可用于衡量同一 GPU 在处理不同网络训练训练或预测时的计算量,计算量越多对应了越高的参数量。参数量反映了模型的复杂程度:一般而言,越复杂的模型对应越高的参数量。

图 8 展示了 9 种语义分割模型的运算效率对比情况,可见 3 类算法都在模型复杂度增加的情况下,

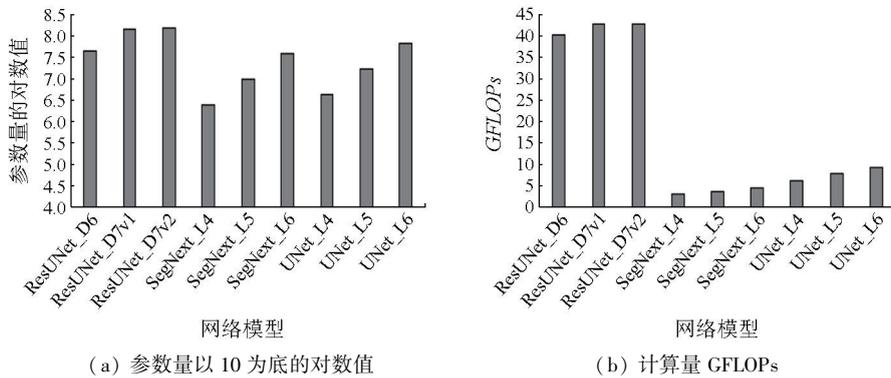


图 8 9 种深度卷积语义分割网络的参数量与计算量对比

Fig.8 Comparison of 9 deep convolutional semantic segmentation networks for the parameter size and computational load

4 结论

本文主要对比了 3 类深度卷积语义分割网络在农田地区遥感影像中的应用效果。通过在预测精度、训练过程与计算效率 3 个方面的实验分析,得出结论如下。

1) 虽然 ResUNet 具有最高的模型复杂度,但在本文的数据集中出现了明显的过拟合现象。

2) SegNext 在训练过程中得到的模型具有平稳的性能表现,但其最优精度低于 UNet。

3) UNet_L6 取得了最高的测试集精度,说明对于本文所采用的农田遥感影像的分类,其适用度相对最佳。

在未来的工作中,有必要在其他场景(例如遥感影像的建筑提取或变化监测)对本文所采用的 9 种模型进行对比研究。此外,在本文实验结论的基础上,可以发展其他网络模型,以进一步优化农田地区遥感影像分类的精度。

参考文献 (References):

[1] Kattenborn T, Leitloff J, Schiefer F, et al. Review on convolutional neural networks (CNN) in vegetation remote sensing [J]. ISPRS

出现了效率降低(对应参数量与计算量的增加)的现象。SegNext_L4 的计算效率是最高的,而计算效率最低的模型是 ResUNet_D7v2。对于在本文实验中取得最佳测试集精度的 UNet_L6,其模型复杂度仅低于 3 种 ResUNet,计算效率也是除 ResUNet 之外最低的,但是其计算效率与 SegNext 和另外 2 种 UNet 相差不大。另外,在层数相同的前提下,UNet 在参数量上略微高于 SegNext(见图 8(a));一般而言,虽然过多的参数量会导致更高的过拟合风险,但也能产生更强大的特征提取能力,这可能也是同层数下 UNet 的精度高于 SegNext 的原因。

Journal of Photogrammetry and Remote Sensing, 2021, 173: 24-49.

[2] You N, Dong J. Examining earliest identifiable timing of crops using all available Sentinel 1/2 imagery and Google Earth Engine [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 161: 109-123.

[3] Hao P, Di L, Zhang C, et al. Transfer learning for crop classification with cropland data layer data (CDL) as training samples [J]. Science of the Total Environment, 2020, 733: 138869.

[4] 梁继, 郑镇炜, 夏婷婷, 等. 高分六号红边特征的农作物识别与评估 [J]. 遥感学报, 2020, 24(10): 1168-1179.

Liang J, Zheng Z W, Xia S T, et al. Crop recognition and evaluation using red edge features of GF-6 satellite [J]. Journal of Remote Sensing, 2020, 24(10): 1168-1179.

