https://doi.org/10.24191/jeesr.v21i1.002

A Deep-learning Semantic Segmentation Approach for Road Segmentation of UAV Images

Mat Nizam Mahmud, Muhammad Khusairi Osman*, Anas Ibrahim, Ahmad Puad Ismail, Fadzil Ahmad, and Azmir Hasnur Rabiani

***Abstract*— Road image segmentation is important for a variety of reasons, including road maintenance, intelligent transportation systems, and urban planning. The method has recently been adapted to road images captured by unmanned aerial vehicles (UAV), because of its low cost, easy manoeuvrability and wide field of view. Due to the complexity of the backgrounds in these images, high-precision road segmentation from UAV images remains difficult. This research proposed an automated road segmentation model utilising a deep learning technique called DeepLab V3+ semantic segmentation to solve the problem. A UAV is utilised to capture and collect road images in the state of Kedah and Selangor, Malaysia. Then, the DeepLab V3+ with Resnet-50 backbone is developed and trained to segment the road from the background. The performance is then evaluated by comparing the segmented images using deep learning to images that have been labelled manually. For the evaluation, three measures are used: pixel accuracy (*PA*), mean area intersection by union (*mIoU*), and mean F1-score (*MeanF1*). Additionally, for benchmarking purposes, the research compares the segmentation performance with two different backbones of DeepLab V3+ called Resnet-18 and Mobile NetV2 backbone. According to simulation findings, the DeepLab V3+ with Resnet-50 outperformed the DeepLab V3+ with Resnet-18 and Mobile NetV2 techniques. Experiments on various road images as well as comparisons with different backbones demonstrate the effectiveness and robustness of the proposed DeepLab V3+ with Resnet-50. The method outperformed the Resnet-18 and Mobile NetV2 by at least 0.22%, 0.14%, and 1.39%, for *PA*, *mIoU* and *MeanF1* respectively.**

***Index Terms*— Road Image Semantic Segmentation, UAV, DeepLab V3+, Resnet-50.**

This manuscript is submitted on 7 April 2022 and accepted on 2 September 2022. This work was supported by Universiti Teknologi MARA through the Geran Insentif Penyeliaan (GIP), UiTM Grant No: GIP5/3 (034/2022).

Mat Nizam Mahmud, Muhammad Khusairi Osman, Ahmad Puad Ismail, Fadzil Ahmad are with Center for Electrical engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, Permatang Pauh Campus, 13500 Pulau Pinang, MALAYSIA. (e-mail: mtnizam86@gmail.com).

Anas Ibrahim is with Center for Civil engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, Permatang Pauh Campus, 13500 Pulau Pinang, MALAYSIA. (e-mail: ceanas@uitm.edu.my).

Azmir Hasnur Rabiani is with THB Maintenance Sdn. Bhd., 204 A, Jalan PSK 5, Pekan Simpang Kuala, 05400 Alor Setar, Kedah Darul Aman. (e-mail: azmir.ahr@gmail.com).

*Corresponding author
Email address: khusairi@uitm.edu.my

I. INTRODUCTION

Recent advancements in remote sensing technology have opened the way for a simpler and better method of monitoring geographical areas [1]. Traditionally, satellite remote sensing (SRS) has provided low spatial and temporal resolution for applications including land cover mapping, weather, meteorology, mineralogy, etc. Specifically, with the increasing use of unmanned aerial vehicle (UAV) for specific remote sensing applications, we have overcome the limitations of SRS with regards to spatial and temporal resolution [2].

The use of unmanned aerial vehicles (UAVs) in road engineering has increased in popularity recently. Analyzing images captured by the UAV has several applications, including vehicle tracking, object identification, road maintenance, and anomaly detection [3][4]. Most of these applications require inferring geographical and contextual information about the images. For example, if road knowledge is provided, vehicle tracking will be simpler.

Semantic segmentation is a method for dividing an image into separate semantic sections and categorizing these parts [4]. Semantic segmentation remains challenging owing to class variation, perspective loss, scene context, and the presence of noise, particularly when segmenting road pavements [5].

The critical problem in road segmentation is accurately identifying pixels in an image as being on or off the road (background). The variety of road areas in terms of their location, size, form, and colour makes developing effective segmentation algorithms more difficult. Additionally, when trees or buildings are covered by shadow in UAV images, the accuracy of road segmentation is compromised [6].

