Deep Learning-based Road Segmentation Using Aerial Images: A Comparative Study

Kamal KC[1], Alaka Acharya[2], Kushal Devkota[3], Kalyan Singh Karki[4], and Surendra Shrestha [5,*]

[1,2,3]*College of Biomedical Engineering and Applied Sciences, Purbanchal University, Kathmandu, Nepal*
[4]*School of Life Science and Technology, Harbin Institute of Technology, Harbin 150001, China*
[5,*]*Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Lalitpur, Nepal*

kamal@biomedical.edu.np [1], *alakacharya@biomedical.edu.np* [2], *kushal@biomedical.edu.np* [3],
kalyankarki2019@hit.edu.cn [4], *surendra@ioe.edu.np* [5,*]

Abstract

Precise road segmentation i.e., extracting underlying road networks, using deep learning architectures continues to present a significant challenge. Numerous state-of-the-art (SOTA) models have been developed for the task of road detection through semantic segmentation. Nevertheless, there is a lack of comprehensive information regarding their comparative analysis, making it challenging to select the most effective model based on performance in this specific domain. Thus, we perform a comparative study of cutting-edge deep learning-based segmentation models in the context of road segmentation using RGB aerial imagery, named Massachussetts Roads Dataset. We utilize SOTA segmentation models U-Net and LinkNet combined with six backbone CNNs (VGG19, Resnet50, Resnet152, MobilenetV2, EfficientnetB0, and EfficientnetB7) to achieve this. The results show that U-Net paired with EfficientnetB7 achieved the highest performance, Intersection of Union, IoU (91.69%), diceloss (4.36%), F1-score (95.65%), accuracy (95.75%), recall (97.97%), and precision (93.43%).

Keywords: *Deep Convolutional Neural Network, Semantic Road Segmentation, Fine-tuning, U-Net, LinkNet.*

1. Introduction

Semantic segmentation in computer vision, which involves classifying pixels at a fine-grained level, has brought about significant advancements in various domains, including medical image analysis and precision agriculture [1,2]. Road segmentation, which helps to identify primary, secondary, and tertiary roads, is considered one of the major areas that directly contribute to improving geospatial analysis. The ability to isolate roads from aerial images comprising complex structures plays an important role in improved navigation systems, autonomous driving, traffic management, disaster mitigation, environmental monitoring, urban planning etc [3]. The shift from manual road extraction to automated extraction using image processing algorithms has decreased the latency and increased effectiveness depending upon the robustness of the computer vision methods. Over time, machine learning and deep learning algorithms have proven to be efficient for semantic segmentation tasks.

The performance of deep learning-based semantic solutions is highly influenced by both the quality and quantity of the data utilized in their training process [4]. With the advent of Convolutional Neural Networks (ConvNets) in deep learning, researchers have explored a range of ConvNet architectures to effectively extract roads from aerial images [5]. Among these architectures, encoder-decoder-based designs such as U-net, Feature Pyramid Network (FPN), PSPNet, Linknet, muti-branch ConvNets etc., are commonly employed because of their capacity to capture substantial spatial context. While these models perform well, their initial weights and bias limit their performances. Moreover, deep learning-based models require a humongous amount of data. To overcome both these problems transfer learning and fine-tuning of pre-trained deep models on an imagenet dataset is widely preferred [6]. A combination of state-of-the-art segmentation models with pre-trained weights results in higher performance.

2. Materials and Methods

2.1 Dataset

This road segmentation dataset, named Massachusetts Roads Dataset [7], which consists of RGB images, comprises 1171 aerial images of the state of Massachusetts. Each image is 1500×1500 pixels in size, covering an area of 2.25 square kilometres. This dataset is randomly split into a training set of 1108 images, a validation set of 14 images and a test set of 49 images. The dataset consists of RGB images and their segmented masks. The images were annotated at the pixel level into the road and background. The dataset covers a wide variety of urban, suburban, and rural regions and covers an area of over 2600 square kilometres. The test set alone covers over 110 square kilometres. The target maps were generated by rasterizing road centerlines obtained from the OpenStreetMap project. A line thickness of 7 pixels and no smoothing was used in generating the labels. All imagery is rescaled to a resolution of 1 pixel per square meter. Figure 1 shows a few data samples.

