Advances in Applied Mathematics 应用数学进展, 2022, 11(6), 3753-3765
Published Online June 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.116403

Hans汉斯

基于迁移学习的大规模遥感图像语义分割与提取

袁　月[1]，汪鹏飞[2]，胡志淘[2]

[1]上海理工大学理学院，上海
[2]南京大学电子科学与工程学院，江苏　南京

收稿日期：2022年5月21日；录用日期：2022年6月11日；发布日期：2022年6月24日

摘　要

图像语义分割是计算机视觉领域的热点研究课题，随着全卷积神经网络的迅速兴起，图像语义分割和全卷积神经网络的融合发展取得了快速发展。本文建立了基于特征融合的大规模卷积神经网络语义分割训练模型。通过迁移学习的监督式训练方式，对目前图像分割领域主流的模型进行了训练与比较，建立了评估指标PR参数以及预测图像的噪声分析模型。在训练模型时将其分成两个分支，利用主流的语义分割模型特性，分别做降噪分支和提取空间语义分支等，引入采集源高度作为权重参数，对不同分支进行特征融合，以提高其鲁棒性。最后进行特征融合，借助机器学习完成遥感图像分割任务并对本文模型的有效性进行验证。

关键词

遥感图像语义分割，特征融合，迁移学习，卷积神经网络

Semantic Segmentation and Extraction of Large Scale Remote Sensing Images Based on Transfer Learning

Yue Yuan[1], Pengfei Wang[2], Zhitao Hu[2]

[1]College of Science, University of Shanghai for Science and Technology, Shanghai
[2]School of Electronic Science and Engineering, Nanjing University, Nanjing Jiangsu

Received: May 21st, 2022; accepted: Jun. 11th, 2022; published: Jun. 24th, 2022

文章引用: 袁月, 汪鹏飞, 胡志淘. 基于迁移学习的大规模遥感图像语义分割与提取[J]. 应用数学进展, 2022, 11(6): 3753-3765. DOI: 10.12677/aam.2022.116403

Abstract

Image semantic segmentation is a hot research topic in the field of computer vision. With the rapid rise of fully convolutional neural networks, the fusion of image semantic segmentation and fully convolutional neural networks has achieved rapid development. In this paper, a large-scale convolutional neural network semantic segmentation training model based on feature fusion is established. Through the supervised training method of transfer learning, the current mainstream models in the field of image segmentation are trained and compared, and the evaluation index PR parameters and the noise analysis model of the predicted image are established. When training the model, it is divided into two branches, using the characteristics of mainstream semantic segmentation models to do noise reduction branch and extraction spatial semantic branch, etc., and introducing the height of the acquisition source as a weight parameter, and perform feature fusion on different branches to improve its robustness. Finally, feature fusion is performed, the remote sensing image segmentation task is completed with the help of machine learning and the effectiveness of the model in this paper is verified.

Keywords

Remote Sensing Image Semantic Segmentation, Feature Fusion, Transfer Learning, Convolutional Neural Network

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

遥感图像目标识别技术是遥感图像应用的基础，识别技术决定了遥感图像应用的效果，这就对遥感图像的目标识别技术提出了更高的要求，在应用遥感图像的过程中，准确、高效地进行目标识别，可以节约大量人力物力，具有很高的经济价值。目标自动识别一直是计算机视觉与模式识别研究领域的重要研究方向，也是当前模式识别和图像处理领域的一个研究热点[1]，通过对遥感影像中感兴趣的类别进行提取和分类，即利用图像分割提取出图像中的房屋，水域，农田等用地类型，这在城乡规划、防汛救灾等领域具有很高的实用价值[2]。识别技术直接决定了遥感图像应用的效果，这就对遥感图像的目标识别技术提出了更高的要求。目前高精度的耕地信息提取主要还是依靠人工解译，耗费大量人力、财力且效率较低，因此，遥感图像的耕地识别算法研究将对耕地遥感制图提供重要帮助[3]。

2. 研究方法

2.1. 数据处理

本文整理了国内外公开的遥感图像数据集，包括 UCMerced，LandUse，WHU-RS19 等。为了满足内存需求和提升训练效率，本文使用滑动窗口对图像数据和标签数据进行了裁剪，统一裁剪为 256*256。为了最大程度地减少其余变量在不同图像中造成的差异。

2.2. 图像增强

由于可用的已标注的高分遥感影像数据较少，难以用于训练一个深度神经网络模型，因此本文在训

练过程中，通过对图像随机旋转，镜像变换，调整 HSV 通道[4]等图像增强操作，丰富数据集，实现数据增广，解决训练过程中数据集匮乏，解决训练过拟合的情况。对现有的数据集，通过分割将 1 个图片样本成若干个样本，再通过图像增强操作，极大地丰富数据集，提高模型鲁棒性，既能使模型在训练时可以学习到图像的特征，又不会图像过大导致硬件环境资源不足。

2.3. 后处理—膨胀与腐蚀

形态学变换膨胀采用向量加法(或 Minkowski 集合加法) [5]，膨胀 $X \oplus B$ 代表所有向量加之和的集合，向量加法的两个操作数分别来自于 X 和 B，并且取到任意可能的组合：

$$X \Theta B = \left\{ p \in \varepsilon^2, p + b \in X, \forall b \in B \right\} \tag{1}$$

腐蚀与膨胀的关系可以描述为：

$$\left(X \Theta B \right)^C = X^C \oplus \tilde{B} \tag{2}$$

其中 $\tilde{B}$ 关于参考点的对称集合，也称转置，可以根据此式，利用膨胀运算来实现腐蚀运算。

2.4. 初步模型

本文对数据集制作相应的标签，并利用图像增强操作对制作的数据集进行数据增广。建立基于卷积神经网络的训练模型，主要包括 FCN，UNET 系列，GCN 模型，结合语义分割相关的评价指标，以及预测结果在连通性和噪声两个方面进行对比分析，根据评估结果进行模型的融合与改进，使得模型的训练效果达到最好。

本文主要使用以深度学习网络为基础的模型，首先对数据集进行监督式的训练处理，再通过迁移学习进行优化调参。采用的图像分割模型主要包括骨干网络，特征训练的编码解码部分，以及最后的分类层。每个网络部分的基础单元包括卷积层，池化层等基本单元。本文提出的语义分割模型及其基础模型以深度神经学习模型，按图 1 所示的深度学习模型的训练流程进行训练[6]。

