Land Cover Mapping Using Semantic Segmentation Models

Guide: [1]Dr. Sarika Zaware, Authors: [2]Vaishnavi Patil, [3]Rajlaxmi Bhosale, [4]Nikita Shetty, [5]Vama Shah

[1]Head, Computer Engineering Department, AISSMS Institute of Information Technology, Pune, India
sarika.zaware@aissmsioit.org
[2]Computer Engineering, AISSMS Institute of Information Technology, Pune, India
vaishnavi.patil4848@gmail.com
[3]Computer Engineering, AISSMS Institute of Information Technology, Pune, India
rajlaxmib22@gmail.com
[4]Computer Engineering, AISSMS Institute of Information Technology, Pune, India
shettynikita277@gmail.com
[5]Computer Engineering, AISSMS Institute of Information Technology, Pune, India
vamashahnamo04@gmail.com

Abstract. Land cover mapping is the process of creating a map that shows the types of land cover in a given area. This includes vegetation, land use, surface features, and other natural and human-made features. Land cover mapping is used to help understand environmental processes, land use, and land management. This also helps identify potential risks from natural disasters, such as floods and landslides. By understanding land cover, land managers can make decisions about how to use best and manage the land. Deep learning has opened up new opportunities for extracting LC information from remotely sensed imagery. Deep learning models are typically applied to satellite imagery and aerial photographs, which are then analyzed with the help of convolutional neural networks (CNNs). CNNs are able to accurately identify and classify land cover types in high-resolution imagery. This paper includes three Semantic Segmentation models, namely PSPNet, UNet, and FPN on the DeepGlobe dataset for land cover mapping, and has achieved IoU scores of 91%, 78%, and 64% respectively.

Keywords: Deep learning, image classification, land cover (LC) mapping, semantic segmentation, Intersection Over Union(IoU), Feature Pyramid Network(FPN), Pyramid Scene Parsing Network(PSPNet).

1 Introduction

Land cover mapping is a tool that provides essential information about Earth's land areas and their patterns. Land cover is used for monitoring agricultural land, urban planning, and observing land development over a time period. This uses satellite images to observe the land cover and then classifies the images. This requires image preprocessing and processing techniques to accurately map land cover patterns
Deep learning is a very significant technique for land cover mapping and image segmentation. Deep learning is a subsection under Machine Learning that replicates the way humans acquire certain types of knowledge and learn representations. This is helpful to data scientists who are responsible for the collection, analysis, and interpretation of large amounts of data; deep learning provides an incentive to these processes, making them faster and easier.

Image segmentation is the process of classifying the image pixel-wise, into one of the many classes provided. This paper compares the models such as PSPNet, FPN, and UNet to get better results for image mapping and get accurate results. This article overcomes the following drawbacks that were not resolved by the articles published before:

- *Low data quality*

Low data quality in land cover mapping can be caused by a variety of factors, including inaccurate data collection, limited spatial or temporal resolution, or inadequate data validation. To improve data quality, it is important to ensure that the data collection process is accurate, the classification system is appropriate for the area being mapped, the spatial resolution is sufficient to capture the desired level of detail, and the data is validated to ensure accuracy.

- *Data-inconsistency*

Data inconsistency can lead to inaccurate land cover maps and can make this difficult to accurately assess the state of the land cover. To address this issue, it is important to ensure that the data used for land cover mapping is accurate, up-to-date, and appropriately processed. Additionally, it may be necessary to use multiple sources of data and combine them to create the most accurate representation of the land cover.

- *Data validation*

Data validation in land cover mapping is an important process that ensures that land cover maps accurately reflect the characteristics of the landscape. This is done by comparing the data collected from various sources, including satellite imagery, aerial photos, field visits, and other methods, to ensure the accuracy of the mapped land cover. Data validation also involves evaluating and monitoring changes in land cover over time to identify any changes or inconsistencies. Additionally, data validation may involve using ground-truth data from field visits to ensure that the mapped land cover accurately reflects the landscape.

- *Unevenly distributed data*

Unevenly distributed data in land cover mapping occurs when there are areas of the map with more land cover classes than other areas. These imbalances may be caused by factors such as varying land use in different parts of the region, limited availability of data for certain land cover classes, or the presence of outliers in the data. Unevenly distributed data can make it difficult to accurately classify land cover classes and produce accurate and reliable results. To address this issue, it is helpful to apply data preprocessing techniques such as smoothing, outlier detection, and data augmentation. Additionally, using an appropriate classification algorithm and a larger dataset can help reduce the effects of unevenly distributed data.