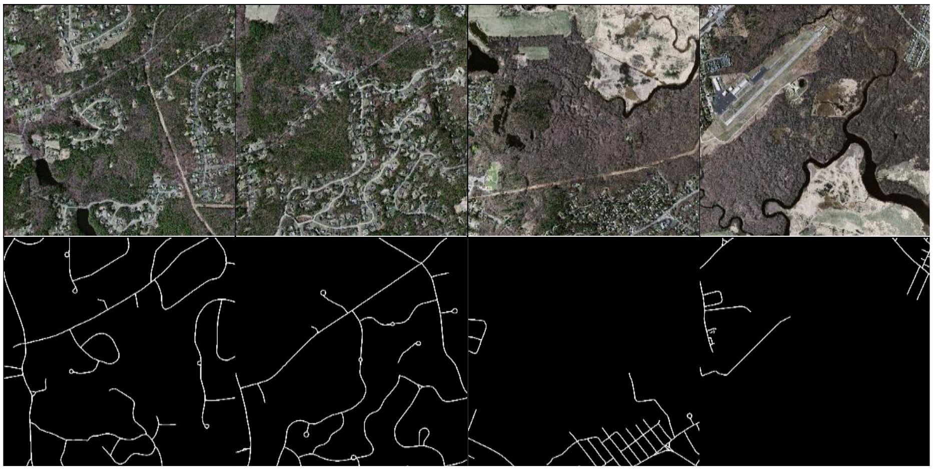

Figure 1. Few samples from the Massachusetts Roads Dataset. Top row images comprise RGB aerial images, and bottom row images are their corresponding segmented masks.

2.2 State-of-the-art Segmentation models

Among various segmentation models such as DeepLabV3+, SegNet, FPN, PSPNet, DLinkNet etc., U-net and LinkNet have been employed with VGG19, Resnet50, Resnet152, MobilenetV2, EfficientnetB0, and EfficientnetB7 backbones. Both the models are fully convolutional as shown in Figure 2. U-nets employ a sequence of downsampled convolutions, succeeded by a sequence of upsampled transpose convolutions. They incorporate residual connections connecting the downsampled and upsampled pathways. In contrast, Linknet is a faster network that follows a similar downsample and upsample procedure, with the difference of utilizing summation operations between the layers [8].

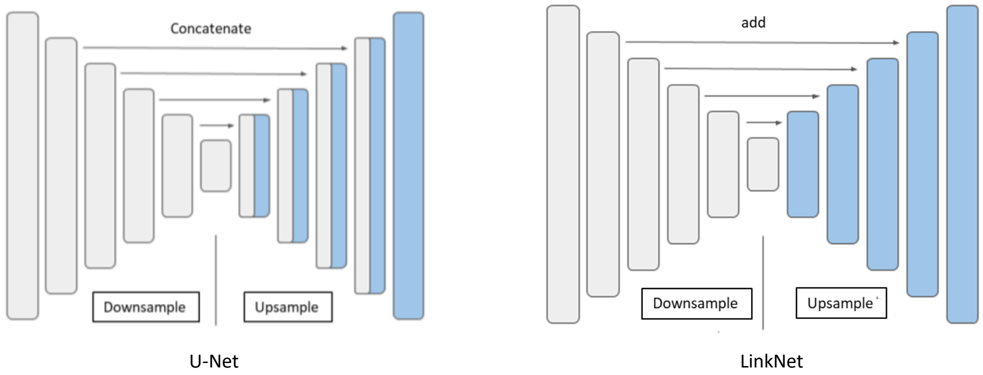

Figure 2. The skeleton of U-Net and LinkNet architectures. The depth and layers depend upon the backbone.