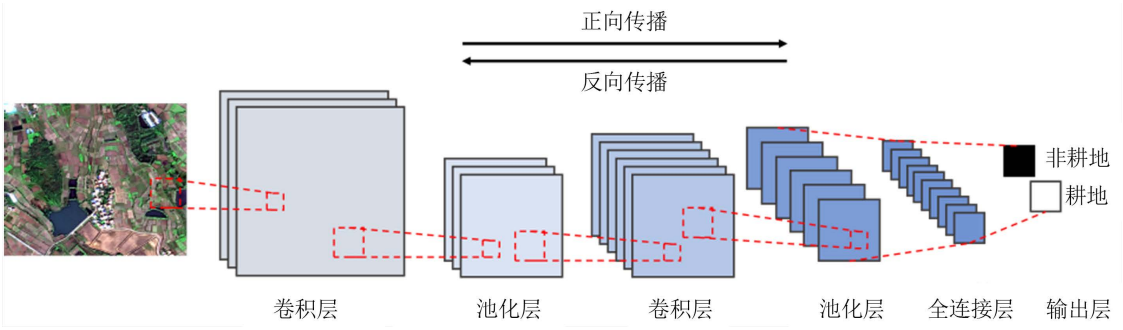


Figure 1. Training process of convolutional neural network
图 1. 卷积神经网络训练过程

由于缺乏相应的数据集，本文采用的迁移学习的训练方式，即利用搜集的数据集优化搭建的模型或者已有公开主流模型进行迁移学习。使用预先训练的模型作为基础，并重新训练优化最后几层的参数，获得一个良好的模型。

经典 CNN 模型在多个卷积层和池化层之后会包含若干个全连接层，全卷积神经网络(Fully Convolution Networks) [7] [8]将全连接层转换为卷积层，通过上采样操作输出与输入图像同等尺寸的特征图并进行预测，最终得到图像的像素级分类结果。具体实现结构如图 2 所示[9]：

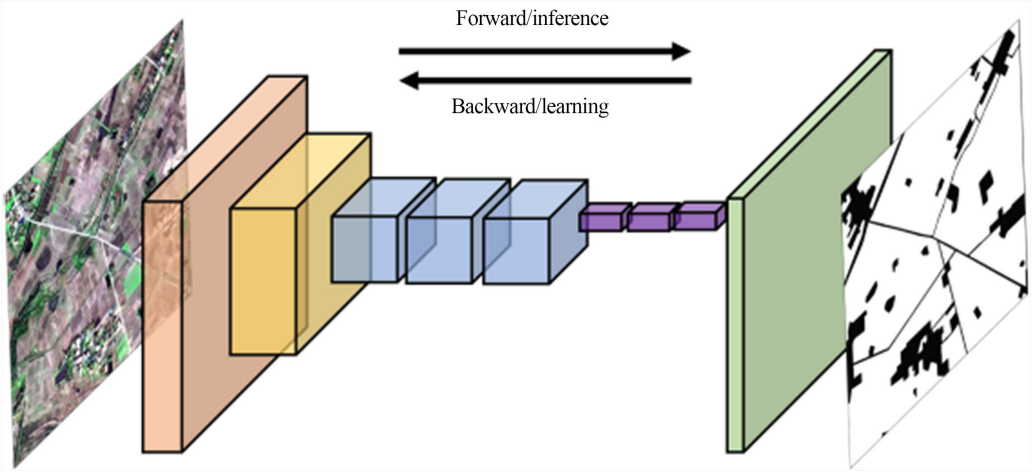


Figure 2. Schematic diagram of FCN model
图 2. FCN 模型示意图

PSP 网络模型(Pyramid Scene Parsing Network)针对 FCN 网络存在分割精度不够精细，没有考虑上下文信息的问题进行了改进，增加了 PSP 模块获取多个 channel 的特征，结构如图 3 所示：