Nowadays, high-resolution visible remote sensing/UAV images are widely used due to the fast development of remote sensing technology. These high-resolution images provide a new challenge for road segmentation algorithms since they include more information and have a more complicated background than conventional images.

Numerous previous studies have indicated that roads may be recognized and segmented using high resolution visible remote sensing/UAV images [6][7]. Cheng et al. [8] used object-based feature extraction to recover coarse road regions and then performed pixel-based road segmentation. This method, however, has trouble segmenting boundary roadways consistently. The newly suggested JointNet uses the focus loss function to enhance road extraction while retaining a wider receptive field at the same time [9]. The JointNet is a mix of

dense connectivity and atrous convolution that successfully extracts both road and building areas.

Deep convolutional neural networks (DCNNs) have recently aided computer vision systems in a variety of tasks, including image classification [8], object identification [10], and semantic segmentation [11]. DCNNs have been successfully used in a variety of scientific and technological areas. This is because the DCNN can collect contextual data, which is important in many applications. The segmentation accuracy is controlled by both local parameters (colour and intensity) and global variables such as (texture, and context). Both features may be learned end-to-end, which contributed to the network's effectiveness in semantic segmentation [11]. Because they rely on learned features, deep learning methods to semantic segmentation are often preferred.

Another recent study suggests the use of a convolutional neural network (CNN) to extract structural features and then use multi-scale Gabor filters and edge-preserving filters to enhance feature extraction [12]. Y-Net, a new deep learning technique, combines feature extraction with a fusion module that can better segment multi-scale roadways in high-resolution images. In contrast with previous techniques, Y-Net performs better in extracting narrow roadways [12].

Few studies have used shallow neural networks for the road extraction [13]; some of them contain larger trainable weights and take advantage of local spatial coherence of the output [14]. The current improvement in computing speed and greater data resources have significantly boosted the use of deep neural networks (DNN). Gao et al. [15], by taking use of high-resolution remote sensing data, presented multi-level semantic characteristics. In this research, a new loss function is proposed to overcome misclassification loss and helps to concentrate on the spare set of actual labeled pixels in the training stage. A convolutional neural network (CNN) with different variants like fusion with a line integral convolution-based algorithm [16], a combination of deep convolutional neural network and finite-state machine [17], derivatives such as the road structure refined convolutional neural network (RSRCNN) [18], and DenseNet methods [19] were successfully applied on road segmentation. Fully convolutional networks (FCN) have lately acquired a lot of popularity and based on the data availability and computing capacity a choice may be made on whether to utilize pre-trained nets like the VGG-Net. Generally, dense networks like VGG-Net and ResNet contain a high number of layers which need a very lengthy training period and thus their pre-trained weights are utilized to perform various tasks [20]. These networks are trained on datasets like ImageNet which contain approximately 1000 classes. Training a network on such a huge number of classes, using an exceptionally big dataset and very deep neural networks is complicated. In most of the above CNN-based deep learning methods, the precision and accuracy of the segmentation improves significantly with the aid of deep Networks architecture. However, are also prone to a lot of computing effort and need big datasets.

In the literature, DeepLabv3+ networks are extensively utilized for image segmentation in a variety of difficult situations, including the automated drawing of retrogressive thaw slump regions, ship image segmentation, and pothole extraction on roads [21][22]. A lot of effort has been done on road extraction from street view but not much work has been undertaken on road extraction from UAV remote sensing images. Moreover, to the authors' knowledge, there is no study that has applied DeepLab V3+ networks to perform road extraction from UAV remote sensing images.

The first three editions of the DeepLab series , v2 , and v3 included a variety of network topologies for semantic segmentation [23]. DeepLab uses dilated convolution [24], which expands the receptive field while maintaining the same processing speed as normal convolutional operations. Additionally, when the feature map was shrunk to the same multiple, the bigger receptive field absorbed more information, increasing the accuracy of the segmentation predictions.

Chen et al.[25] presented an encoder-decoder structure with atrous convolution for fine-grained segmentation. This structure gathers high-level semantic information and facilitates segmentation at a finer level. It used a downsampling factor (output stride) of eight in DeepLabv3+, which means that the encoder part of the algorithm downsamples the input image by a factor of up to eight.