DeepGlobe

The DeepGlobe dataset is a publicly available satellite image dataset created for the DeepGlobe Satellite Image Understanding Challenge. This consists of over 1 million high-resolution satellite images from DigitalGlobe, annotated with labels for land cover, land use, and building damage. The dataset is divided into three categories: Land Cover, Land Use, and Building Damage. Each category contains an equal number of images, with labels for each image. The Land Cover category contains labels for classifying land covers such as open water, urban land, or barren land. The Land Use category contains labels for classifying land use such as agriculture, forests, or residential areas. The Building Damage category contains labels for classifying the extent of damage to buildings, such as no damage, minor damage, or major damage. The dataset is intended to be used to train algorithms to classify satellite images, leading to improved understanding and analysis of satellite imagery.

Sr. No	CLASS	COLOR(R,G,B)	COLOR(hex code)
1	Urban Land	(0,255,255)	00ffff
2	Agricultural Land	(255,255,0)	ffff00
3	Range Land	(255,0,255)	ff00ff
4	Forest Land	(0,255,0)	00ff00
5	Water	(0,0,255)	0000ff
6	Barren Land	(255,255,255)	ffffff
7	Unknown	(0,0,0)	000000

Table. 1. Dataset Classes

2 Related Works

Paper [1] deduces MFF and Attention mechanism for accurate Image Segmentation. This initially uses an encoder network to extract feature maps and feed them to MFF modules to get aggregate features. Then feed these to Attention Modules to extract local and global features. Lastly, it uses the decoder to project the features onto a larger dimension space.

Paper[2] implements ATD-LinkNet with a replacement for the residual part in the downsampling block with an AT block. Here, the author uses two datasets PostDam and the DeepGLobe road extraction dataset. Also, compares two models, the generic D-LinkNet and the proposed model, ATD-LinkNet.The proposed model performs better on both datasets.

[3]A dilated CNN stack block with a merging module as output makes up a dense dilated convolutional network. Dilated convolution is the first step in a basic DC block, which is followed by a PReLu block. This feeds the top layer, with the output and input stacked together. DDCM-Net only fuses low-level signal features just once before

the outcome is predicted. Instead of combining multiscale characteristics obtained at numerous encoder layers, CNN layers are used. When compiling and predicting the segmentation maps, it employs the Adam optimizer.

In paper [4], GAN is used for full space domain adaptation and subsequent prediction of segmentation maps using source and target images. To close the gap between source images with labels and other original images, it first constructs image space domain adaptation and feature space domain modules. The output image style transfer image and use it to train an FCN model to predict a segmentation map.

Paper [5] deduces a deformable attention network using a deformable attention module. It is more adaptable to HRRS image structures that capture each pixel's neighboring information. Raw ResNet50's standard convolutional layers come with a DAM that allows them to manage sampling across a wide variety of feature levels and aggregate multi-scale context data. With fewer mechanisms, it operates better than global spatial attention.

[6] The Grassland class appears to be underrepresented, yet most models, aside from unsupervised k-means, can make accurate predictions about it. Here, the standard methods without dedicated adaptations have limited capabilities. The most successful algorithm for predicting the grassland class appears to be the Random Forest algorithm. This algorithm is known for its ability to handle high-dimensional datasets, which is likely why it is so successful at predicting the Grassland class. Other models such as SVM and Neural Networks can also be used to predict the grassland class, but they tend to have lower accuracy rates compared to the Random Forest algorithm. Additionally, feature engineering is important for improving the accuracy of models when predicting the grassland class.

[7] observes that a high degree of uncertainty is obtained in pixels where the output does not match the ground truth. This is because the model is uncertain about its predictions for these pixels, as it does not know for sure whether the prediction is correct or not. As a result, it assigns a higher level of uncertainty to these pixels, to indicate that its prediction may not be reliable.

[8] The UNet approach, as opposed to SwinUNet, creates smoother segmentation masks, according to the initial observation. On the other hand, SwinUNet's segmentation masks are significantly more precise and thorough. Due to these observations, it is decided to combine these two methodologies and compute the final prediction by averaging the predictions from the two models. The second observation was that the UNet model was much more computationally expensive than the SwinUNet, while still producing results that were not as good as the SwinUNet. This motivated the idea of using the SwinUNet as the primary model, and only using the UNet when the SwinUNet failed to produce an adequate result. This allows for a reduction in computational cost while still producing good results.

[9] In this paper, the authors used AlexNet as the backbone network to extract deep semantic features from the different modalities of data, including RGB, infrared, and LiDAR. Finally, they used the bilinear pooling model to combine the several modal compact feature maps efficiently and produce discriminative bilinear fusion features for classifying land cover. The authors also used a support vector machine (SVM) classifier to classify the fused feature map. The results showed that the method achieved better performance than the traditional methods in land cover classification.