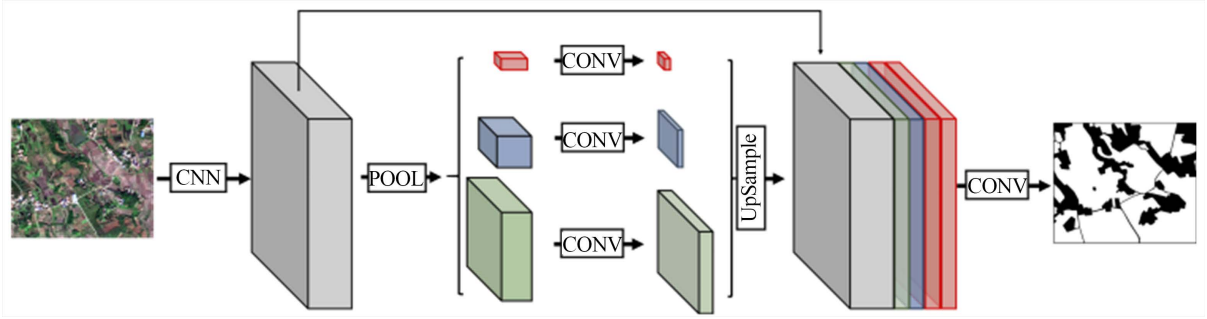


Figure 3. Schematic diagram of the PSP model
图 3. PSP 模型示意图

UNET 是在 FCN 的基础上改进的网络结构，结构比较简单，Encoder 部分做特征提取，Decoder 部分上采样，结构如图 4 所示[10]。

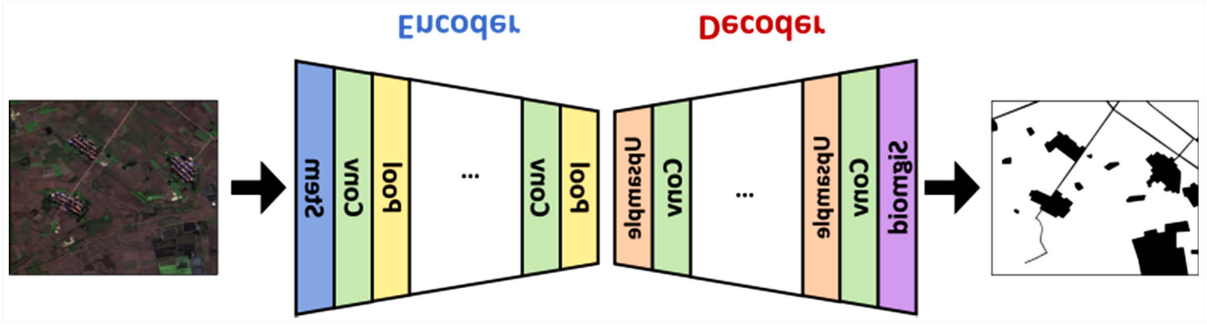


Figure 4. Schematic diagram of the UNET model
图 4. UNET 模型示意图

UNET++相比 UNET 使用了多个编码器来为输入图像生成强特征，如图 5 所示：

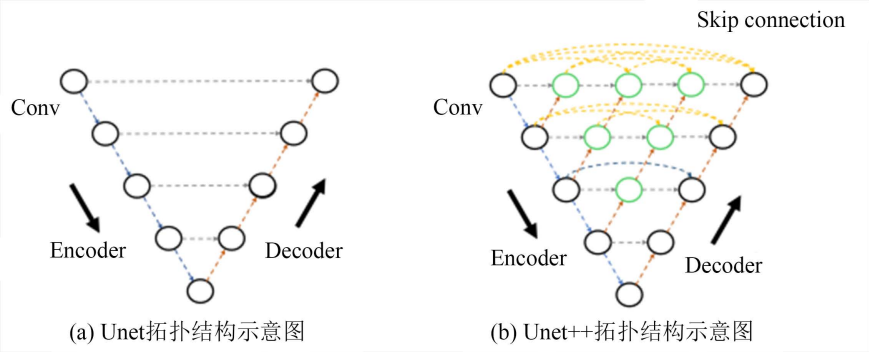


Figure 5. UNET and UNET++ topology diagram
图 5. UNET 和 UNET++拓扑结构图

图卷积神经网络(GCN)可以处理 CNN 在空间语义信息上不足之处[11]，可以通过空间拓扑结构建立空间特征进行机器学习，简要结构如图 6 所示。

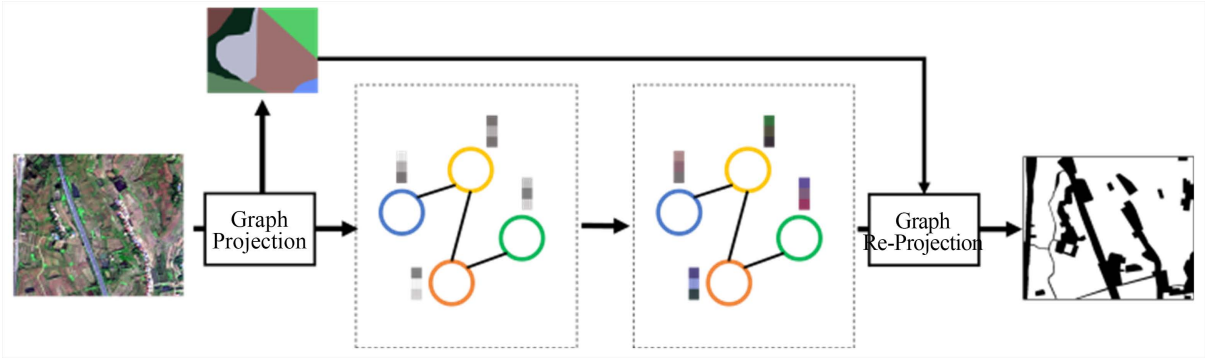


Figure 6. Schematic diagram of GCN model
图 6. GCN 模型示意图

本文基于以上的初步模型建立特征融合的大规模卷积神经网络语义分割训练模型。通过迁移学习的监督式训练方式，对目前图像分割领域主流的模型进行了训练与比较，建立评估指标 PR 参数以及预测图像的噪声分析模型。

3. 训练结果

3.1. 参数设置

本文在训练时以 BCE Loss 与 Dice Loss 相结合作为反向传播的优化方式。在图像语义分割任务中建立了基于 Dice 系数的损失函数[12]，二值交叉熵损失函数为：

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log\left(p\left(y_i\right)\right) + \left(1-y_i\right)\cdot \log\left(1-p\left(y_i\right)\right)$$

其中 y 是标签(绿色点为 1，红色点为 0)，$p(y)$是 N 个点为绿色的预测率。

根据空间位置的不同，可以分别捕获表示每一行的全局上下文信息即高度上下文信息来估计信道的权重。高度驱动的估计信道权重是由卷积层获得的，这些由 N 个卷积层组成的运算可以写成：

$$\mathbf{A} = \mathbf{G}_{\text{up}}\left(\sigma\left(\mathbf{G}_{\text{Conv}}^{N}\left(\cdots\delta\left(\mathbf{G}_{-\mathbf{y},\mathbf{Z}}^{1}\left(\hat{\mathbf{j}}\right)\right)\right)\right)\right)$$

因此该向量与其获得的相应权重一起进行计算，公式为：

$$\tilde{\mathbf{X}}_h = \omega_m\left(\mathrm{X}_\ell\right)\odot\mathbf{X}_h = \mathbf{A}\bigcap\mathbf{X}$$

其中，$\odot$ 代表的是矩阵的乘积运算符号，加权项是典型得分权重乘以特征，计算两个像素点之间的高度关系的亲密度，具体计算过程是这样的：

$$h\left(f_i, f_j\right) = \omega_1 \exp\left(-\frac{\left\|p_i - p_j\right\|^2}{2\sigma_\alpha^2} - \frac{\left\|I_i - I_j\right\|^2}{2\sigma_j^2}\right) + \omega_2 \exp\left(-\frac{\left\|p_i - p_j\right\|^2}{2\sigma_i^2}\right)$$

其中 p 是像素的 2 维位置，I 是图像的 3 维像素值，从这里可以看出：像素距离越近，颜色越接近，特征就越强；

因此融合模型中不同子模型的权重占比与其不同的高度驱动特征融合的计算公式如下：

$$\theta_{ij}\left(x_i, y_j\right) = \mu\left(x_i, y_j\right)\sum_{m=1}^{K}\omega_m \cdot h\left(f_i, f_j\right)$$

3.2. 训练结果分析

我们首先使用 4 类不同的模型进行训练，训练参数如表 1 所示：

Table 1. Training parameters
表 1. 训练参数

数据集	参数	FCN	PSP	UNET++	GCN
Custom	lr	0.02	0.02	0.01	0.02
	epoch	196	200	200	200
	size	64	64	64	64
	batch	256	256	128	256
Test	lr	0.0003	0.0003	0.0001	0.0003
	epoch	198	148	200	298
	size	64	128	64	64
	batch	32	32	32	32

在基本训练参数保持相当时，UNET 系列网络的训练损失函数波动更大一些，这与 UNET 模型的特征有一定关系，相比之下，GCN 利用其图神经提取空间与语义空间的特性，可以更快达到收敛。

迁移学习是将从源域(source domain)学习到的知识迁移到目标域(target domain) [13]，并将允许现有的公开数据集在其他领域重新利用。当设置的批尺寸过小，训练的模型错误率较大[14]。图 7 为使用划分的样本对迁移学习的训练模型进行评估 loss 值变化与 miou 值变化。图 8(b)为使用划分的样本对迁移学习的训练模型进行评估的 miou 值，从图 8 可以看出，模型基本训练达到 0.74 左右，训练时的 miou 值比评估略低，约在 0.67 左右，这表明在迁移学习时使用图像增强的操作对模型训练精度的提升有一定效果。