He et. al. [26] established the Residual Network (ResNet) framework for deep residual learning. It is made up of several residual blocks that pass the feature mappings (activations) of a layer to a deeper layer. Sandler et al. [27] introduced the MobileNetV2 as a mobile-optimized neural network. It is smaller and lighter than ResNet. DeepLabv3+ is widely regarded as the cutting-edge framework for semantic segmentation.

This study proposed an automated road segmentation model utilising a deep learning technique called DeepLab V3+ semantic segmentation method. The method is used for segmenting roads from high-resolution UAV images, intending to achieve high segmentation accuracy even when the background is complex. In this research, the DeepLabv3+ with ResNet-50 is utilized and benchmarked against two different backbones namely Resnet-18 and MobileNetV2. The proposed methodology will be thoroughly explained in the following section.

II. METHODOLOGY

The main aim of this study is to propose an automated road segmentation model based on a deep learning method for segmenting roads and backgrounds from UAV images. The segmented road images may be utilized for a variety of purposes, including road maintenance, intelligent transportation systems, and urban planning. The suggested approach is divided into five stages which are image acquisition and data collection using UAV, image processing and labelling, image augmentation, road and background segmentation, and performance evaluation and comparison. The procedure is summarised in Fig. 1. The following sections will discuss each of the proposed steps in more detail.

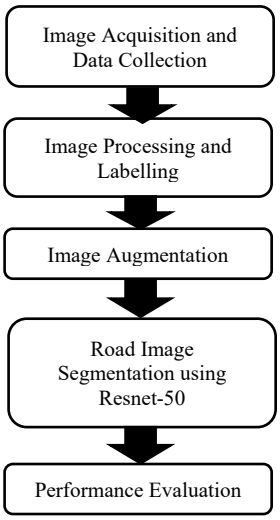


Fig. 1: Block diagram of the proposed automated road image segmentation using deep learning.

A. Image Acquisition and Data Collection

Image acquisition refers to the process of capturing road images using a UAV. The UAV used in this study is the DJI Mavic 2, as shown in Fig. 2. The UAV is equipped with a 1" CMOS sensor that is able to capture images with a resolution of up to 5472×3648.

The UAV is set to manoeuvre during the daytime along a road centre-line at a height of 30m to obtain clear road images. The camera was set to the highest resolution in order to acquire high-resolution road photographs that could clearly visualise the crack in the pavement. These road images are collected and further presented to deep learning for training and testing purposes. Several federal roads in the state of Selangor and Kedah were selected for this study. The road images also consist of non-road regions, refers to as background. The background usually includes buildings, trees, vehicles, and shadows.

Figure 3 shows an example of a road images that have been captured by the UAV. For this work, 200 images with 5472×3648 resolutions have been captured in Selangor and Kedah, Malaysia.

Fig. 2: The DJI Mavic 2 that has been used for the road image capturing process.

Fig. 3: Example of road pavement image captured by the UAV

B. Image Processing and Labelling

Image processing involves resizing the road image to a lower resolution. This process was carried out in order to fit the road images into the DeepLab **V3+** model and speed up the training process. In this study, the road images capture by the UAV were resized from 5472×3648 to a smaller resolution of 224×224 pixels. Figure 4 gives an example of a road image and its result after resizing process. After the images been collected, all images were resized from 5472×3648 as shown in 4(a) to 224×224 as shown in 4(b).

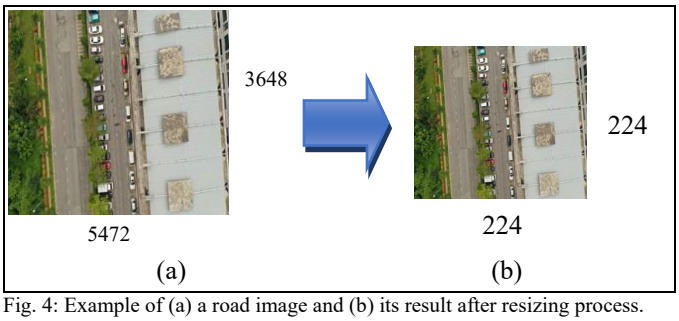


Fig. 4: Example of (a) a road image and (b) its result after resizing process.

After the image processing, manual labelling process was performed to label the images into two categories, road, and background. Manual labelling of road images is achieved by using an annotation tool called Image Labeler. The tool can be found in MATLAB under Image Processing and Computer Vision apps. In this study, the road area is labelled in blue, whereas the background is labelled black. Some example of road images and their result after manual labelling process are shown in Figure 5.