[10] Fusing multimodal, multisource, and multitemporal data for Local Climate Zone classification can be beneficial in a variety of ways. First, it can help to improve the accuracy of the classification, by providing more comprehensive information about the local climate. Combining multiple data sources, including satellite imagery, ground-based observations, and weather forecasts, can provide a more comprehensive view of the local climate and its associated factors. It can also provide a more accurate representation of the local climate over time, as the data from different sources can be used to create a historical record of the local climate. Finally, combining multiple data

sources can also improve the accuracy of the classification by reducing the likelihood of false positives or false negatives.

In [11] it is observed that overall, the NN performed better in cases of large ground areas with fewer sharp edges. It was able to accurately classify the classes with a high degree of accuracy. The NN also performed better on images with a higher resolution. The NN was able to detect and classify different classes in the image with a high degree of accuracy. However, the NN faced difficulties in cases of small ground areas and sharp edges.

[12] The Land-cover Domain Adaptive Semantic Segmentation (LoveDA) dataset is introduced in this research in order to promote semantic and transferable learning. Both land-cover semantic segmentation and unsupervised domain adaptation (UDA) tasks can be completed using the LoveDA dataset. Three issues with large-scale remote sensing mapping are highlighted by the LoveDA dataset: multi-scale objects, complicated background samples, and erratic class distributions.

[13] The primary concept of the semi-automatic auxiliary approach for classifying land cover proposed in this paper is the usage of an interactive segmentation network. The user can then click within and outside of objects to direct the model as it performs the task of segmenting the patches using an interactive segmentation method. To better integrate the characteristics of various scales, this model also incorporates several interactive modules.

[14] The mapping issues addressed in this study include the lack of annotated datasets, the inaccessibility of some satellite imagery, and the disparity in the spectral bands that are readily available.

In [15] presents a new model called DPPNet (Depth-wise Pyramid Pooling Network), which makes use of a dense block with several dilated depth-wise residual connections and the newly constructed Depth-wise Pyramid Pooling (DPP) block. Using the Land-cover and WHDLD high-resolution Space-borne Sensor (HRS) datasets, the proposed DPPNet model is assessed and contrasted with the benchmark semantic segmentation models.

Paper [16] develops a hybrid deep learning model that combines DenseNet and U-advantages Net into a potent new tool. Pixel-by-pixel land cover classification is accomplished in this manner. While U-long-range Net's connections combine the encoder and decoder paths to keep low-level features, DenseNet replaces U-contraction Net's path to extract features at various sizes. It demonstrates that the suggested hybrid network beats other current models by a significant margin and operates at the cutting edge.

In [17], the FCSN (fully convolutional segmentation network) for HSI (Hyperspectral image) classification is the suggested HSI cube-generating approach that cannot accurately imitate true geographical land-cover distributions, according to experimental findings on original databases.

In [18] semantic segmentation of high-resolution aerial images is proposed using a unique coarse-to-fine CCENet.CaEM assists CCENet in improving object boundary results utilizing prior knowledge that is class-aware. The outcomes of the comparative experiments and ablation studies attest to the suggested framework's superiority in terms of high labeling accuracy and computation efficiency. Using three publicly available datasets, quantitative and qualitative demonstrate that CCENet delivers more accurate labeling outcomes than competing well-known algorithms. To close the domain gap between the two ISPRS datasets and address the domain shift issue, the following research will focus on domain adaptation utilizing metrics learned from class representations.

In [19] a new MRF-MEO approach to the inclusion of pixel and object granularities should be optimized. Since the MRF- MEO takes into account the specific information at the finer granularity, the advantages of bigger granularities can be maintained. Many remote-sensing picture types confirm the efficiency of the suggested approach.