通过不同的网络对数据集进行训练，然后利用训练好的网络对给定的图像进行图像分割，将分割后的标签与给定的标签进行对比，得到精确率和召回率。通过图 9 的纵向对比可以更清晰的看出 UNET++耕地识别更完整，准确率也更高。UNET++网络和 GCN 网络有一些耕地的细节部分无法识别，FCN 网络识别较好，虽然有一些噪声，但是总体识别效果更优。

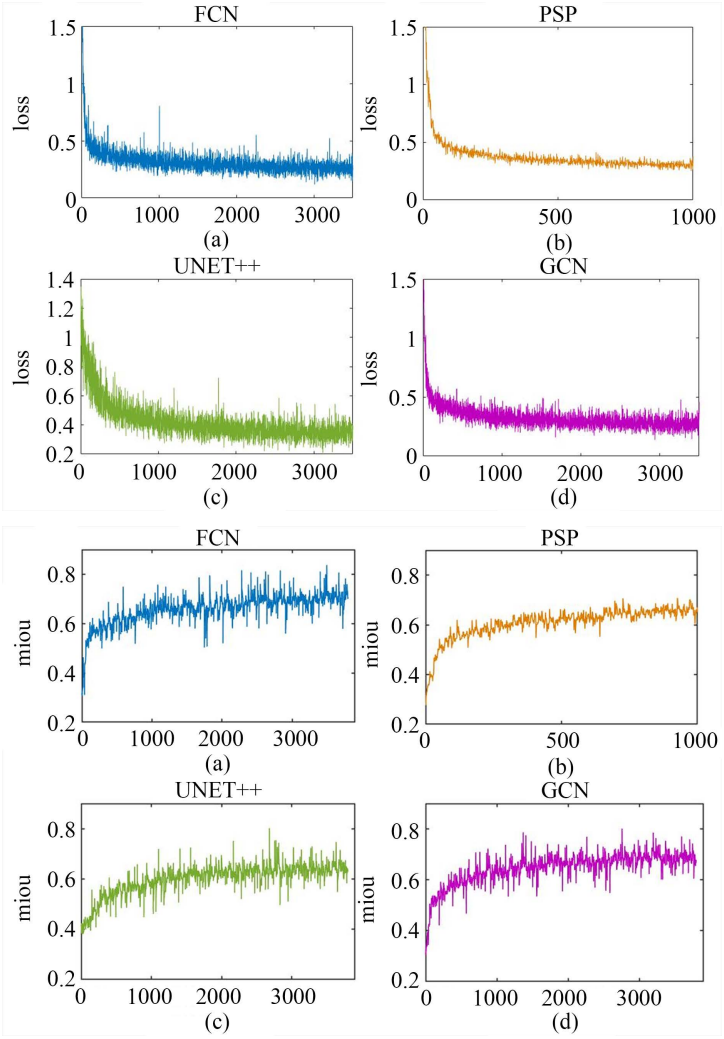


Figure 7. The change of the loss value of the transfer learning training (top), the change of the miou value (bottom)
图 7. 迁移学习训练 loss 值变化(上)，miou 值变化(下)

