

*A Comparative Analysis of CNN-based Deep Learning Models
for Semantic Segmentation of PolSAR data*

*Сравнительный анализ моделей глубокого обучения на основе CNN
для семантической сегментации данных PolSAR*

基於 CNN 的 PolSAR 資料語意分割深度學習模型的比較分析

Anil Kumar¹, Rajat Garg², Manish Prateek³, Shashi Kumar⁴

¹*School of Computer Science, University of Petroleum and Energy Studies, Dehradun*

²*Department of Computer Science and Engineering, IILM University, Greater Noida G.D.*

³*Goenka University, Panipat, India*

⁴*Indian Institute of Remote Sensing, ISRO, Dehradun*

The advent of Deep Learning models for computer vision problems has opened the floodgates of opportunities. The same deep learning models have been highly effective in solving many challenges in the field of remote sensing and geoinformatics[1]. Earth's ecology is undergoing tremendous changes every day and this mandates the continuous monitoring of the Earth's ecosystem. It is an important task since it allows the government and other responsible bodies to take a planned set of actions that allow them to overcome the challenges being faced due to climate change and the natural hazards arising out of it. Remote sensing data is highly effective in monitoring the Earth's ecology by providing accurate Land Use Land Cover maps through high-resolution data. One of the major challenges faced by the remote sensing research community is the unavailability of open-source high-resolution PolSAR data with ground truth. This work aims at comparing the semantic segmentation results of multiple deep learning models. The results are also compared by varying the backbone network used in the different deep learning models to get a clear understanding of which combination works the best and why. Similar work by Yuan et. al. [2] shows that the DenseNet169 performs the best. The initial comparative results between FPN[3], UNET[4], and DeeplabV3+[5] on the L band Uninhabited Ariel Vehicle Synthetic Aperture Radar (UAVSAR) dataset of the Houston region in the USA show that the DeeplabV3+ outperforms the FPN and UNET models with Xception as a backbone. However, it is noteworthy that different models are sensitive to different classes in the PolSAR imagery.

As it can be observed from figure 1, the regions enclosed by black boxes A and B highlight that the DeeplabV3+ (figure 1e) correctly distinguishes the field from the forest while both FPN (figure 1c) and UNET (figure 1d) wrongly identify the field as forest region.

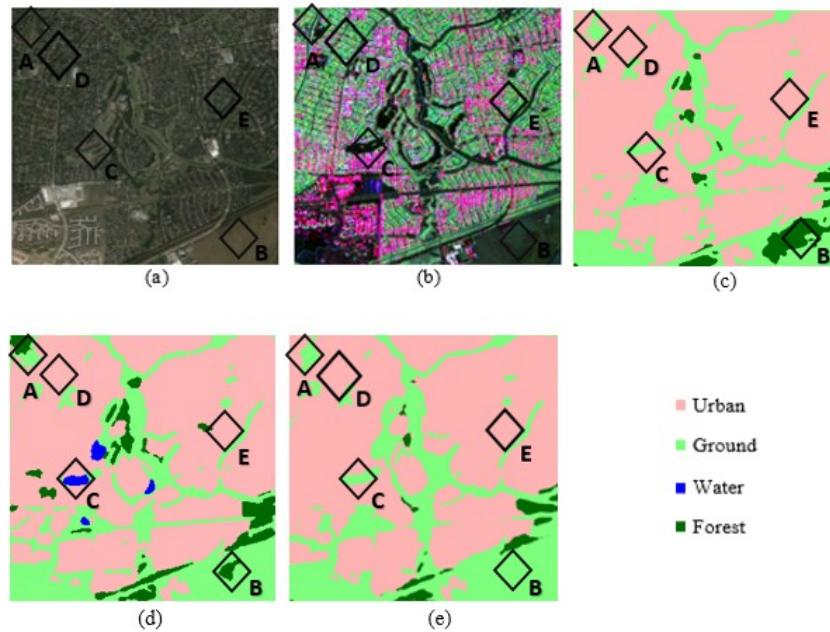


Fig. 1 (a) Google Earth image of test patch for Ground Truth (b) False color composite of UAVSAR image of a part of Houston City, Texas (USA) and the corresponding image segmentation results for (c) FPN, (d) UNET, and (e) DeepLabv3+ respectively.