2.3 Experimental Setup

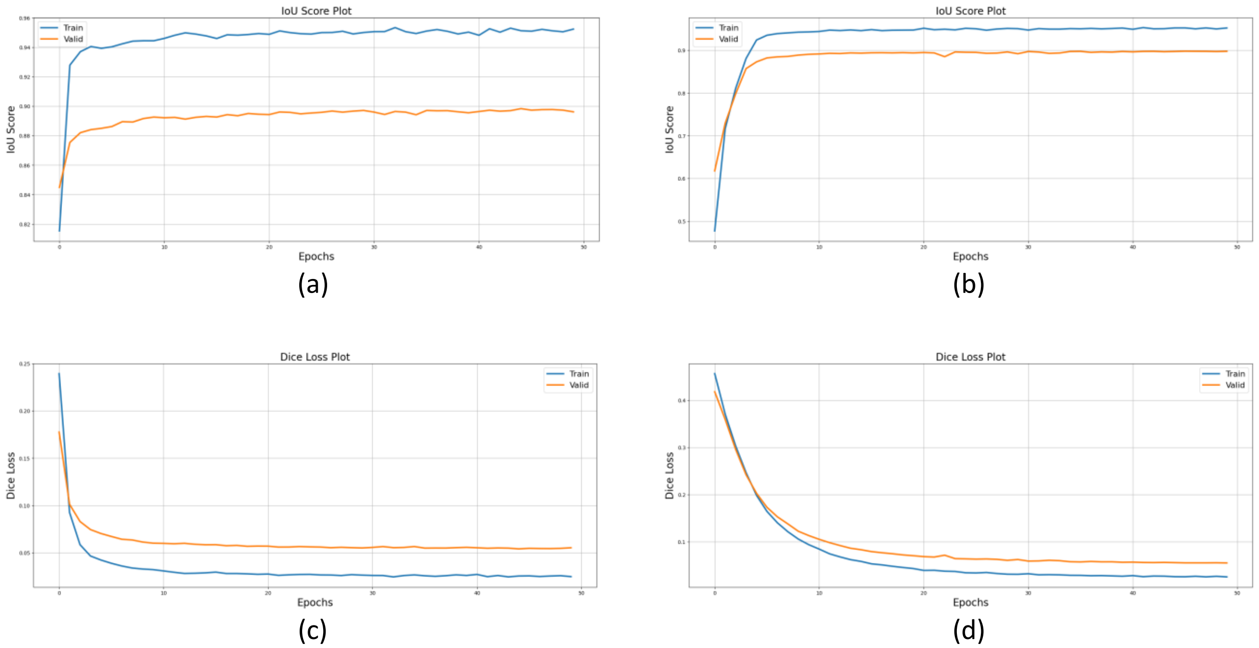

Figure 3. The training and validation curves for UNet and Linknet with EfficientNetB7 as a backbone. (a ,c) represent the model IOU score and Dice loss curves for Unet and (b,d) represent the model IOU score and Dice loss curves for Linknet.

The experimental setup to conduct the road segmentation was built on a Python-based Keras package [9]. All the experiments were carried out on Google Colab cloud computing platform which utilized an NVIDIA T4 GPU with 12 GB RAM. For maintaining consistency in comparison, each segmentation model was trained up to maximum epochs of 50, with a learning rate of 0.0001 and Adam optimizer. The Intersection of Union (IoU) score and f-score were used as evaluation functions while training the model. A sample plot of IoU score and Diceloss achieved per epoch while training the U-net model with EfficienetNetB7 and Linknet model with EfficientNetB7 as the backbone is reported in Figure 3. Performance metrics such as accuracy, recall, and precision have also been reported in this experiment.

3. Results

Pre-trained segmentation models Unet and Linknet with various CNN backbones delivered a commendable performance. The highest performing model (Unet + EfficientnetB7) has a mean IoU of 91.69, F1-score of 95.65, accuracy of 95.75, and precision of 95.43 on test data. However, in terms of recall, Linknet + EfficientnetB7 outperformed all models with a recall value of 97.99. Table 1 shows the performance metrics obtained while employing six backbone CNNs on two segmentation models.

Table 1. The segmentation results on test data of two segmentation models with six backbone CNNs for road vs. background segmentation task. Note that the bold values represent the highest performance.

