GRADIVA

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

Aerial Image Segmentation Using Deep Learning

Abdelghani ROUINI [1] & Messaouda LARBI [2]

1. Department of Science and Technology, Ziane Achour University of Djelfa, Algeria.
1, 2. Applied Automation and Industrial Diagnostic Laboratory, Ziane Achour University of Djelfa, Algeria.
2. Department of Computer Science, Ziane Achour University of Djelfa, Algeria.

Abstract

Image segmentation of aerial images using deep learning has gained significant attention due to its potential for extracting valuable information from high-resolution imagery. This study focuses on the application of deep learning techniques for image segmentation in the context of aerial images. Specifically, popular architectures such as Unet, PSPNet, and LinkNet are utilized with different Feature Extraction Networks including EfficientNet-B4 and ResNet50. The models are trained and evaluated on aerial imagery of Dubai obtained by MBRSC satellites dataset, and the results are assessed using Intersection over Union (IoU) metrics for training and validation sets. The findings reveal that Unet with EfficientNet-B4 achieves the highest IoU scores, with a training IoU of 0.65708 and a validation IoU of 0.64002. PSPNet and LinkNet also demonstrate competitive performance, with EfficientNet-B4 as the preferred Feature Extraction network. These results highlight the effectiveness of deep learning approaches for aerial image segmentation and provide valuable insights for selecting suitable models and architectures for this task.

Keywords: *Image Segmentation, Deep Learning, Aerial imagery, U-Net, PSPNet, LinkNet.*

1. INTRODUCTION

Image segmentation plays a vital role in various computer vision tasks, such as object recognition, scene understanding, and medical image analysis. It involves partitioning an image into distinct regions or segments, each corresponding to a meaningful object or region of interest. Accurate image segmentation is a challenging problem due to the complex nature of images and the variability in object shapes, sizes, and appearances.[1]

Traditional image segmentation approaches heavily rely on handcrafted features and heuristic algorithms, which often struggle to handle diverse and complex scenarios. However, with the advancements in deep learning, specifically convolutional neural networks (CNNs), image segmentation has witnessed remarkable progress in recent years.

In the ever evolving landscape of technology, significant advancements have been made in the field of artificial intelligence (AI), reshaping our world and revolutionizing various domains. One such remarkable application of AI is in the domain of computer vision, which has rapidly gained importance in recent years. [2-4]

Computer vision allows machines to perceive and interpret visual data, like to how humans process images and videos. Through the fusion of AI and computer vision, a wide range of ground-breaking applications have emerged, permeating diverse industries such as healthcare, transportation, entertainment, and more.

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

2. MATERIALS

a. Presentation of the datasets

We utilized aerial imagery of Dubai obtained by MBRSC satellites datasets, the dataset is publicly available at Dataset.[5]

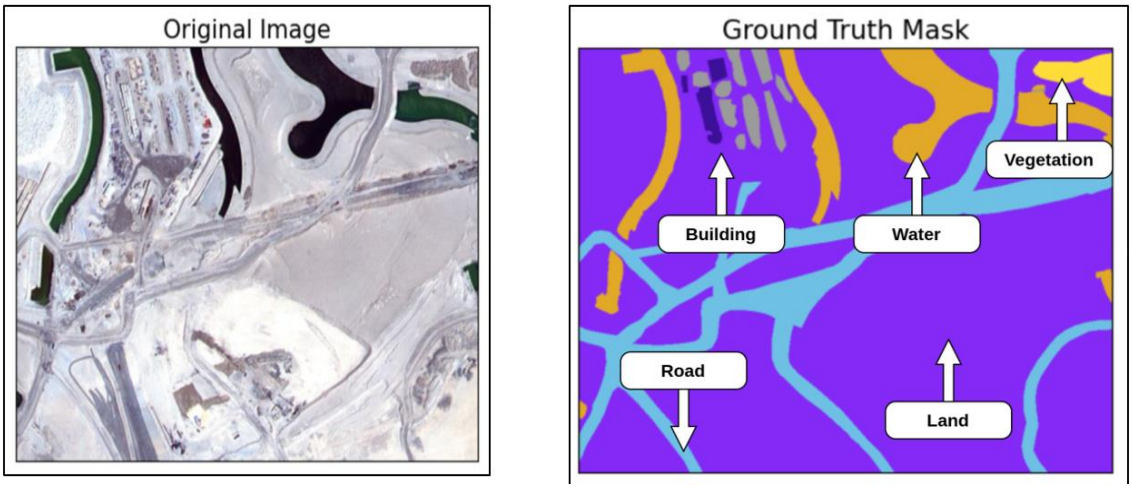

Figure 1: Example from aerial imagery datasets

The dataset consists of aerial imagery of Dubai obtained by MBRSC satellites and annotated with pixel-wise semantic segmentation in 6 classes. The total volume of the dataset is 72 images grouped into 6 larger tiles. The classes are:

1. Building : #3C1098
2. Land : #8429F6
3. Road : #6EC1E4
4. Vegetation : #FEDD3A
5. Water : #E2A929
6. Unlabeled : #9B9B9B

b. Data Pre-processing

Images must be the same size when fed into the neural network input layer. Therefore, before model training we will;

- Split images into patches, the patch size chosen is 160 px. (Divisible by 32)
- Split of the datasets into three sets [train_set /validation_set/test_set] 80% for training, 10% for validation and 10% for testing.

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

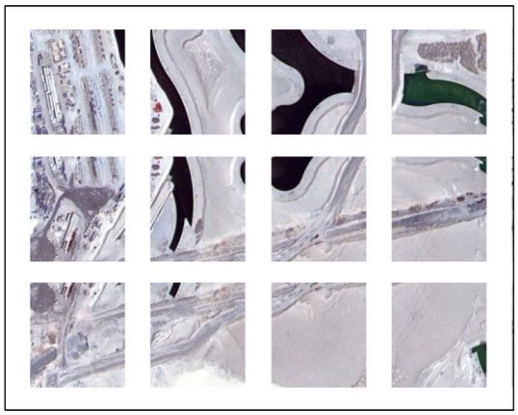

Figure 2: Patches from splinted image

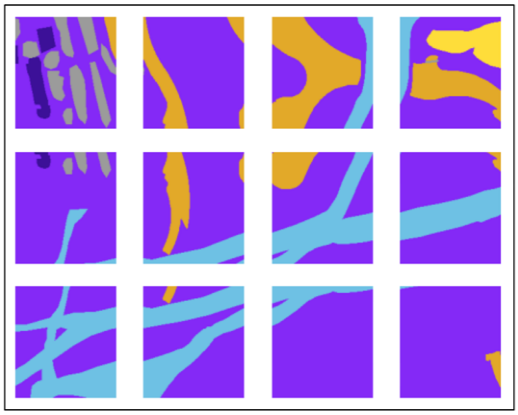

Figure 3: Patches from splinted mask

C. Presentation of the CNN pretrained backbons that we used

During our experiments we used two backbones are:

- Backbons 1: RasNet50.
- Backbons 2: EfficientNet-b4

D. Optimizer ADAM (Adaptive Moment Estimation)

Adam is an optimization algorithm that can be used instead of the classical stochastic gradient descent procedure to update network weights iterative based in training data.

3. THE PROPOSED METHODS

In this section, we will explain the subject on which we worked, the different hardware and software resources that we used, the various experiments that we carried out, we will compare the performance of multiple Deep learning algorithms in the context of Aerial Imagery Segmentation and finally, a discussion of the results of the evaluation acquired.

We used CNN architectures to Segmentation of Aerial Imagery, We will use two pre trained networks for feature extraction (ResNet, EfficientNet-B4) and three image segmentation architectures:

- U-Net [6-8]
- PSPNet [9-10]
- Linknet [11-12]

Configuration used in the implementation:

Computer:

- **Processor:** Intel(R) Xeon(R) CPU @ 2.30GHz with 2 vCPUs.
- **RAM:** 13GB of RAM.
- **Graphics Card:** NVIDIA Tesla T4 GPU with 16GB of VRAM.
- **Hard disk:** 100 GB.
- **Operating system:** Ubuntu 20.04.6 LTS.

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

Programming language

Python was chosen as the programming language for this study due to several key reasons. Firstly, Python is widely recognized as one of the most popular programming languages in the field of data science and machine learning. Its extensive community support and active development ecosystem make it a reliable choice for implementing complex algorithms and models.

Python is a high-level, interpreted programming language known for its simplicity and readability. It was created by Guido van Rossum and first released in 1991. Python emphasizes code readability and uses whitespace indentation to delimit code blocks instead of relying on braces or keywords. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming.