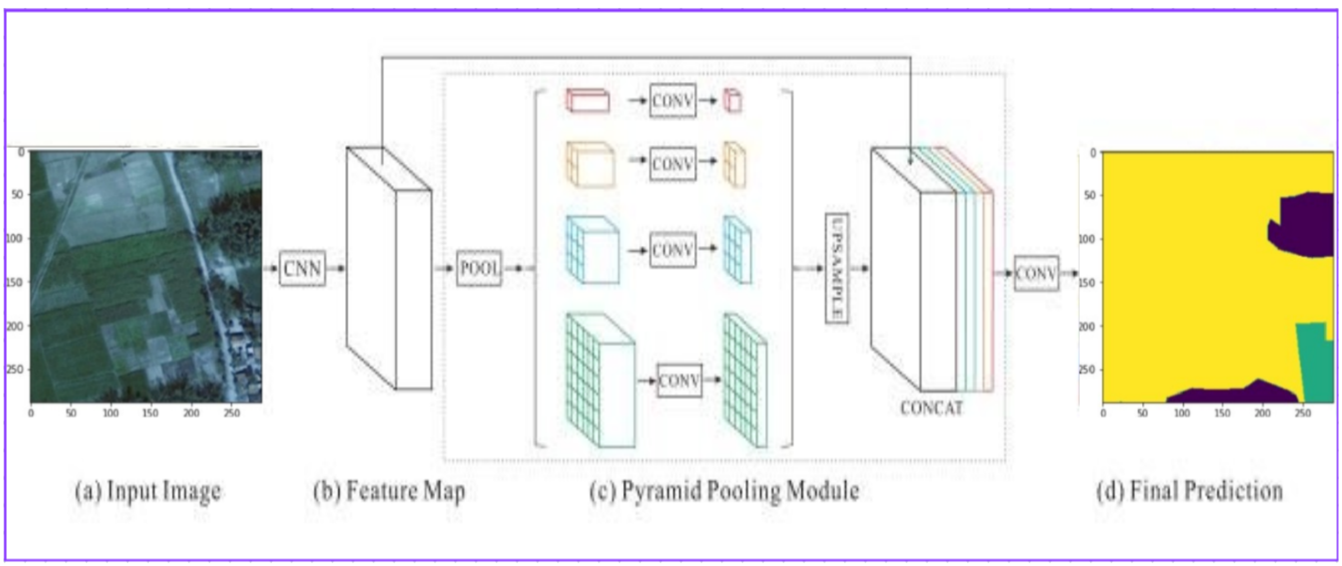
[20] RSI-Net is implemented. It is aimed at improving the effectiveness of semantic segmentation by incorporating both image-level and graph-level information, where deconvolution processes are inserted to retrieve the initial spatial resolution of the feature maps. Finally, comparisons with multiple cutting-edge RSI semantic segmentation techniques are made using three publicly available RSI datasets.Several experimental findings indicate that contextual information is crucial for improving the segmentation in pixel-wise semantics. the pixel-superpixel correlation.Further research should examine the feature map encoding model and its decoding model.

[21]This research proposes a Swin Transformer fusion with a Gabor filter-based semantic segmentation approach for remote sensing images. First, various levels of picture information are extracted using a Swin Transformer as the backbone network. Then, a Gabor filter is used to extract the input image's texture and edge features, and a feature aggregation module (FAM) and an attentional embedding module are introduced to combine the multilayer features (AEM). The fully linked conditional random field is used to maximize the segmentation outcome (FC-CRF).