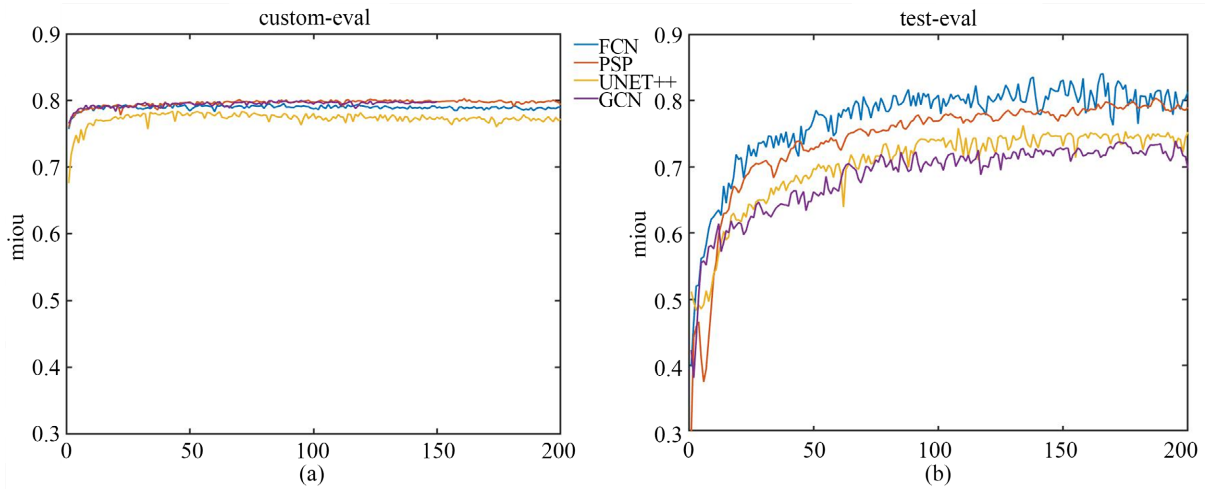


Figure 8. Changes in miou value of the evaluation model
图 8. 评估模型的 miou 值变化

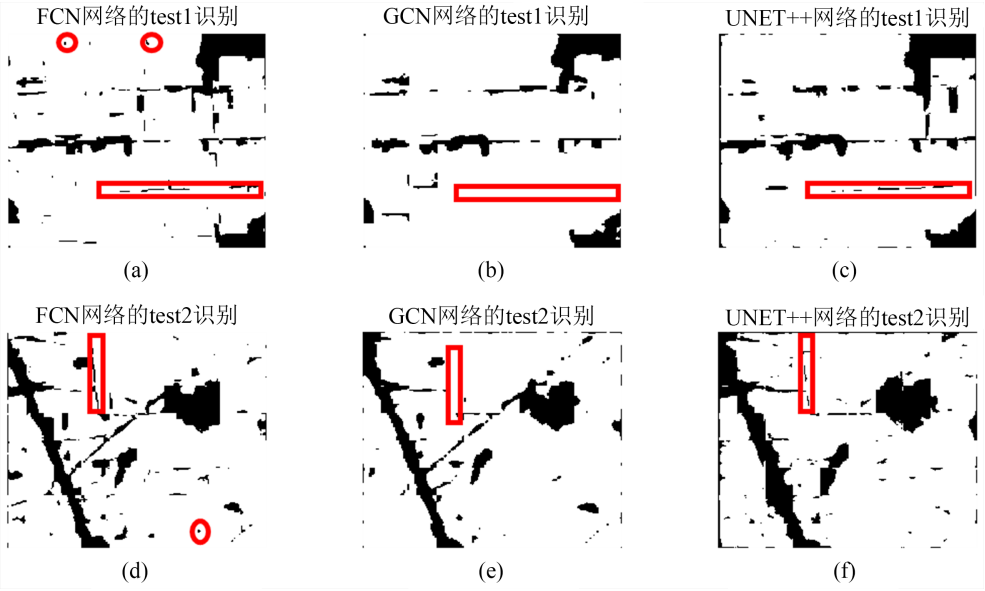


Figure 9. Results label comparison
图 9. 结果标签对比

4. 融合模型结果分析

由于遥感图像拍摄高度不同时，不同的网络分割效果差异较大，因此考虑将分割效果最好的两种网络得到的特征值进行融合分割，将 FCN 网络和 UNET++网络训练得到的参数，分别用于测试图像的分割，得到特征值后，按照不同的权重将特征值叠加，然后进行分割，经过测试，得到 FCN 网络与 UNET++网络的特征值比值为 10/7 时，分割效果最好。

Table 2. Precision and recall
表 2. 精确率和召回率

	precision			recall		
	FCN	UNET++	融合	FCN	UNET++	融合
Data set1	0.9405	0.9402	0.9458	0.9559	0.9563	0.9367
Data set 2	0.9170	0.9413	0.9509	0.9839	0.9873	0.9682
Data set 3	0.8847	0.9155	0.9206	0.9343	0.9466	0.8994
Data set 4	0.9277	0.9335	0.9357	0.9437	0.9574	0.9189
Data set 5	0.9010	0.9031	0.8837	0.8973	0.9247	0.9005
Data set 6	0.9578	0.9509	0.9614	0.9813	0.9890	0.9670
Data set 7	0.9312	0.9408	0.9440	0.9452	0.9647	0.9189
Data set 8	0.9595	0.9620	0.9671	0.9771	0.9860	0.9659

根据表 2 的精确率和召回率可以看出融合后的特征值的得到的分割图像指标更好一些，其中精确率要比单一网络的数值要高，而召回率因为 FCN 网络和 UNET++网络差异较大，融合后的召回率处于 UNET++网络和 FCN 网络之间。从数值上观察还可以发现，融合后的模型精确率提高了 9%，更重要的是融合后的模型适用于分割更多的场景，不会出现某些场景精确率和分割率特别低的情况，鲁棒性得到明显提高。图像之间的差异可以通过观察测试图像分割效果来分析。

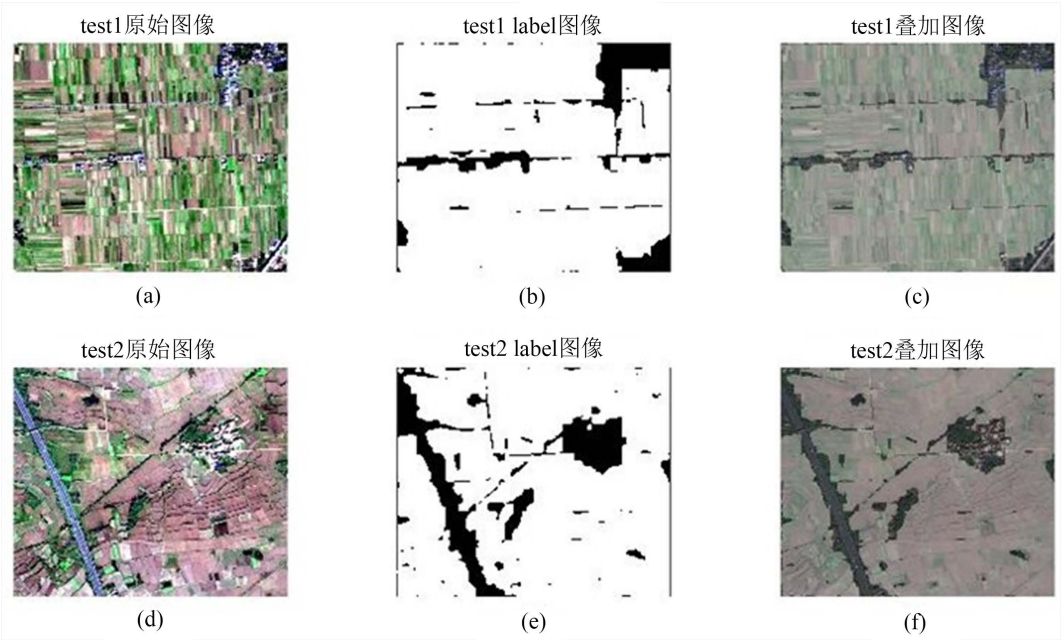
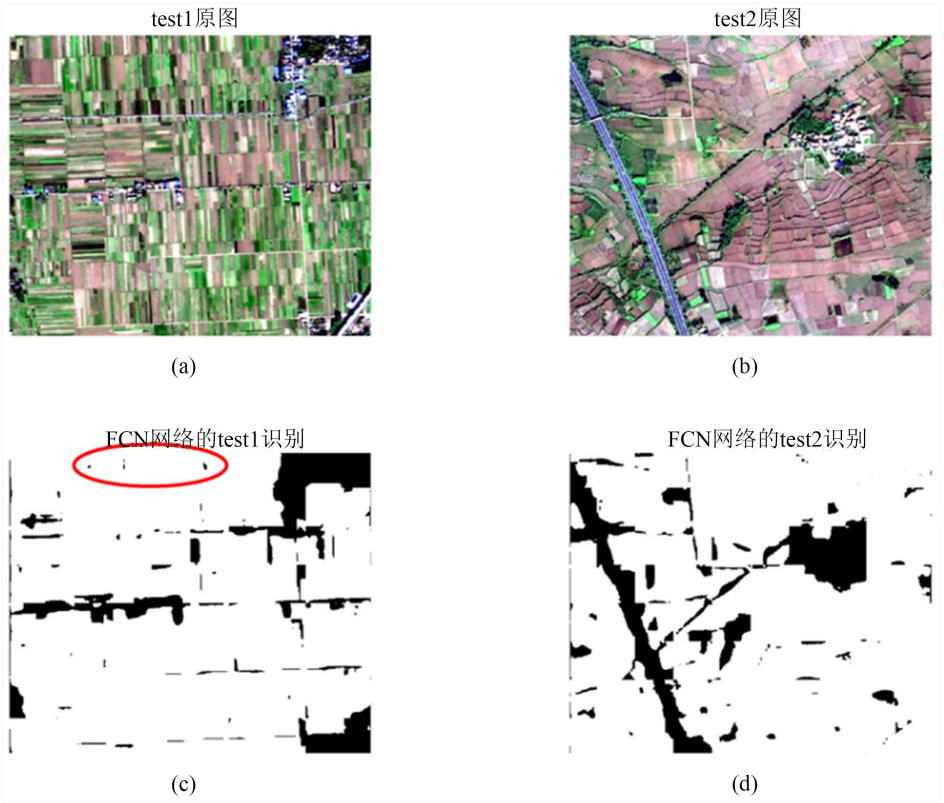