An important advantage of using Python for deep learning is the availability of cloud-based platforms such as Google Colab. These platforms provide free access to high-performance GPUs, which are crucial for training deep learning models on large-scale image datasets. The ability to leverage the power of cloud computing significantly speeds up the training process, enabling faster iterations and experimentation.

4. RESULTS AND DISCUTIONS

A. The training of our models

We trained all models for 30 Epochs with a change each time the backbone.

a. U-Net

1. First with EfficientNet-B4 as feature extractor:

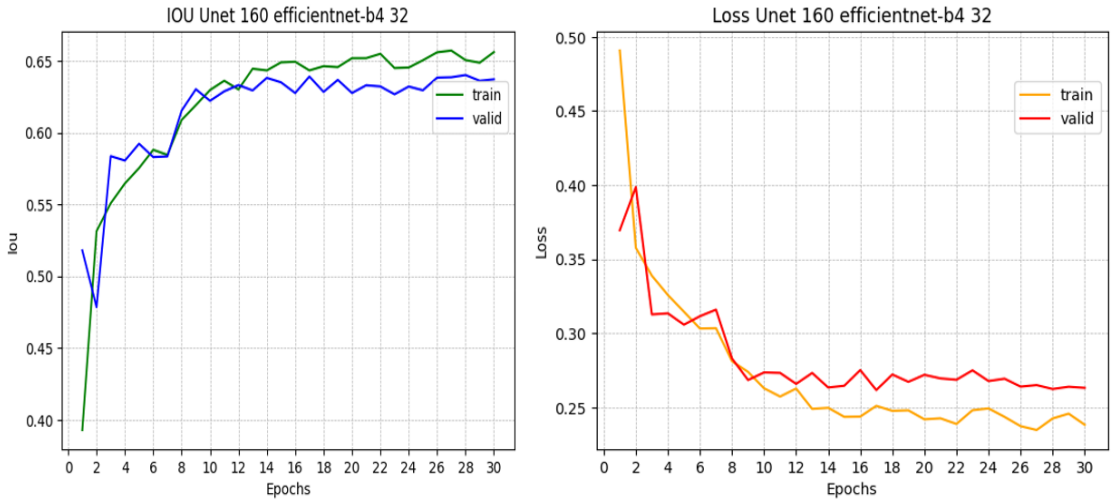

Figure 4: U-net Training with EfficientNet-B4 as feature extractor (A) IOU score achieved for train and validation sets vs epoch number during training.(B) Loss on training and validation sets vs epoch number during training

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

2. Second with RasNet50 as feature extractor:

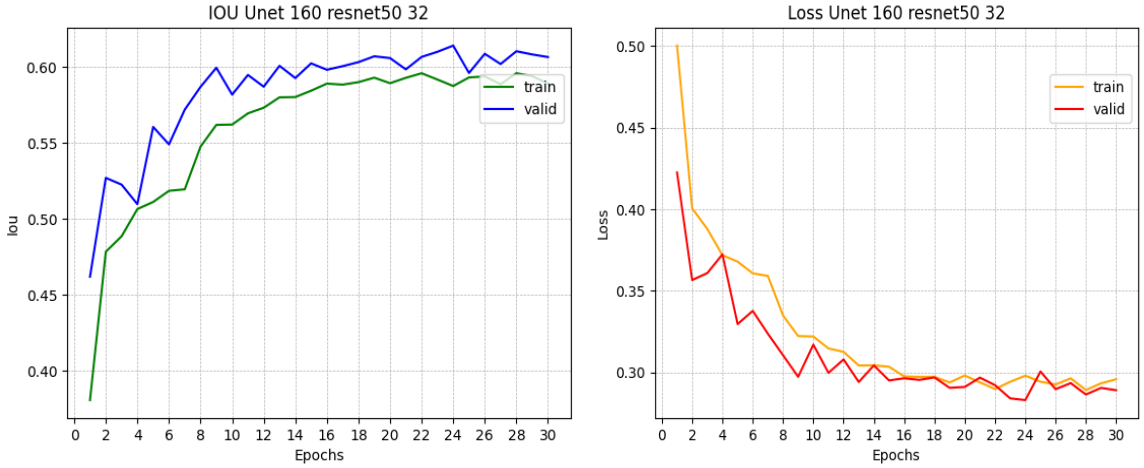

Figure 5: U-net Training with ResNet50 as feature extractor (A) IOU score achieved for train and validation set vs epoch number during training. (B) Loss on training and validation sets vs epoch number during training