3 Methodology

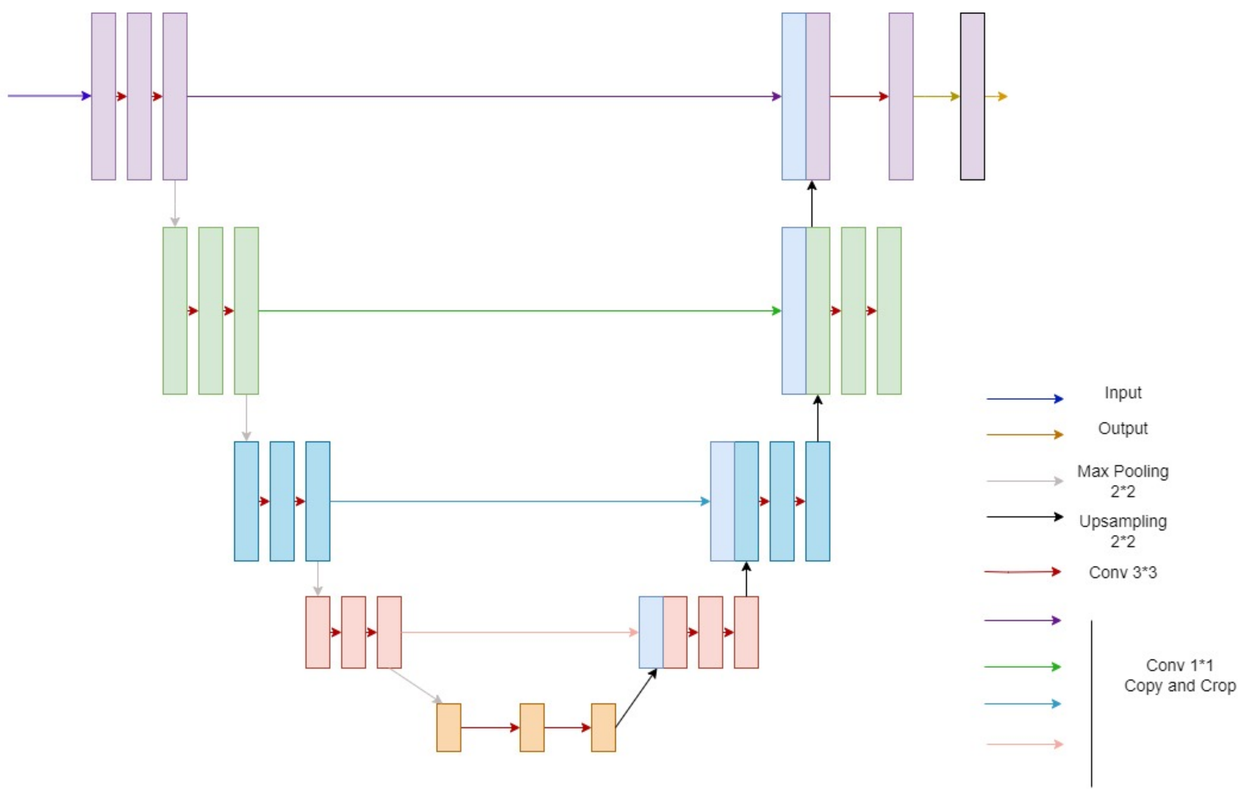
3.1 PSPNet

In this paper, PSPNet is used and trained on the DeepGlobe Land Cover Segmentation dataset. The dataset has images and the corresponding masks consisting of 7 classes of land cover. In PSPNet, the global context information is exploited using different region-based context aggregation methods via the pyramid pooling module [22]. The pyramid pooling module is the primary component of this model since it enables the model to recognize the global context in the image and classify the pixels according to that context. [22]. This approach gives a better improvement in the Intersection over Union (IoU) score as compared to the other models. It sophisticates the pixel-level prediction tasks [22]. It provides a fairly good IoU score along with a low inference time, which is as low as 18 ms [22].



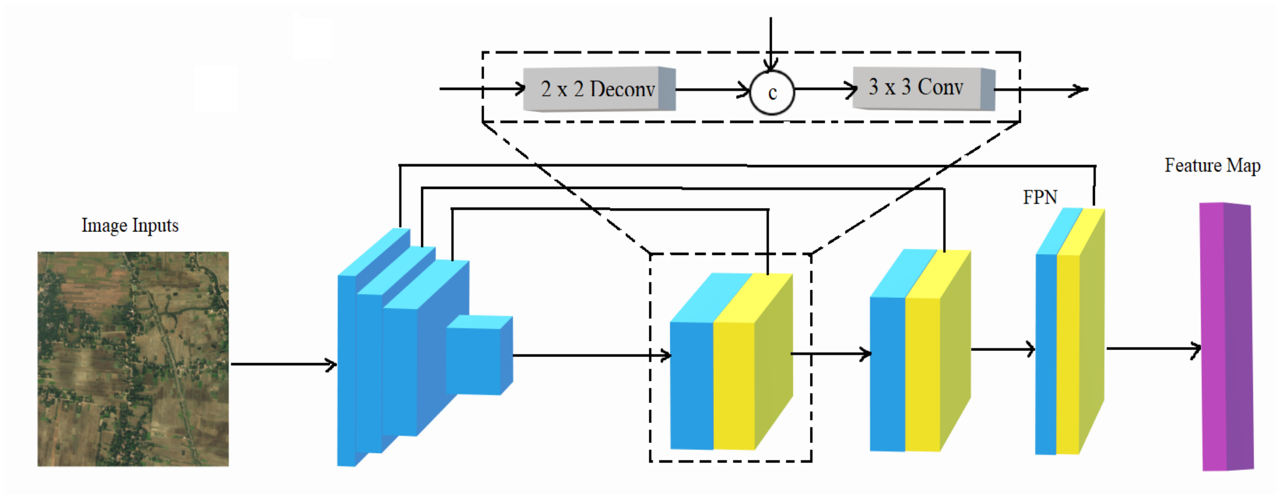
3.2 UNet

In this study, UNet was employed, and the DeepGLobe Land Cover Segmentation dataset was used to train it. It features a total of 800 photos and the seven classes of masks that go with them. The Convolutional Network Model is the foundation of UNet. The UNet architecture has two routes. The contraction path, often referred to as the encoder, is the first path that is utilized to record the context of the image [23]. The symmetric expanding path, sometimes referred to as a decoder, is utilized to provide exact localization via transposed convolutions [23]. UNet performs similarly to PSPNet in terms of performance, although PSPNet is outperformed by UNet inference time, which is roughly 195 ms.



3.3 FPNeT

FPNET was trained on the DeepGlobe Land Cover Segmentation dataset and used in this paper. The collection includes photos and matching masks for each of the seven land cover classifications. A feature extractor called Feature Pyramid Network (FPN) was created with accuracy and speed in mind for such a pyramid concept [24]. For segmentation, it takes the place of detectors like Faster. R-feature CNN's extractor produces many feature map layers (multi-scale feature maps) with higher-quality data than the standard feature pyramid.