Backbone CNN	Pre-trained Segmentation model	Mean IoU	Diceloss	F1-score	Accuracy	Recall	Precision
VGG19	Unet	91.50	4.47	95.55	95.65	97.92	93.29
	Linknet	91.52	4.52	95.56	95.66	97.92	93.30
Resnet50	Unet	91.36	4.55	95.47	95.57	97.84	92.31
	Linknet	91.41	4.57	95.50	95.60	97.87	93.24
Resnet152	Unet	91.36	4.54	95.47	95.57	97.79	93.26
	Linknet	91.38	4.55	95.48	95.58	97.80	93.27
MobilenetV2	Unet	91.00	4.75	95.27	95.38	97.64	93.01
	Linknet	90.13	5.25	94.78	94.90	97.08	92.59
EfficientnetB0	Unet	91.41	4.53	95.49	95.59	97.68	93.24
	Linknet	90.74	4.97	95.12	95.23	97.43	92.92
EfficientnetB7	Unet	**91.69**	**4.36**	**95.65**	**95.75**	97.97	**93.43**
	Linknet	91.65	4.48	95.63	95.72	**97.99**	93.37

Table 2 shows a comparison of our results with previous studies. Pre-trained U-Net with EfficientnetB7 backbone exceptionally outperformed all the previous models. All the pretrained models surpassed the results of prior research. Notably, the precision and recall values showed substantial improvements, while the F1-score and IoU values reached exceptionally high levels.

Table 2. Comparing performance metrics of Massachusetts Roads Dataset.

Method	Precision	Recall	F1-score	IoU
U-Net scratch [1]	82.46	84.34	83.39	60.97
SPIN Road Mapper [10]	83.9	85.06	84.47	65.24
D-Linknet [11]	91.35	93.41	77.90	63.80
Pre-trained U-Net + EfficientnetB7 (ours)	**93.43**	**97.97**	**95.75**	**91.69**