b. PSPNet

1. First with EfficientNet-B4 as feature extractor:

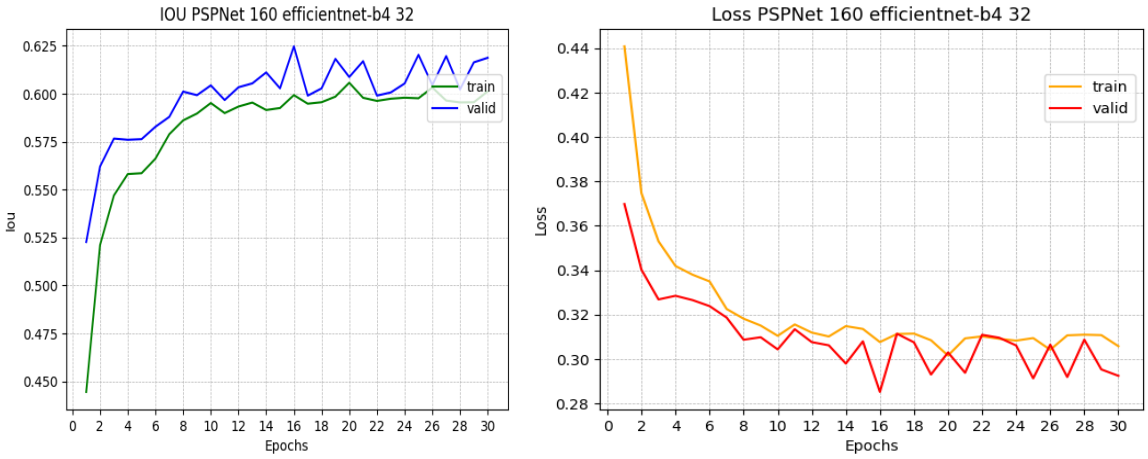

Figure 6: PSPnet Training with EfficientNet-B4 as feature extractor (A) IOU score achieved for train and validation sets vs epoch number during training. (B) Loss on training and validation sets vs epoch number during training

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

2. Second with RasNet50 as feature extractor:

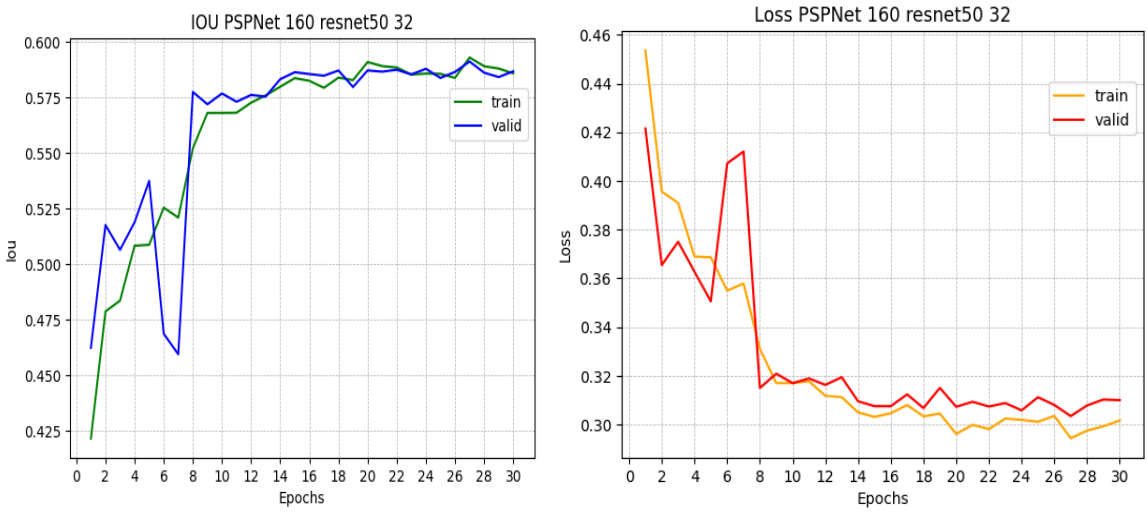

Figure 7: PSPnet Training with ResNet50 as feature extractor (A) IOU score achieved for train and validation sets vs epoch number during training. (B) Loss on training and validation sets vs epoch number during training