4 Results

MODELS	IOU SCORE	EPOCHS	INFERENCE TIME
PSPNet	0.91	50	18 ms
UNet	0.78	50	195 ms
FPN	0.64	50	140 ms

Table. 2. Results of Segmentation Models

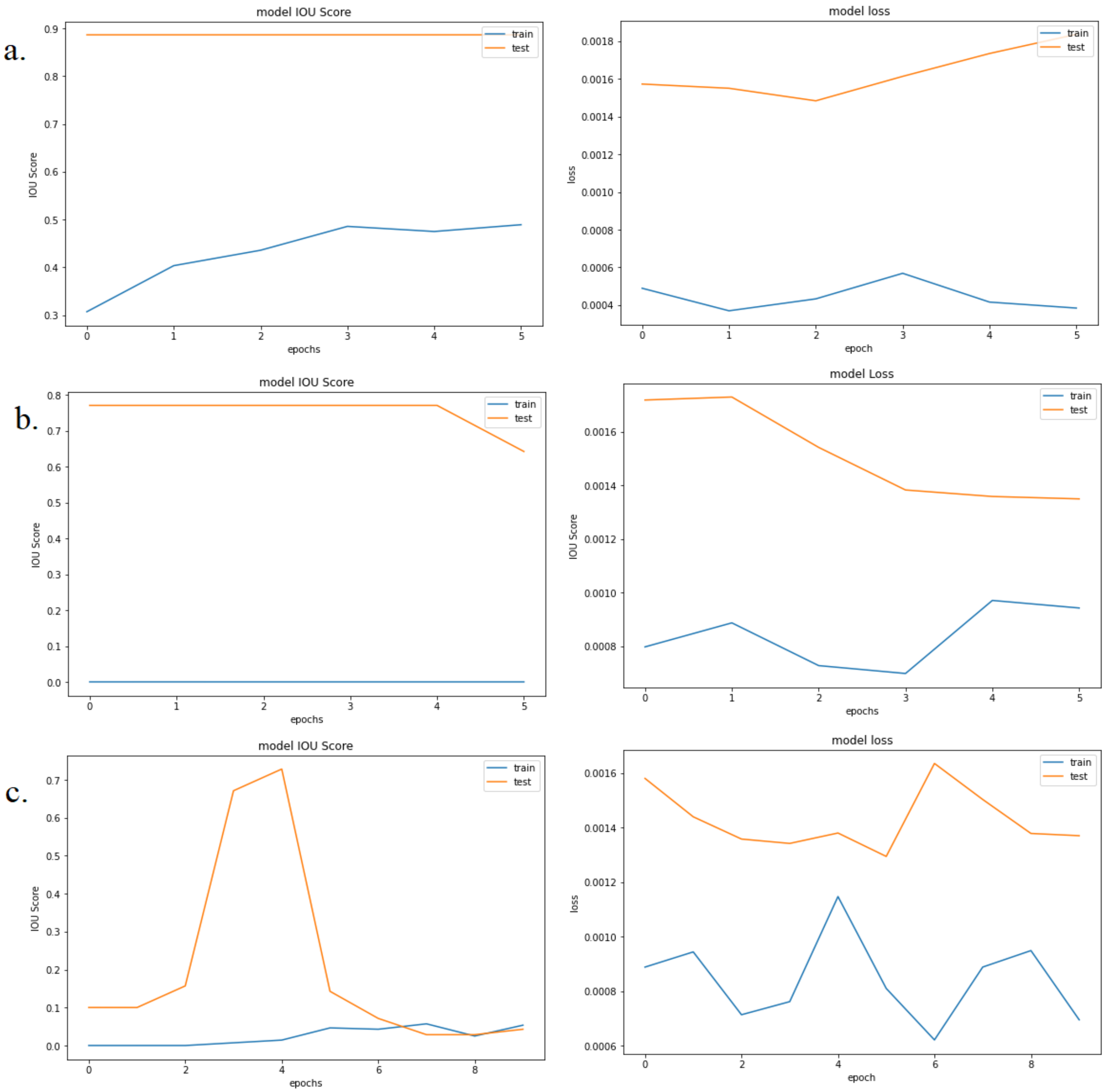


Fig 2. a) Results of PSPNet, b) Results of UNet, c) Results of FPN

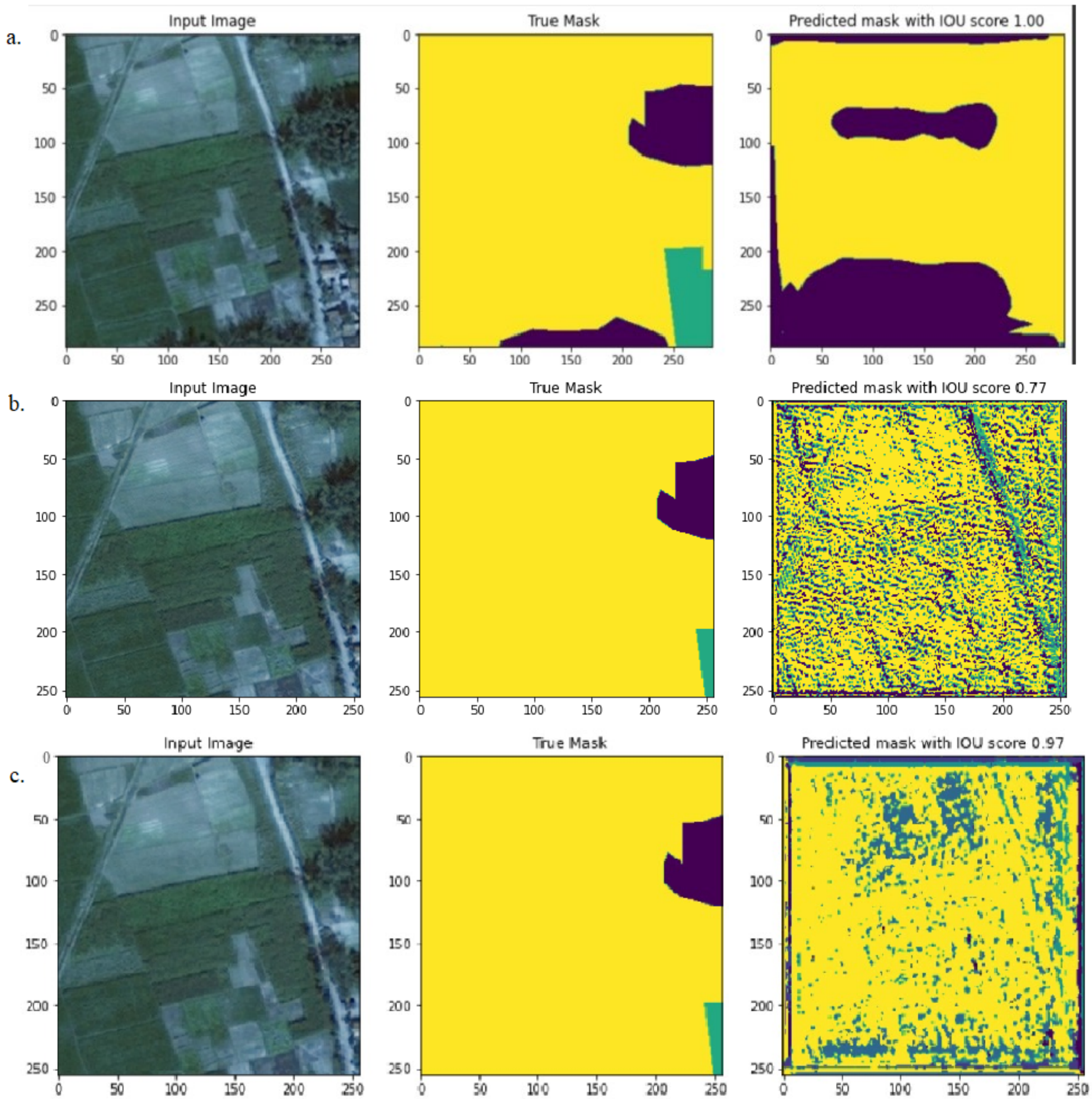


Fig 3. a) Prediction of PSPNet, b) Prediction of UNet, c) Prediction of FPN

5 Conclusion:

In conclusion, the results of this study demonstrate that semantic segmentation models can be effectively used for wide-area land cover mapping with satellite imagery. The best-performing models for this task were PSP, FPNet, and UNet, which achieved an IoU score of 91%, 78%, and 64%, respectively, with the reference land cover maps. The models further demonstrated good generalization ability, with the best model (PSPNet) having a much shorter inference time. The given results offer a standard of accuracy by which the recently suggested models should be measured.