Figure 10. Noise comparison effect after fusion of FCN and UNET
图 10. FCN 和 UNET 融合后的噪声对比效果

综合图 10 和图 11，可以发现 FCN 在分割拍摄距离与地面更近的图像时效果较好，UNET++网络在分割拍摄距离与地面更远的图像时效果更优，特征值融合后的结果综合了单一模型分割的优点，提升了鲁棒性[15]。可以看出建立与高度有关的权重后融合的分割图像比单一网络的分割图像效果要好。



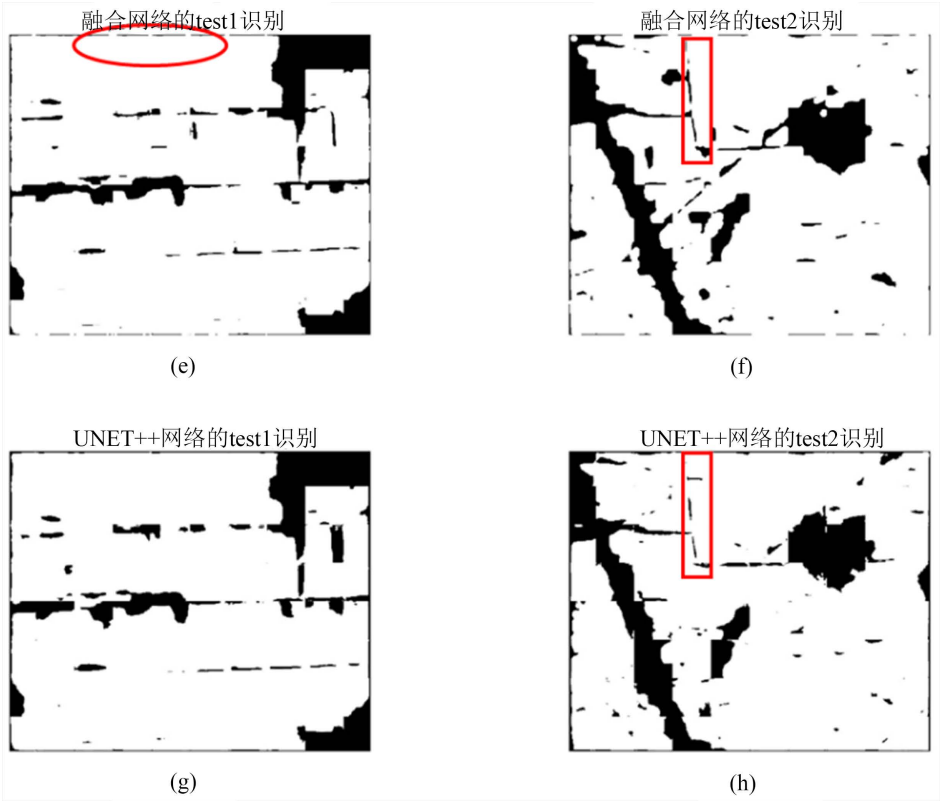


Figure 11. UNET++ fusion comparison noise analysis
图 11. UNET++融合对比噪声分析

本文将 FCN 与 GCN 模型在模型训练时进入融合，作为训练的两个分支，根据 GCN 和 FCN 的模型特性，分别对训练集进行增强空间语义信息，减弱噪声的干扰[16]，如图 12 所示。

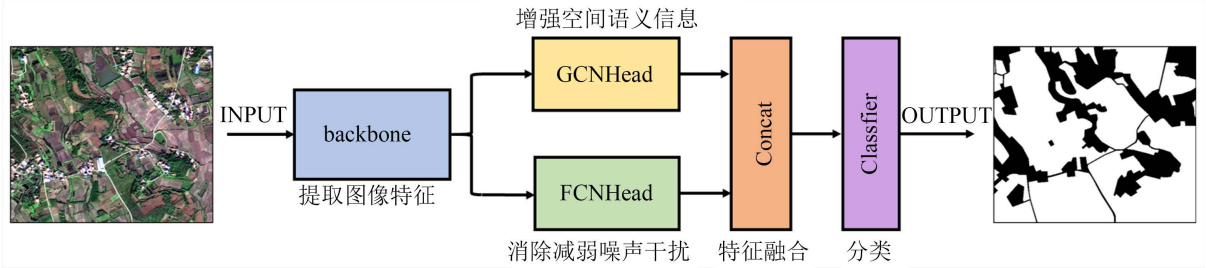


Figure 12. The fusion structure of this paper
图 12. 本文融合结构图

与经典的卷积神经网络(CNN)在卷积层之后使用完全连接的层来获得固定长度的特征向量以进行分类不同，FCN 可以接受任何大小的输入图像，并使用反卷积层对最后一个卷积层的特征图进行上采样，以将其恢复到与输入图像相同的大小，从而可以为每个像素生成预测，同时保留原始输入图像中的空间信息，从而在语义级别上解决图像分割问题。本文借鉴了 FCN 网络，其网络结构由两部分组成。前一部分的收缩网络使用 3 × 3 卷积和合并下采样来捕获图像中的上下文信息；后一部分的扩展网络采用 3 × 3 卷积和上采样，以达到精确定位图像所需分割部分的目的。此外，网络中还使用了特征融合，将前一部分的下采样特征和后一部分的上采样特征进行拼接融合，获得更准确的上下文信息，达到更好的分割效

果。UNET++将不同规模的 UNET 结构集成到一个网络中，捕捉不同级别的特征，并通过特征叠加将它们整合成更浅的 UNET 结构，使得融合时特征图的尺度差异更小。