c. Linknet

1. First with EfficientNet-B4 as feature extractor:

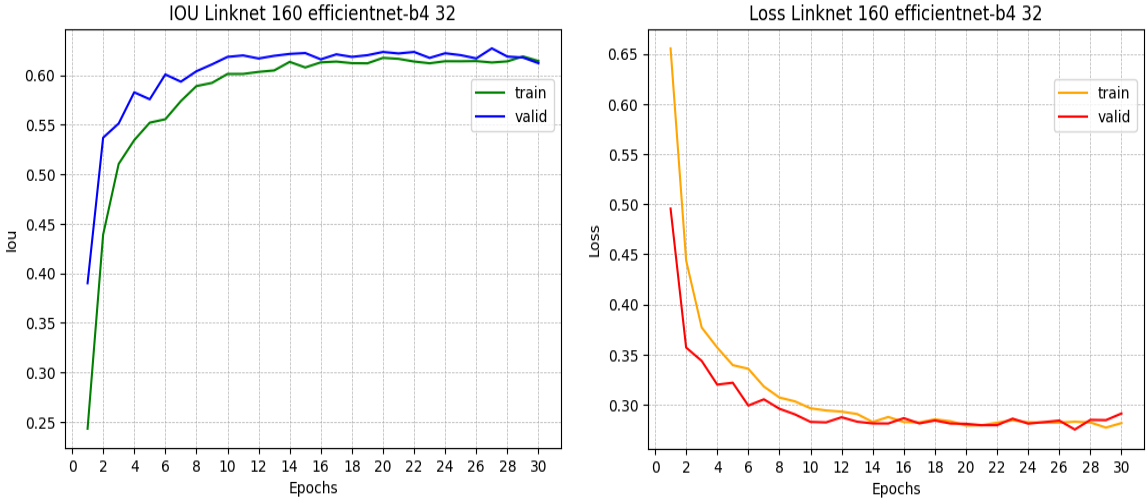

Figure 8: Linknet Training with EfficientNet-B4 as feature extractor (A) IOU score achieved for train and validation sets vs epoch number during training. (B) Loss on training and validation sets vs epoch number during training

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

2. Second with RasNet50 as feature extractor:

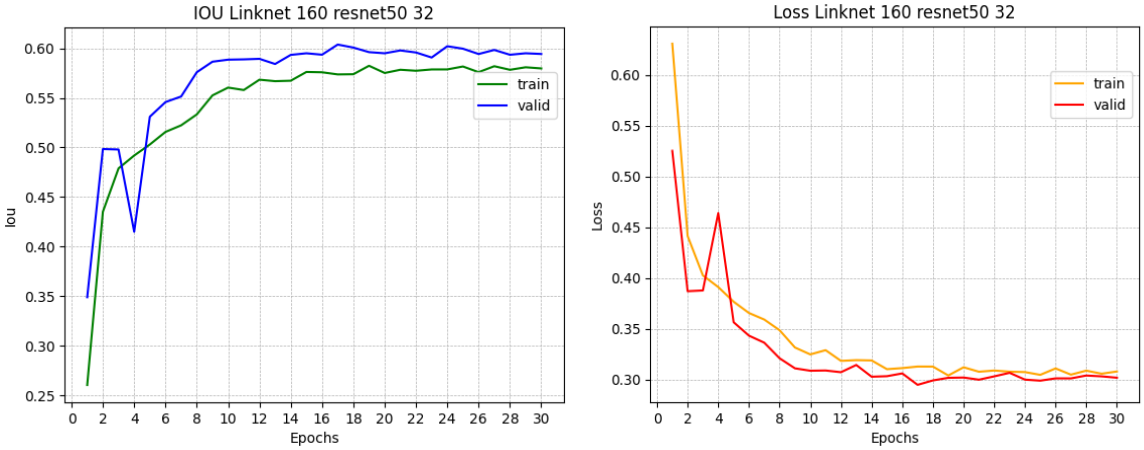

Figure 9: Linknet Training with ResNet50 as feature extractor (A) IOU score achieved for train and validation sets vs epoch number during training. (B) Loss on training and validation sets vs epoch number during training

B. The performance of our models

The Table 1 presents the results of evaluating different models for image segmentation using two different backbones: EfficientNet-B4 and ResNet50. The performance of each model is measured in terms of IoU on both the training and validation datasets.