[5] Gao H, Wang C, Wang G, et al. A novel crop classification method based on ppSVM classifier with time-series alignment kernel from dual-polarization SAR datasets [J]. Remote Sensing of Environment, 2021, 264: 112628.

[6] 马战林, 薛华柱, 刘昌华, 等. 基于主被动遥感数据和面向对象的大蒜识别 [J]. 农业工程学报, 2022, 38(2): 210-222.

Ma Z L, Xue H Z, Liu C H, et al. Identification of garlic based on active and passive remote sensing data and object-oriented technology [J]. Transactions of the Chinese Society of Agricultural Engineering, 2022, 38(2): 210-222.

[7] Zhong Y, Hu X, Luo C, et al. WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural

- network with CRF[J]. Remote Sensing of Environment, 2020, 250: 112012.
- [8] Liu S, Zhou Z, Ding H, et al. Crop mapping using sentinel full-year dual-polarized SAR data and a CPU-optimized convolutional neural network with two sampling strategies[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021, 14: 7017-7031.
- [9] Gallo I, Ranghetti L, Landro N, et al. In-season and dynamic crop mapping using 3D convolution neural networks and sentinel-2 time series[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2023, 195: 335-352.
- [10] Li G, Cui J, Han W, et al. Crop type mapping using time-series Sentinel-2 imagery and U-Net in early growth periods in the Hetao irrigation district in China[J]. Computers and Electronics in Agriculture, 2022, 203: 107478.
- [11] 许晴, 张锦水, 张风, 等. 深度学习农作物分类的弱样本适用性[J]. 遥感学报, 2022, 26(7): 1395-1409.
- Xu Q, Zhang J S, Zhang F, et al. Applicability of weak samples to deep learning crop classification[J]. National Remote Sensing Bulletin, 2022, 26(7): 1395-1409.
- [12] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[J/OL]. arXiv, 2015. <https://arxiv.org/abs/1505.04597.pdf>.
- [13] Diakogiannis F I, Waldner F, Caccetta P, et al. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2020, 162: 94-114.
- [14] Guo M H, Lu C Z, Hou Q, et al. SegNext: Rethinking convolutional attention design for semantic segmentation[J/OL]. arXiv: 2022(2022-9-18)[2023-8-26]. <https://arxiv.org/abs/2209.08575>.
- [15] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual U-net[J/OL]. arXiv, 2017. <https://arxiv.org/abs/1711.10684.pdf>.
- [16] Baatz M, Schäpe A. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation[C]// Angewandte Geographische Informations Verarbeitung XII, Wichmann Verlag, 2000: 12-23.

A comparative study on semantic segmentation-orientated deep convolutional networks for remote sensing image-based farmland classification: A case study of the Hetao irrigation district

SU Tengfei

(College of Water Conservancy and Civil Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China)

Abstract: In the management of modern agriculture production, the spatial distribution of different crop types is identified as important information about agricultural conditions. Identifying crop types from satellite remote sensing imagery serves as a fundamental method for acquiring such information. Although there exist various algorithms for identifying surface features from remote sensing imagery, reliable farmland classification remains challenging. This study selected three representative semantic segmentation-orientated deep convolutional models, i. e., UNet, ResUNet, and SegNext, and compared their performance in crop classification using remote sensing images of the Hetao irrigation district from the Gaofen-2 satellite. Using the three algorithms, nine network models with varying complexities were developed to analyze the differences in the performance of various network structures in classifying crops in farmland based on remote sensing imagery, thus providing optimization insights and an experimental basis for future research on relevant models. Experimental results indicate that the six-layer UNet achieved the highest identification accuracy (88.74%), while the six-layer SegNext yielded the lowest accuracy (84.33%). The ResUNet displayed the highest complexity but serious over-fitting with the dataset used in this study. Regarding computational efficiency, ResUNet was significantly less efficient than the other two model types.

Keywords: deep convolution; semantic segmentation; crop field classification; Hetao irrigation district

(责任编辑: 张 仙)