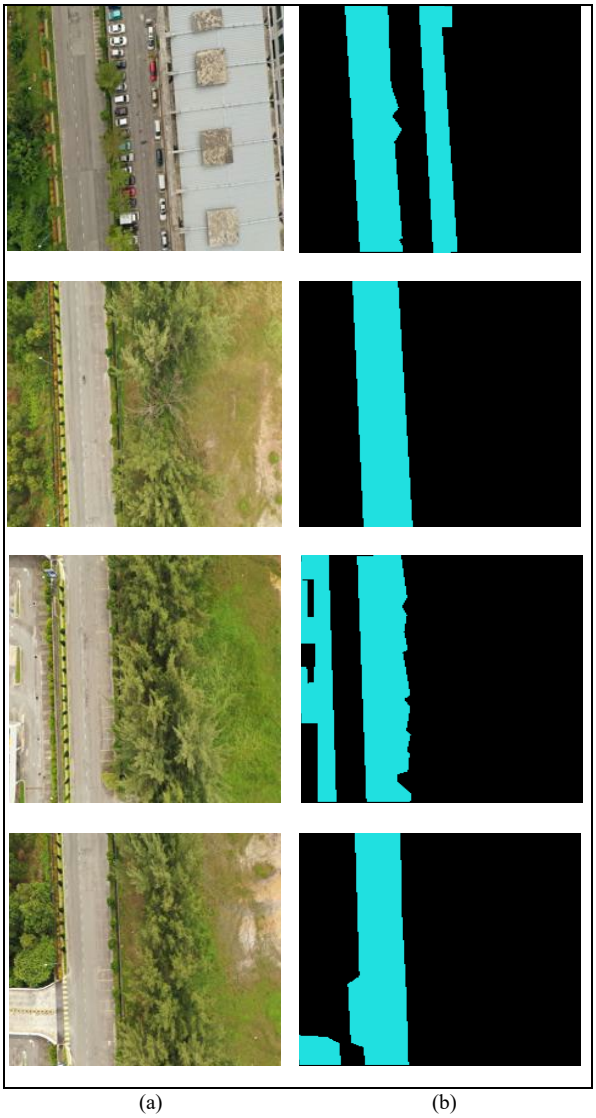
C. Image Augmentation

Image augmentation was applied to the manual labelling images to increase the number of road images without capturing

new images. Several image augmentation techniques including horizontal, vertical, and diagonal flipping were implemented. By implementing image augmentation, the diversity of road images can be increase significantly. This will assist in enhancing the network performance during the training phase [28]. Through the image augmentation process, the number of road images was significantly increased to 600 images.

D.Road Image Segmentation using DeepLab V3+

The road images are segmented using a deep learning-based image semantic segmentation model called DeepLab V3+ with a ResNet50 backbone. In general, the model consists of two main components which are encoder and decoder configuration, as illustrates in Figure 6. The encoder module handles multiscale contextual information by using dilated convolution neural network (DCNN) at several scales. The encoder starts with five conventional convolution modules and then transfers the output to four Atrous convolution modules and one average max-pooling module simultaneously. Normal



(a) (b)

Fig. 5: (a) Example of road images after the resizing process and (b) their results of manual image labelling

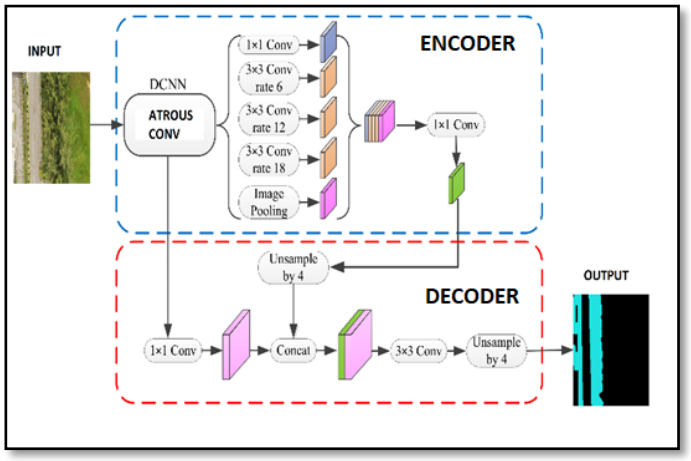


Fig.6: DeepLab V3+ architecture.

convolution's first five modules are generated using a pre-trained backbone model. Meanwhile, the decoder module refines the segmentation outputs at object boundaries. Decoder structure progressively reconstructs spatial information to better capture object boundaries.