Рис. 1 (а) Изображение Google Earth тестового участка для Ground Truth (б) Композиция в искусственных цветах изображения БПЛА части Хьюстон-Сити, штат Техас (США) и соответствующие результаты сегментации изображения для (в) FPN, (г) UNET и (е) DeepLabv3+ соответственно.

圖 1 (a) Ground Truth 測試補丁的 Google Earth 影像 (b) 美國德州休士頓市部分 UAVSAR 影像的偽彩色合成以及 (c) FPN 的相應影像分割結果, (d) 分別為 UNET 和 (e) DeepLabv3+。

However, in the case of the water class, UNET has correctly identified the water as bounded by the black box C whereas FPN and DeeplabV3+ failed to identify the same. It is also to be observed that all the models have correctly classified the oriented urban targets bounded by boxes D and E, which appear in green color in the decomposed false-color composite image due to the diffused type of scattering coming from the oriented urban targets as observed in figure 1b. The overall accuracy of FPN, UNET, and DeeplabV3+ is 84.62%, 86.52%, and 87.78% respectively. This accuracy can be further enhanced by varying the backbone network and also by training the models from scratch instead of relying on the transfer learning method. The only problem with training a model from scratch is that too large a dataset is required and the creation of a dataset is a very time-consuming and laborious task that requires domain experts. Generally, the improvement in the accuracy achieved by deploying extra computational power is not worth the time and resources consumed and it is better to go with the transfer learning approach.

Появление моделей глубокого обучения для решения задач компьютерного зрения открыло шлюзы возможностей. Одни и те же модели глубокого обучения оказались весьма эффективными при решении многих задач в области дистанционного зондирования и геоинформатики [1]. Экология Земли претерпевает огромные изменения каждый день, и это требует постоянного мониторинга экосистемы Земли. Это важная задача, поскольку она позволяет правительству и другим ответственным органам предпринять запланированный набор действий, которые позволят им преодолеть проблемы, с которыми сталкиваются из-за изменения климата и связанных с ним стихийных бедствий. Данные дистанционного зондирования очень эффективны для мониторинга экологии Земли, поскольку предоставляют точные карты землепользования и растительного покрова на основе данных высокого разрешения. Одной из основных проблем, с которыми сталкивается исследовательское сообщество в области дистанционного зондирования, является недоступность открытых данных PolSAR с высоким разрешением и достоверной информацией. Целью этой работы является сравнение результатов семантической сегментации нескольких моделей глубокого обучения. Результаты также сравниваются путем изменения магистральной сети, используемой в различных моделях глубокого обучения, чтобы получить четкое представление о том, какая комбинация работает лучше всего и почему. Аналогичная работа Yuan et. al. [2] показывает, что DenseNet169 работает лучше всего. Первоначальные сравнительные результаты между FPN[3], UNET[4] и DeeplabV3+[5] в наборе данных необитаемого радара Ariel с синтезированной апертурой (UAVSAR) L-диапазона в регионе Хьюстон в США показывают, что DeeplabV3+ превосходит FPN и UNET. модели с Xception в качестве основы. Однако примечательно, что разные модели чувствительны к разным классам изображений PolSAR.

Как видно из рисунка 1, области, заключенные в черные прямоугольники A и B, подчеркивают, что DeeplabV3+ (рисунок 1e) правильно отличает поле от леса, в то время как FPN (рисунок 1c) и UNET (рисунок 1d) ошибочно идентифицируют поле. как лесной регион. Однако в случае с классом воды UNET правильно определил, что вода ограничена черным ящиком C, тогда как FPN и DeeplabV3+ не смогли это идентифицировать. Также следует отметить, что все модели правильно классифицировали ориентированные городские цели, ограниченные прямоугольниками D и E, которые отображаются зеленым цветом на разложенном составном изображении в искусственных цветах из-за диффузного типа рассеяния, исходящего от ориентированных городских целей. как показано на рисунке 1b. Общая точность FPN, UNET и DeeplabV3+ составляет 84,62%, 86,52% и 87,78% соответственно. Эту точность можно еще больше повысить, изменив магистральную сеть, а также обучив модели с нуля вместо того, чтобы полагаться на метод трансферного обучения. Единственная проблема с обучением модели с нуля заключается в том, что требуется слишком

большой набор данных, а создание набора данных — очень трудоемкая и трудоемкая задача, требующая экспертов в предметной области. Как правило, повышение точности, достигаемое за счет использования дополнительных вычислительных мощностей, не стоит затраченных времени и ресурсов, и лучше использовать подход трансферного обучения.