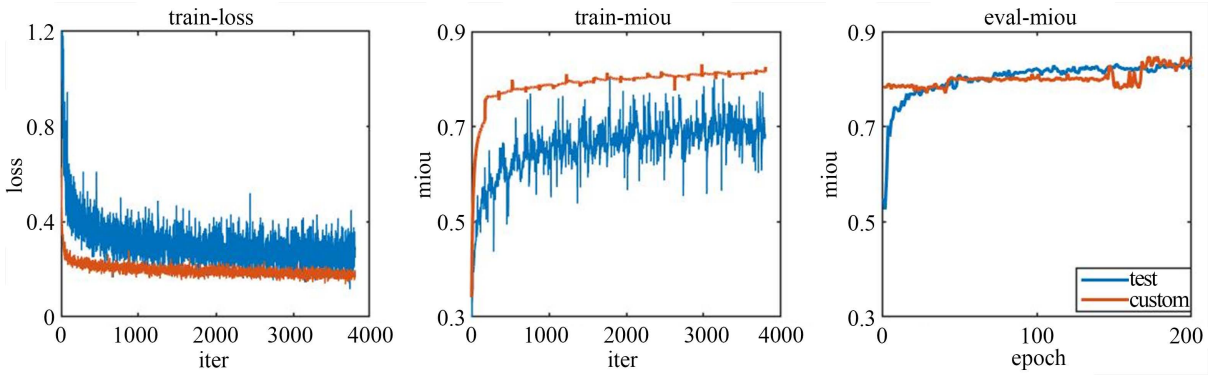


Figure 13. Training flow chart
图 13. 训练流程图

本文模型训练过程如图 13 所示，训练参数和 GCN 模型保持一致，训练时 loss 参数可以较快达到收敛，miou 值稳定到 0.8 左右，迁移学习因为样本数量过少，因此在学习时收敛 loss 波动都比较大。测试集评估的 miou 值比最终比训练时高 0.1，表明图像增强起到作用。

融合后模型的分割图像与叠加图像如图 14 所示：

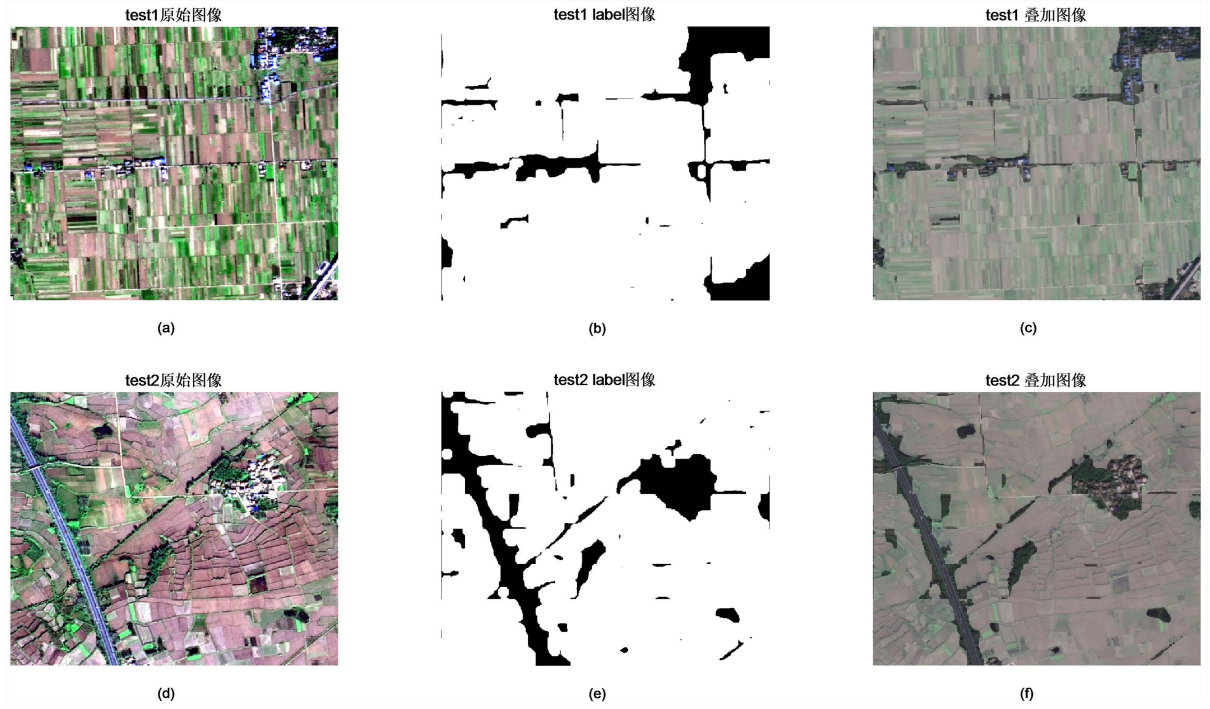


Figure 14. Output renderings
图 14. 输出效果图

可以看出两张测试图片中的噪声都得到明显的减少，而且道路，房屋等非耕地部分大的轮廓更加明显，与测试图像更加贴近。对比融合模型网络和单独的 FCN 和 GCN 网络如图 15 所示：