The proposed model was also compared with two different backbones, Resnet-18 and MobileNetV2, to segment the road pavement target and eliminate irrelevant background noise, thus decreasing the effect of background noise on UAV images of road pavement. DeepLabv3+ is generally considered as the state-of-the-art semantic segmentation framework. In all experiments, the mini-batch size and epoch were set to 16 and 30 respectively. Epoch refers to the number of full passes through the training set while the mini-batch refers to a portion of the training patches used to update the network parameters (weights and biases). To find the optimised network parameters, the stochastic gradient descent was used as its training algorithm. The DeepLabv3+ architecture as well as the three backbones were developed using MATLAB 2020a version and Deep Learning Toolbox. The experiments were conducted on a personal computer (PC) equipped with Windows 10 Operating System (OS) and has the following hardware specifications: Intel i-5 10400 @ 2.90 GHz processor, 16 GB of DDR4 3200 MHz RAM and NVIDIA RTX 2070, 8 GB GPU.

E. Performance Assessment

A quantitative assessment was selected to evaluate the segmentation performance of DeepLab V3+. The assessment was carried out by comparing the network's pixel to pixel image segmentation output with the manual labelling image. Three performance measures were used in this study, which are Pixel Accuracy (*PA*), mean intersection over union (*mIoU*) , and MeanF1 score (*MeanF1*), as suggested by Baheti et al. [11].

The pixel accuracy (*PA*) value shows the number of correctly recognized and misclassified pixels, as well as the percentages associated with each. The following formula is used to determine the *PA*:

$$PA = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

where *TP, FP, TN*, and *FN* represent the number of true positive, false positive, true negative, and false negative pixels in that class, respectively. The positive class denotes the pixels on the road, while the negative class denotes the background.

Additionally, each class's performance is quantified in terms of the mean intersection over union (*mIoU*). The *IoU* for a certain class is defined as follows:

$$IoU = \frac{TP}{TP+FP+FN} \tag{2}$$

where *mIoU* is the average of all class *IoUs*. Additionally, the MeanF1-score has been used as a performance measure (*MeanF1*). This metric is an extremely efficient way to get the harmonic mean of P and R. *MeanF1* is obtained by averaging all F1-score classes. It is calculated using the equation below:

$$F1 = 2 \text{ x } \frac{PxR}{P+R} \tag{3}$$

where the following equations determine the precision and recall:

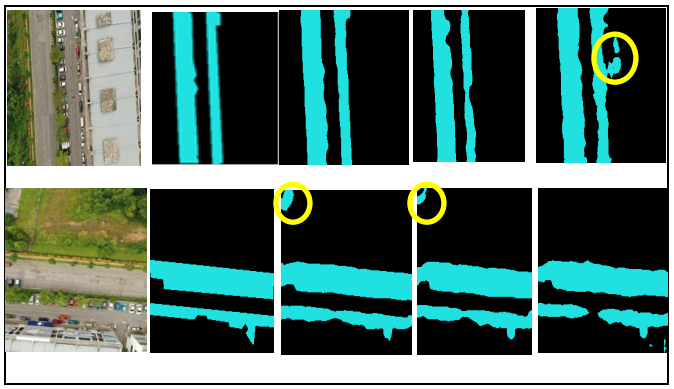
$$P = \frac{TP}{TP+FP} \tag{4}$$

$$R = \frac{TP}{TP+FN} \tag{5}$$

III. Results And Discussion

Road segmentation from UAV images is challenging due to the complexity of the images. Due to the high resolution, noise sources such as trees, people, buildings, and shadows affected the detection performance and created significant difficulties for road segmentation.

Figure 7 illustrates road images captured with the UAV against a background of trees and structures, as well as the image segmentation results. Figure 7(a) shows the original road images. Meanwhile, Figure 7(b)–(e) illustrates the results of manual labelling image, DeepLab V3+ with Resnet-50, DeepLab V3+ with Resnet-18 and DeepLab V3+ with the Mobile NetV2 backbone network. As can be observed, there are instances of false negatives for the background class, which is caused to the buildings' presence. This effect is clear in Figure 7 (c)-(e) when DeepLab V3+ predicts objects labelled as background as road pixels. A yellow circle indicates the false negative in Figure 7 (c)-(e).