Table 1: Result

Model	Backbon	Train IoU	Validation IoU
Unet	EfficientNet-B4	0.65708	0.64002
	ResNet50	0.59594	0.61410
PSPNet	EfficientNet-B4	0.60565	0.62464
	ResNet50	0.59287	0.59106
LinkNet	EfficientNet-B4	0.61888	0.62704
	ResNet50	0.58243	0.60382

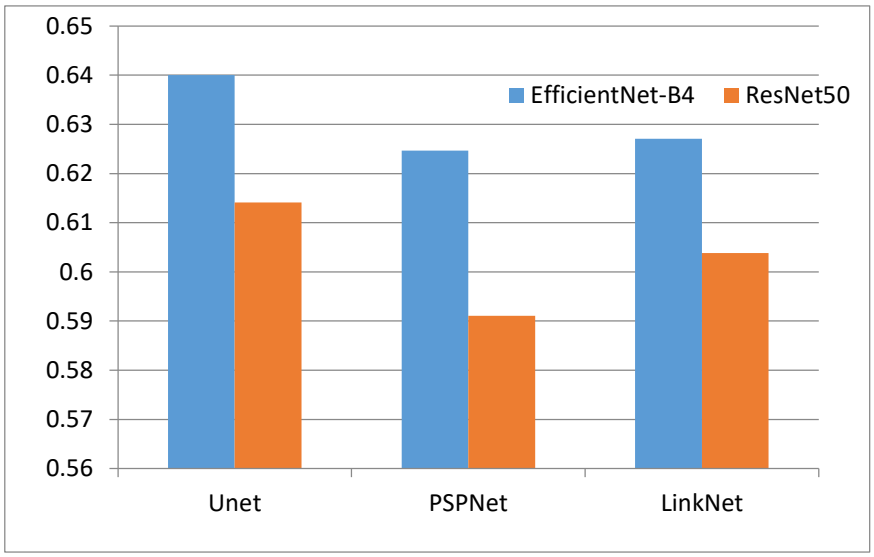

Figure 10: Validation results (IoU)

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

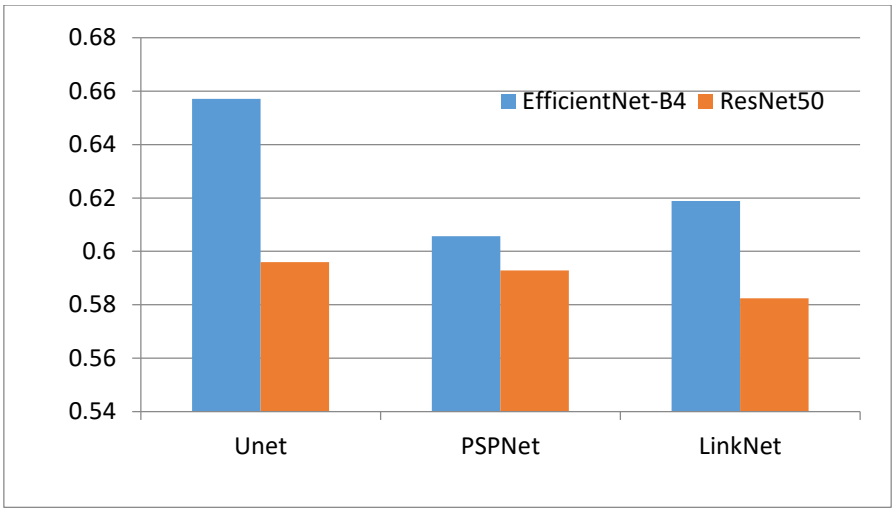

Figure 11: Training results (IoU)

Among the models evaluated, Unet with EfficientNet-B4 backbone achieved the highest IoU scores, with a training IoU of 0.65708 and a validation IoU of 0.64002. Unet with ResNet50 backbone also performed well, although slightly lower, with a training IoU of 0.59594 and a validation IoU of 0.61410.

PSPNet with EfficientNet-B4 backbone obtained a training IoU of 0.60565 and a validation IoU of 0.62464, demonstrating competitive performance. On the other hand, PSPNet with ResNet50 backbone yielded lower IoU scores, with a training IoU of 0.59287 and a validation IoU of 0.59106.