References

[1] S. Wang, X. Mu, D. Yang, H. He and P. Zhao,
"Attention Guided Encoder-Decoder Network With Multi-Scale Context Aggregation for Land Cover Segmentation," in *IEEE Access*, vol. 8, pp.
215299-215309, 2020, doi: 10.1109/ACCESS.2020.3040862.
[2] X. Qi, K. Li, P. Liu, X. Zhou, and M. Sun, "Deep Attention and Multi-Scale Networks for Accurate Remote Sensing Image Segmentation," in *IEEE Access*, vol. 8, pp. 146627-146639, 2020, doi: 10.1109/ACCESS.2020.3015587
[3] liu, Qinghui & Kampffmeyer, Michael & Jenssen, Robert & Salberg, Arnt-Børre. (2019). Dense Dilated Convolutions Merging Network for Semantic
Mapping of Remote Sensing Images. 1-4. 10.1109/JURSE.2019.8809046.
[4] S. Ji, D. Wang and M. Luo, "Generative Adversarial Network-Based Full-Space Domain Adaptation for Land Cover Classification From Multiple-Source Remote Sensing Images," in *IEEE Transactions on Geoscience and Remote Sensing*,
vol. 59, no. 5, pp. 3816-3828, May 2021, doi: 10.1109/TGRS.2020.3020804.
[5] R. Zuo, G. Zhang, R. Zhang, and X. Jia, "A Deformable Attention Network for High-Resolution Remote Sensing Images Semantic Segmentation," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-14, 2022, Art no. 4406314, doi: 10.1109/TGRS.2021.3119537.
[6] Schmitt, Michael, Jonathan Prexl, Patrick Ebel, Lukas Liebel, and Xiao Xiang Zhu. "Weakly supervised semantic segmentation of satellite images for land cover mapping--challenges and opportunities." arXiv preprint arXiv:2002.08254 (2020).
[7] Pascual, Guillem, Santi Seguí, and Jordi Vitria. "Uncertainty gated network for land cover segmentation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.
[8] Scheibenreif, Linus, Joëlle Hanna, Michael Mommert, and Damian Borth. "Self-supervised vision transformers for land-cover segmentation and classification." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1422-1431. 2022.
[9] Li, Xiao, Lin Lei, Yuli Sun, Ming Li, and Gangyao Kuang. "Multimodal bilinear fusion network with second-order attention-based channel selection for land cover classification." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13 (2020): 1011-1026.
[10] Liu, Shengjie, Zhixin Qi, Xia Li, and Anthony Gar-On Yeh. "Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and SAR data." *Remote Sensing* 11, no. 6 (2019): 690.
[11] Onim, Md Saif Hassan, Aiman Rafeed Bin Ehtesham, Amreen Anbar, AKM Nazrul Islam, and AKM Mahbubur Rahman. "LULC classification by semantic segmentation of satellite images using FastFCN." In *2020 2nd International Conference on Advanced Information and Communication Technology (ICAICT)*, pp. 471-475. IEEE, 2020.
[12] Wang, J., Zheng, Z., Ma, A., Lu, X., Zhong, Y. LoveDA: A Remote Sensing Land-Cover Dataset for Domain Adaptive Semantic Segmentation. arXiv 2021. arXiv preprint arXiv:2110.08733.
[13] Xu, L., Liu, Y., Shi, S., Zhang, H., Wang, D. (2022). Land-Cover Classification with High-Resolution Remote Sensing Images Using Interactive Segmentation. IEEE Access.
[14] S. Ji, D. Wang, and M. Luo, "Generative Adversarial Network-Based Full-Space Domain Adaptation for Land Cover Classification From Multiple-Source Remote Sensing Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 59, no. 5, pp. 3816-3828, May 2021, doi: 10.1109/TGRS.2020.3020804.

[15] Sravya, N., Lal, S., Nalini, J., Reddy, C. S., Dell'Acqua, F. (2022). DPPNet: An Efficient and Robust Deep Learning Network for Land Cover Segmentation From High-Resolution Satellite Images. IEEE Transactions on Emerging Topics in Computational Intelligence.
[16] Khan, Sultan Daud, Louai Alarabi, and Saleh Basalamah. "Deep Hybrid Network for Land Cover Semantic Segmentation in High-Spatial Resolution Satellite Images." Information 12.6 (2021): 230
[17]Sun, Hao, Xiangtao Zheng, and Xiaoqiang Lu. "A supervised segmentation network for hyperspectral image classification." *IEEE Transactions on Image Processing* 30 (2021): 2810-2825.
[18]Wang, Qixiong, et al. "CCENet: Cascade Class-Aware Enhanced Network for High-Resolution Aerial Imagery Semantic Segmentation." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 15 (2022): 6943-6956.
[19]Zheng, Chen, et al. "An MRF-based multi-granularity edge-preservation optimization for semantic segmentation of remote sensing images." *IEEE Geoscience and Remote Sensing Letters* 19 (2021): 1-5.
[20]He, Shuang, et al. "RSI-Net: Two-Stream Deep Neural Network for Remote Sensing Images-Based Semantic Segmentation." *IEEE Access* 10 (2022): 34858-34871.
[21] D. Feng, Z. Zhang, and K. Yan, "A Semantic Segmentation Method for Remote Sensing Images Based on the Swin Transformer Fusion Gabor Filter," in IEEE Access, vol. 10, pp. 77432-77451, 2022, doi: 10.1109/ACCESS.2022.3193248.
[22] https://arxiv.org/abs/1612.01105
[23] https://towardsdatascience.com/understanding-semantic-segmentation-with-unet-6be4f42d4b47