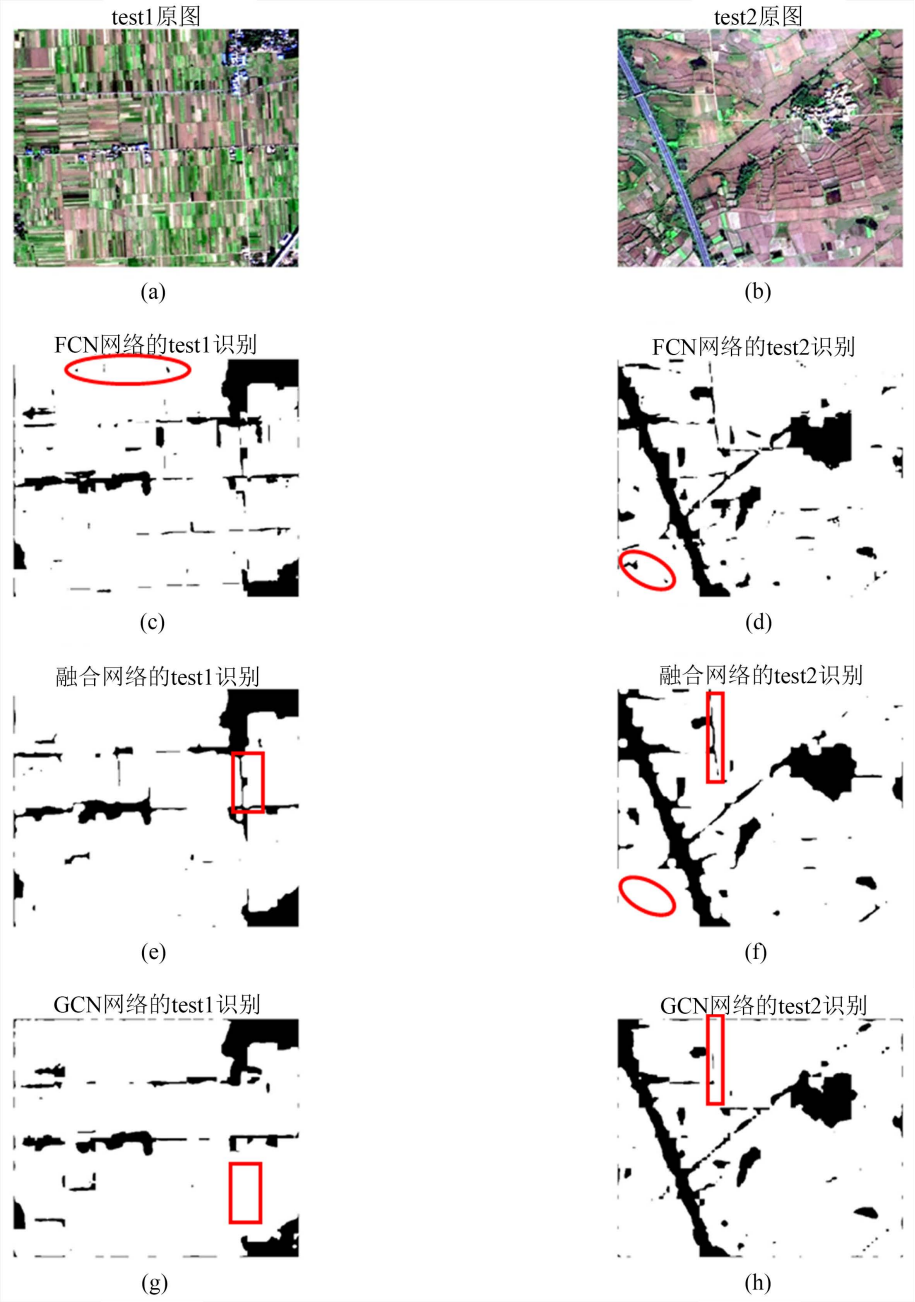


Figure 15. Comparison of FCN and GCN fusion effects
图 15. FCN 与 GCN 融合效果对比图

因此 FCN 与 GCN 网络融合后模型的分割结果更好，对比发现，融合网络分割结果中的噪声要比 FCN 网络中的噪声更少，耕地轮廓比 GCN 网络的分割结果更加明显，对比同样可以发现，融合模型中的噪声比 FCN 网络更少，道路以及耕地轮廓的识别比 GCN 网络更加清晰。而且相对于两种场景下的遥感图像，融合后的模型对两种场景下的识别得到改善，鲁棒性得到提高。

5. 总结与展望

本文主要通过以卷积神经网络模型进行深度学习，通过先验数据集训练出较好的模型。在模型训练

的选择方式上，本文通过迁移学习的监督式训练方式解决缺乏庞大数据集的问题，具体来讲，主要先通过整理的数据集训练到模型符合评估指标要求后，再用样图分割成多个数据样本进行迁移学习。在训练模型的构建上，首先利用目前图像分割领域主流的模型进行训练并对比评价，主要包括全卷积模型 FCN 或 PSP，改进的 UNET 系列，以及图神经的模型 GCN。通过对比及分析，由于不同模型对于遥感图像在不同比例尺采集图像的预测效果鲁棒性不高，本文根据模型的特点，赋予两种模型不同权重融合结果，权重与采集时的高度建立函数关系，本文在在训练模型时就将训练模型分成两个分支，利用 FCN 和 GCN 的模型特性，分别做降噪分支和提取空间语义分支，最后进行特征融合。

但是本文亦需要借助多种评价指标，如连通性，噪声等来评价模型真实的优劣，靠得非常近的耕地边缘会被错误地合并在一起。因此，下一步将着手解决这些问题。

参考文献

[1] 余帅, 汪西莉. 含多级通道注意力机制的 CGAN 遥感图像建筑物分割[J]. 中国图象图形学报, 2021, 26(3): 686-699.

[2] Yang, G., Zhang, Q. and Zhang, G. (2020) EANet: Edge-Aware Network for the Extraction of Buildings from Aerial Images. *Remote Sensing*, **12**, Article No. 2161. https://doi.org/10.3390/rs12132161

[3] 季顺平, 魏世清. 遥感影像建筑物提取的卷积神经元网络与开源数据集方法[J]. 测绘学报, 2019, 48(4): 448-459.

[4] 朱明胜. 图像增强技术研究与实现[D]: [硕士学位论文]. 合肥: 安徽大学, 2014.

[5] 邓仕超, 黄寅. 二值图像膨胀腐蚀的快速算法[J]. 计算机工程与应用, 2017, 53(5): 207-211.

[6] 张刚. 基于深度学习的遥感图像语义分割关键技术研究[D]: [博士学位论文]. 成都: 中国科学院大学(中国科学院光电技术研究所), 2020.

[7] Lu, Y., Chen, Y., Zhao, D., *et al*. (2019) Graph-FCN for Image Semantic Segmentation. In: Cham, Ed., *Advances in Neural Networks*, Springer, Berlin, 97-105. https://doi.org/10.1007/978-3-030-22796-8_11

[8] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 3431-3440. https://doi.org/10.1109/CVPR.2015.7298965

[9] Chen, L.C., Papandreou, G., Kokkinos, I., *et al*. (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

[10] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Munich, 5-9 October 2015, 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[11] 王健宗, 孔令炜, 黄章成, 肖京. 图神经网络综述[J]. 计算机工程, 2021, 47(4): 1-12.

[12] 徐文娜. 基于高分辨率全卷积网络的遥感影像耕地提取方法研究[D]: [硕士学位论文]. 深圳: 中国科学院大学(中国科学院深圳先进技术研究院), 2020.

[13] Pan, S.J. and Yang, Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, **22**, 1345-1359. https://doi.org/10.1109/TKDE.2009.191

[14] Wu, Y.X. and He, K.M. (2018) Group Normalization. *International Journal of Computer Vision*, **128**, 742-755. https://doi.org/10.1007/s11263-019-01198-w

[15] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., *et al*. (2018) A survey on Deep Learning Techniques for Image and Video Semantic Segmentation. *Applied Soft Computing*, **70**, 41-65. https://doi.org/10.1016/j.asoc.2018.05.018

[16] Zhang, W., Li, R., Deng, H., *et al*. (2015) Deep Convolutional Neural Networks for Multi-Modality Isointense Infant Brain Image Segmentation. *NeuroImage*, **108**, 214-224. https://doi.org/10.1016/j.neuroimage.2014.12.061