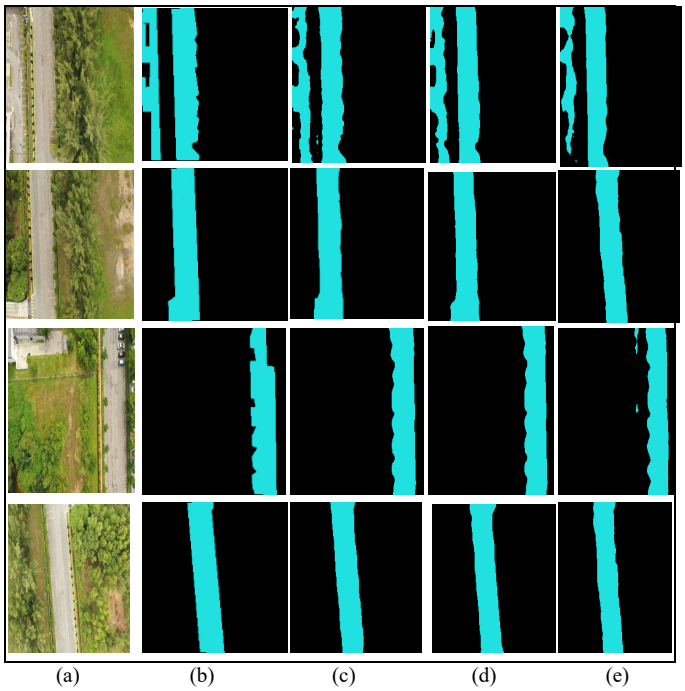


(a) (b) (c) (d) (e)

Fig. 7. Segmentation Results with tree and building background (a) original image (b) manual labelling images (c) Segmented image with Resnet-50 (d) Segmented image with Resnet-18 (e) Segmented image with MobileNetV2

Figure 8 illustrates road images captured with the UAV and the resulting image segmentation results. Figure 8 has the original road photos (a). Meanwhile, Figure 8(b)–(e) illustrates the results of manual labelling images, DeepLab V3+ with Resnet-50, DeepLab V3+ with Resnet-18, and DeepLab V3+ with the Mobile NetV2 backbone network. In this situation, the optimal segmentation technique is determined based on the result, with the optimal approach visually reflecting manual segmentation.

As shown in Figure 8, a few pixels within the road area indicate false positives for the road class. This uncertainty is increased by background noise, such as building and shadow. In general, segmentation performance was improved when the



(a) (b) (c) (d) (e)

Fig. 8. Some results of semantic segmentation using DeepLab V3+ architecture. (a) Original images (b) manual labelling images (c) segmented images using Resnet-50 (d) Segmented image with Resnet-18 (e) Segmented image with MobileNetV2

DeepLab V3+ with Resnet-50 was used instead of the Resnet-18 and Mobile NetV2.

To further verify the effectiveness of segmentation, the outcomes of all segmentations were analyzed quantitatively. A total of 240 road images were examined, and their performance was quantified and averaged in terms of *PA*, *mIoU*, and mean *meanF1*. The performance of both techniques is summarized in Table 1.

As shown by the data in Table 1, the *PA* of the DeepLab v3+ with Resnet-50 architecture is much greater than the *PA* of the DeepLab v3+ with Resnet-50 and Mobile NetV2 architectures. Resnet-50 has a *PA* of 97.1%, Resnet-18 has a *PA* of 96.88%, and mobile-NetV2 has a *PA* of 95.83%, respectively. Similarly, the *mIoU* for Resnet-50 produced somewhat more accurate findings, at 93.33%. Resnet-18 and mobile-NetV2, on the other hand, have a *mIoU* of 96.88 % and 95.83%, respectively. Additionally, the *MeanF1* for Resnet-50 produced more precise findings, with an accuracy of 90.14 %. However, the MeanF1 for Resnet-18 and mobile-NetV2 are 88.75 and 82.89 %, respectively.