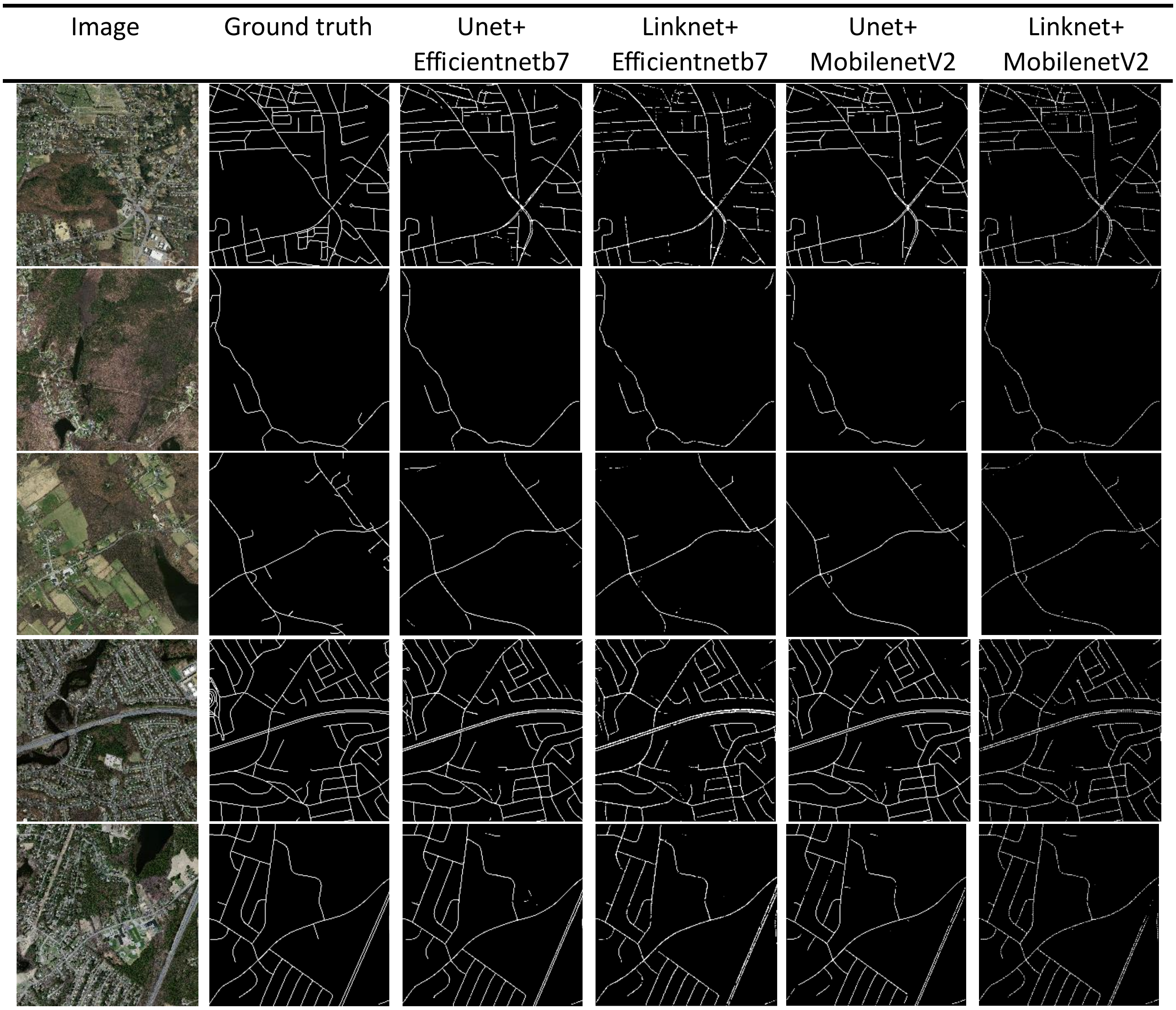

Figure 4. Segmented mask obtained from various segmentation models.

A few samples of test images, their corresponding ground truth masks, and masks obtained from Unet + efficientnetB7 and mobilenetV2, and Linknet + efficientnet and mobilenetV2 can be seen in Figure 4. This figure gives a visual confirmation of best performing CNN backbone vs. least performing backbone as can be seen in the table above. It can be seen that the mask corresponds to the input image in the second row where certain white pixels (i.e. road pixels) are missing. In contrast, additional pixels can be observed for the mask corresponding to the input image in the last row. Both result in reduced performance.

4. Conclusion

In this work, we extensively studied the outcome of deep learning-based pre-trained segmentation models with various CNN backbones for road detection. The results indicate that efficientnetB7 paired with U-net pre-trained on the Imagenet dataset outperforms all other models in all performance metrics except for recall in which Linknet paired with efficientnetB7 outperformed all. Furthermore, the comparison of six backbone CNNs on the Massachusetts Roads Dataset demonstrates similar performance for the semantic segmentation task. Also, pre-trained models with imagenet weights proved to be a reliable source domain to transfer learned features for the road segmentation domain.

References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

[2] Waldner, F., & Diakogiannis, F. I. (2020). Deep learning on edge: Extracting field boundaries from satellite images with a convolutional neural network. Remote sensing of environment, 245, 111741.

[3] Blaschke, T., Hay, G. J., Weng, Q., & Resch, B. (2011). Collective sensing: Integrating geospatial technologies to understand urban systems—An overview. Remote Sensing, 3(8), 1743-1776.

[4] Ozturk, O., Isik, M. S., Kada, M., & Seker, D. Z. (2023). Improving Road Segmentation by Combining Satellite Images and LiDAR Data with a Feature-Wise Fusion Strategy. Applied Sciences, 13(10), 6161.

[5] Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., & Alamri, A. (2020). Deep learning approaches applied to remote sensing datasets for road extraction: A state-of-the-art review. Remote Sensing, 12(9), 1444.

[6] Kamal, K. C., Yin, Z., Li, B., Ma, B., & Wu, M. (2019). Transfer Learning for Fine-Grained Crop Disease Classification Based on Leaf Images. 2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), 1–5.

[7] Mnih, V. (2013). Machine learning for aerial image labelling. University of Toronto (Canada).

[8] Iakubovskii, P. (2019). Segmentation Models Pytorch. GitHub Repository.

[9] Chollet, F.; Zhu, Q.S.; Rahman, F.; Qian, C.; Jin, H.; Gardener, T.; Watson, M.; Lee, T.; de Marmiesse, G.; Zabluda, O.; et al. Keras. 2015. Available online: https://github.com/keras-team/keras (accessed on 1 August 2023).

[10]Bandara, W. G. C., Valanarasu, J. M. J., & Patel, V. M. (2022, May). Spin road mapper: Extracting roads from aerial images via spatial and interaction space graph reasoning for autonomous driving. In 2022 International Conference on Robotics and Automation (ICRA) (pp. 343-350). IEEE.

[11]Xu, H., He, H., Zhang, Y., Ma, L., & Li, J. (2023). A comparative study of loss functions for road segmentation in remotely sensed road datasets. International Journal of Applied Earth Observation and Geoinformation, 116, 103159.