For LinkNet, the results showed a training IoU of 0.61888 and a validation IoU of 0.62704 when using the EfficientNet-B4 backbone. LinkNet with ResNet50 backbone achieved slightly lower scores, with a training IoU of 0.58243 and a validation IoU of 0.60382.

Based on these results, it can be observed that models utilizing the EfficientNet-B4 backbone generally outperformed those using the ResNet50 backbone. This suggests that the EfficientNet-B4 architecture captures more informative features for the image segmentation of aerial (satellite) imagery.

Additionally, the comparison between the different models indicates that Unet consistently produced the best results in terms of IoU, regardless of the backbone used. This highlights the effectiveness of the Unet architecture for image segmentation tasks.
Overall, these results provide insights into the performance of various models with different backbones for aerial image segmentation. The findings can guide further investigations and optimizations to achieve even better segmentation accuracy and generalization capabilities.

5. CONCLUSION

In conclusion, this study focused on investigating the feasibility and effectiveness of using deep learning architectures for satellite image segmentation. The objectives were to assess the performance of different deep learning architectures (Unet, PSPNet, and LinkNet) when utilizing transfer learning with two different backbone networks (EfficientNet-B4 and ResNet50).

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

The main findings and contributions of this thesis are summarized as follows:

In this work we train and compare multiple deep learning architectures using transfer learning techniques. The results revealed variations in performance across the architectures and backbone networks. Notably, the Unet architecture with EfficientNet-B4 backbone achieved the highest validation IoU of 0.64002, showcasing its effectiveness in satellite image segmentation tasks when combined with transfer learning.

The methodology employed in this study involved collecting satellite image data and preprocessing it into 160x160 pixel patches. Transfer learning was then applied by utilizing pre-trained models as the starting point and fine-tuning them on the specific satellite image segmentation task. This approach allowed for efficient training and evaluation of the models, taking advantage of the learned features from larger, general-purpose datasets.

The study focused on a specific set of deep learning architectures and backbone networks, and the evaluation was based on a specific dataset. Further investigations using different architectures, backbones, and datasets would enrich the understanding of transfer learning in satellite image segmentation.

Future research directions should explore advanced techniques to further improve the segmentation accuracy and generalization of the models. This could include investigating different transfer learning strategies, incorporating data augmentation techniques, and exploring ensemble methods to boost performance.

In conclusion, this study has advanced the field of satellite image segmentation by demonstrating the feasibility and effectiveness of transfer learning with deep learning architectures

Reference

1) Long Jonathan, Evan Shelhamer and Trevor Darrell, "Fully convolutional networks for semantic segmentation", *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431-3440, 2015.

2) D. Guo, A. Weeks and H. Klee, "Robust approach for suburban road segmentation in high-resolution aerial images", *International Journal of Remote Sensing*, vol. 28, no. 2, pp. 307-318, 2007.

3) Luo Haifeng, Chongcheng Chen, Lina Fang, Xi Zhu and Lijing Lu, "High-resolution aerial images semantic segmentation using deep fully convolutional network with channel attention mechanism", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3492-3507, 2019.

4) Li Xiang, Yuchen Jiang, Hu Peng and Shen Yin, "An aerial image segmentation approach based on enhanced multi-scale convolutional neural network", *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, pp. 47-52, 2019.

5) Datasource:https://humansintheloop.org/resources/datasets/semantic-segmentation-dataset-2/

6) Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv preprint arXiv:1505.04597.

Volume 63 | Issue 01 | January 2024
DOI: 10.5281/zenodo.10669323

7) Jun Hee Kim, Haeyun Lee, Seonghwan J. Hong, Sewoong Kim, Juhum Park, Jae Youn Hwang, et al., "Objects segmentation from high-resolution aerial images using U-Net with pyramid pooling layers", *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 115-119, 2018.

8) Cai, Y., & Wang, Y. (2020). MA-Unet: An improved version of Unet based on multi-scale and attention mechanism for medical image segmentation. arXiv preprint arXiv:2012.10952.

9) Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid Scene Parsing Network. arXiv preprint arXiv:1612.01105.

10) H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, ‘‘Pyramid scene parsing network,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2881–2890.

11) Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. arXiv 2015, arXiv:1411.4038

12) Chaurasia, A., & Culurciello, E. (2017). LinkNet: Exploiting encoder representations for efficient semantic segmentation. In 2017 IEEE Visual Communications and Image Processing (VCIP) (pp. 1-4). IEEE. doi: 10.1109/vcip.2017.8305148.