In general, the DeepLab v3+ design with Resnet-50 outperformed the DeepLab v3+ architecture with Resnet-18 and Mobile NetV2 in terms of *PA*, *mIoU*, and *MeanF1*. This is because the Resnret-18 and MobileNetV2 encoders are tiny and shallow in DeepLabV3+, with a limited residual structure.

TABLE I
SEGMENTATION PERFORMANCE FOR 240 ROAD IMAGES USING DEEPLAB V3+ WITH RESNET-50 RESNET-18 AND MOBILE-NETV2.

Methods	*PA*	*mIoU*	*MeanF1*
ResNet-50	97.10%	93.33%	90.14%
ResNet-18	96.88%	93.19%	88.75%
Mobile-NetV2	95.83%	90.17%	82.89%

IV. CONCLUSION

This study makes a significant contribution by implementing a DeepLab V3+ network for road segmentation from UAV images with noisy backgrounds. The sophisticated architecture of the network enables it to cope with complex noise conditions. As shown by the comparison results, DeepLab V3+ with Resnet-50 backbone beat DeepLab V3+ with Resnet-18 and Mobile NetV2 backbone. For *PA*, *mIoU*, and *MeanF1*, DeepLab V3+ with Resnet-50 beats DeepLab V3+ with Resnet-18 and Mobile NetV2. DeepLab V3+ with Resnet-50 beats DeepLab V3+ with Resnet-18 and Mobile NetV2 by 0.22% and 1.27%, respectively, in the *PA* category. While DeepLab V3+ with Resnet-50 beat DeepLab V3+ with Resnet-18 and Mobile NetV2 by 0.14% and 3.16%, respectively, in the *MIoU*. Additionally, DeepLab V3+ with Resnet-50 beat DeepLab V3+ with Resnet-18 and Mobile NetV2 by 1.39% and 7.25%, respectively, for *MeanF1*.

This effort is the first stage in creating an automated pavement analysis and monitoring system for this research. Future work will concentrate on automatically detecting and classifying road severity features such as cracks and potholes from segmented images to automate pavement distress assessments.

V. ACKNOWLEDGEMENT

The authors acknowledge with gratitude the support and involvement of all parties, especially Universiti Teknologi MARA, Cawangan Pulau Pinang. This research is funded by the Geran Insentif Penyeliaan (GIP), UiTM Grant No: GIP5/3 (034/2022).

REFERENCES

[1] J. Senthilnath, D. Kumar, J. A. Benediktsson, and X. Zhang, "A novel hierarchical clustering technique based on splitting and merging," *Int. J. Image Data Fusion*, vol. 7, no. 1, pp. 19–41, 2016, doi: 10.1080/19479832.2015.1053995.

[2] Y. Guo, J. Senthilnath, W. Wu, X. Zhang, Z. Zeng, and H. Huang, "Radiometric calibration for multispectral camera of different imaging conditions mounted on a UAV platform," *Sustain.*, vol. 11, no. 4, 2019, doi: 10.3390/su11040978.

[3] H. Zhu, F. Meng, J. Cai, and S. Lu, "Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 12–27, 2016, doi: 10.1016/j.jvcir.2015.10.012.

[4] S. Girisha, M. M. Manohara Pai, U. Verma, and R. M. Pai, "Semantic segmentation of UAV aerial videos using convolutional neural networks," *Proc. - IEEE 2nd Int. Conf. Artif. Intell. Knowl. Eng. AIKE 2019*, pp. 21–27, 2019, doi: 10.1109/AIKE.2019.00012.

[5] Y. Li, B. Peng, S. Member, and L. He, "Road Segmentation of Unmanned Aerial Vehicle Remote Sensing Images Using Adversarial Network With Multiscale Context Aggregation," *IEEE J. Sel. Top. Appl. EARTH Obs. Remote Sens.*, vol. 12, no. 7, pp. 1–9, 2019.

[6] Y. Li, L. Guo, J. Rao, L. Xu, and S. Jin, "Road segmentation based on hybrid convolutional network for high-resolution visible remote sensing image," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 613–617, 2019, doi: 10.1109/LGRS.2018.2878771.

[7] N. Audebert *et al.*, "Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps To cite this version : HAL Id : hal-01523573 Joint Learning from Earth Observation and OpenStreetMap Data to Get Faster Better Semantic Maps," *arXiv*, pp. 67–75, 2017.