針對電腦視覺問題的深度學習模型的出現開啟了機會之門。同樣的深度學習模式在解決遙感和地理資訊學領域的許多挑戰方面非常有效[1]。地球的生態每天都在發生巨大的變化，這要求對地球生態系統進行持續監測。這是一項重要的任務，因為它使政府和其他負責機構能夠採取一系列有計劃的行動使他們能夠克服氣候變遷及其引起的自然災害所面臨的挑戰。遙感資料透過高解析度資料提供準確的土地利用土地覆蓋圖，在監測地球生態方面非常有效

遙感研究界面臨的主要挑戰之一是無法取得具有地面實況的開源高解析度 PolSAR 資料。這項工作旨在比較多個深度學習模型的語義分割結果。也透過改變不同深度學習模型中使用的主幹網路來比較結果，以便清楚了解哪種組合效果最好以及原因。袁等人的類似工作。等人。[2]顯示 DenseNet169 表現最佳。FPN[3]、UNET[4]、DeeplabV3+[5]在美國休士頓地區 L 波段無人機合成孔徑雷達（UAVSAR）資料集上的初步對比結果表明，DeeplabV3+ 優於 FPN 和 UNET 以 Xception 作為骨幹的模型。然而，值得注意的是，不同的模型對 PolSAR 影像中的不同類別敏感。

從圖 1 可以看出，黑框 A 和 B 包圍的區域突出表明 DeeplabV3+（圖 1e）正確區分了田野和森林，而 FPN（圖 1c）和 UNET（圖 1d）都錯誤地識別了田野作為林區。然而，就水類而言，UNET 已正確識別出以黑盒 C 為界的水，而 FPN 和 DeeplabV3+ 未能識別出相同的水。還可以觀察到，所有模型都正確分類了由框 D 和 E 包圍的定向城市目標，由於來自定向城市目標的散射類型，這些目標在分解的假色合成圖像中顯示為綠色如圖 1b 所示。FPN、UNET 和 DeeplabV3+ 的整體準確率分別為 84.62%、86.52% 和 87.78%。透過改變主幹網路以及從頭開始訓練模型而不是依賴遷移學習方法，可以進一步提高這種準確性。從頭開始訓練模型的唯一問題是需要太大的資料集，而資料集的建立是一項非常耗時費力的任務，需要領域專家。一般來說，透過部署額外的運算能力來提高準確性並不值得消耗時間和資源，最好採用遷移學習方法。

Литература

References

參考書目

1. Xin Zhang et. al., “How well do deep learning based methods for land cover classification and object detection perform on high resolution remote sensing imagery?,” Remote Sensing, vol. 12(3), 2020, <https://doi.org/10.3390/rs12030417>

2. J. Yuan, L. Ru, S. Wang and C. Wu, “WH-MAVS: A Novel Dataset and Deep Learning Benchmark for Multiple Land Use and Land Cover Applications,” in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 1575-1590, 2022, <https://doi.org/10.1109/JSTARS.2022.3142898>
3. Tsung-Yi Lin et. al., “Feature Pyramid Networks for Object Detection”, CVPR, 17 April 2017, <https://doi.org/10.48550/arXiv.1612.03144>
4. Olaf Ronneberger et. al., “U-Net: Convolutional Networks for Biomedical Image Segmentation, CVPR, 18 May 2015, <https://doi.org/10.48550/arXiv.1505.04597>
5. L. C. Chen et. al., “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation”, CVPR, 22 Aug 2018, <https://doi.org/10.48550/arXiv.1802.02611>