[8] S. S. Baek *et al.*, "Identification and enumeration of cyanobacteria species using a deep neural network," *Ecol. Indic.*, vol. 115, no. April, p. 106395, 2020, doi: 10.1016/j.ecolind.2020.106395.

[9] Z. Zhang and Y. Wang, "JointNet: A Common Neural Network for Road and Building Extraction," *Remote Sens.*, vol. 11, no. 6, p. 696, 2019, doi: 10.3390/rs11060696.

[10] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, and B. Ibragimov, "Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database," *Comput. Electr. Eng.*, vol. 78, pp. 388–399, 2019, doi: 10.1016/j.compeleceng.2019.08.004.

[11] B. Baheti, S. Innani, S. Gajre, and S. Talbar, "Semantic scene segmentation in unstructured environment with modified DeepLabV3+," *Pattern Recognit. Lett.*, vol. 138, pp. 223–229, 2020, doi: 10.1016/j.patrec.2020.07.029.

[12] R. Liu *et al.*, "Multiscale road centerlines extraction from high-resolution aerial imagery," *Neurocomputing*, vol. 329, pp. 384–396, 2019, doi: 10.1016/j.neucom.2018.10.036.

[13] M. Mokhtarzade, "Road Detection from High Resolution Satellite Imagery Using Texture Parameters in Neural Network," *Entropy*, no.

January, 2007.

[14] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6316 LNCS, no. PART 6, pp. 210–223, 2010, doi: 10.1007/978-3-642-15567-3_16.

[15] X. Gao *et al.*, "An End-to-End Neural Network for Road Extraction from Remote Sensing Imagery by Multiple Feature Pyramid Network," *IEEE Access*, vol. 6, pp. 39401–39414, 2018, doi: 10.1109/ACCESS.2018.2856088.

[16] P. Li *et al.*, "Road network extraction via deep learning and line integral convolution," *Int. Geosci. Remote Sens. Symp.*, vol. 2016-Novem, pp. 1599–1602, 2016, doi: 10.1109/IGARSS.2016.7729408.

[17] J. Wang, J. Song, M. Chen, and Z. Yang, "Road network extraction: a neural-dynamic framework based on deep learning and a finite state machine," *Int. J. Remote Sens.*, vol. 36, no. 12, pp. 3144–3169, 2015, doi: 10.1080/01431161.2015.1054049.

[18] Y. Wei, Z. Wang, and M. Xu, "Road Structure Refined CNN for Road Extraction in Aerial Image," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 709–713, 2017, doi: 10.1109/LGRS.2017.2672734.

[19] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road extraction from high-resolution remote sensing imagery using deep learning," *Remote Sens.*, vol. 10, no. 9, 2018, doi: 10.3390/rs10091461.

[20] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, 2017, doi: 10.1109/TGRS.2016.2636241.

[21] M. Majurski *et al.*, "Cell image segmentation using generative adversarial networks, transfer learning, and augmentations," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, vol. 2019-June, pp. 1114–1122, 2019, doi: 10.1109/CVPRW.2019.00145.

[22] S. Du, S. Du, B. Liu, and X. Zhang, "Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high resolution remote sensing images," *Int. J. Digit. Earth*, vol. 14, no. 3, pp. 357–378, 2021, doi: 10.1080/17538947.2020.1831087.

[23] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," 2017, [Online]. Available: http://arxiv.org/abs/1706.05587.

[24] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *4th Int. Conf. Learn. Represent. ICLR 2016 - Conf. Track Proc.*, 2016.

[25] T. H. Kim, M. S. M. Sajjadi, M. Hirsch, and B. Sch, *Encoder-Decoder with Atrous Separable Convolution for Semantic*. Springer International Publishing, 2018.

[26] V. Sangeetha and K. J. R. Prasad, "Syntheses of novel derivatives of 2-acetylfuro[2,3-a]carbazoles, benzo[1,2-b]-1,4-thiazepino[2,3-a]carbazoles and 1-acetyloxycarbazole-2- carbaldehydes," *Indian J. Chem. - Sect. B Org. Med. Chem.*, vol. 45, no. 8, pp. 1951–1954, 2006, doi: 10.1002/chin.200650130.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4510–4520, 2018, doi: 10.1109/CVPR.2018.00474.

[28] L. Perez and J. Wang, "The Effectiveness of Data Augmentation in Image Classification using Deep Learning," 2017, [Online]. Available: http://arxiv.org/abs/1712.04621.