SEGMENTATION FOR AERIAL IMAGE WITH DEEP LEARNING METHOD

Agnivesh Pandey[1], Rohit Raja[2]

[1,2] SOSE&T, IT GGV KONI BILASPUR C.G. India 495009, www.ggu.ac.in

**Contact e-mail: agnitu1984@gmail.com*

Abstract. *Deep Convolution Neural Network (DCNN) have been utilized to accomplish cutting edge performance on Computer Vision Task like object recognition object discovery and Semantic Segmentation, in this paper we borrow lately popular frame U-Net in computer vision for pixel position bracket Semantic Segmentation of Aerial Imagery. In this paper model is trained with image patches of size 256x256x3 of Images and their Masks patches are cropped from original image which aren't same size, to estimate the model we crop the images due to different in their size, resize will lead the loose the original image objects we will crop them to nearest size separable by 256 and also divide all images in to patches not to resize the image to minimize the noise. We estimated our network on data set comprises of upstanding symbolism of Dubai attained by MBRSC satellites and clarified with pixel-wise semantic division in 6 classes. Result Mean IoU (Intersection over Union) is calculated and achieved good results.*

Key words: *Deep Learning, convolution, Neural Network, Semantic Segmentation, Aerial Imagery component.*

1. INTRODUCTION

Images are most powerful media for carry lots of information and the top view images aerial images are captured easily due to advancement in drone technology. Unmanned aerial vehicles are used extensively in the military space and are turning out to be progressively popular for monitoring and surveillance of such area where human cannot reach easily, flying symbolism is pictures taken from airplanes like robots, planes, and helicopters and huge amount of data in the form of images are generated these images need to analyze for useful information which need new method technology and algorithm for object detection, classification and segmentation. These aerial Images are used to train a machine learning model to detect objects.

Aerial imagery technologies are very popular for image capturing and computer vision. Because aerial detection for the human eye is difficult in comparison to a highly trained computer vision model, Application Specific data set are available on line for the computer vision task for pattern of agriculture by UAV imagery is reported by Author [1] Ge-otagged images for efficient and Accurate Aerial imagery [2] RGB Aerial Imagery are process for precision agriculture[3] [4] Thermal and visual image registration are used for real time surveillance using light weight UAV [5]vegetation information extraction Application[6] aerial image segmentation is very useful at the time of Disaster management for the early and quick action for example flood location segmentation and forest fire detection in the scenario human cannot easily access , In medical science diseases like detection and location of diseases is identified by semantic segmentation utilizing the deep learning technique R-CNN and U-NET [7] [8] [9].

So need an efficient and fast segmentation model for Aerial imagery. As a result, we pro-pose Deep learning model based on the U-Net with different optimal hyper parameters for detecting and segmenting the Aerial images. We have used the publicly available kaggle data set named MBRSC dataset exist under the CCO license in this data set pixel wise labeled in six classes. MBRSC dataset consists of 72 limited numbers of images in HEX code and Mask is in RGB. Our assumption for this paper is summarized below.

- We use high resolution multispectral Aerial images for semantic segmentation which is supervised deep learning based on the labels of annotated mask of different classes.

- In the dataset of aerial images are annotated in six predefined classes and do not consider any class beyond the given mask.

- It is assumed that single object class is present in one place and other class objects are not overlap.

Rest of the work is coordinated into following Parts; Part-2 about the previous works and different procedures for Image processing in computer for classification and segmentation. Part-3 Describe the dataset and technique for the assessment. Part-4 is the discussion on experiment results. Part-5 closes the paper with future work.

2. PREVIOUS WORK

Aerial image segmentation and classification task are very essential for the application of urban planning aerial surveillance and Disaster management the segmentation task is challenging in aerial imagery due to variable shape, scale and appearance in Aerial view. The most approaches of solving this problem suppose the usage of deep learning algorithms. In the process of training the Machine Learning model network extract the features automatically. Lately convolutional neural Networks (CNNs) were proposed for the task of segmentation. The fundamental thought of FCNs is theUtilization of completely associated layers with a convolution layer at end, while different layers separate fundamental highlights from input information. It permits to send off this kind of organization for image segmentation [15].

Deep CNNs (DCNNs) were successful in many high-level computer vision tasks, ranging from image classification [10][11] and detection of small object [12], visual detection and image Segmentation [13]DCNNs address teachable errands in a start to finish style, which ordinarily includes joint gaining of a progression of component extractions from crude information to a last, task-explicit result. DCNNs have additionally been applied to remote detecting. [14][15]

Author propose a multi-class, high-precision identification strategy for UAV pictures Faster RCNN[16][17]In semantic segmentation approach the fundamental step is assigning each pixel in an image to one of several semantic classes it is a supervised learning in contrast to standard unsupervised segmentation in which similar regions of pixels are groped based on basic low level features such as colour or texture. In this paper we have six semantic classes is incorporated in broad view semantic segmentation is one of the great level undertakings which give the comprehension of complete scene. Semantic segmentation of ethereal symbolism has been utilized in different applications like peril ID and aversion, traffic the executives and assessment, and metropolitan region arranging and observing[18].The u-net is convolutional network architecture for fast and precise segmentation of biomedical images [19]. In the literature various Aerial image dataset are available publically for the image processing task summarise in table (1).

Table 1.List of dataset and description

Name of Dataset	Total images	Image size	classes	labels	operation	channels	Model used
Agri vision[1]	**94986**	**512x512**	**9**	**169086**	**Segmentation**	**RGB,NIR**	**DeepLabV3+**
DOTA [20]	**2806**	**4000x4000**	**14**	**188282**	**Detection**	**RGB**	**R-CNN**
iSAID[21]	**2806**	**4000x4000**	**15**	**655451**	**Segmentation**	**RGB**	**Mask R-CNN and PANet**
AID[22]	**10000**	**600x600**	**30**	**10000**	**Classification**	**RGB**	**VGG-VD-16**
SAT-6[23]	**40500**	**28x28**	**6**	**40500**	**Classification**	**RGB,NIR**	**Deep Belief Network**
Traffic image[24]	**15070**	**1920x1080**	**2**	**155328**	**Detection**	**RGB**	**YOLOV5M**

Source: Realised by authors

3. TOOLS AND TECHNOLOGIES

Aerial images are obtained from the satellites by high-resolution camera so these images are high spectral images which are usually 16-bit colour scheme in which most of objects are appear very small and cannot easily recognize by human eyes this motivate to extract the useful feature from the images using machine learning model and use for object classification on pixel level which is known as semantic segmentation.

A. Dataset

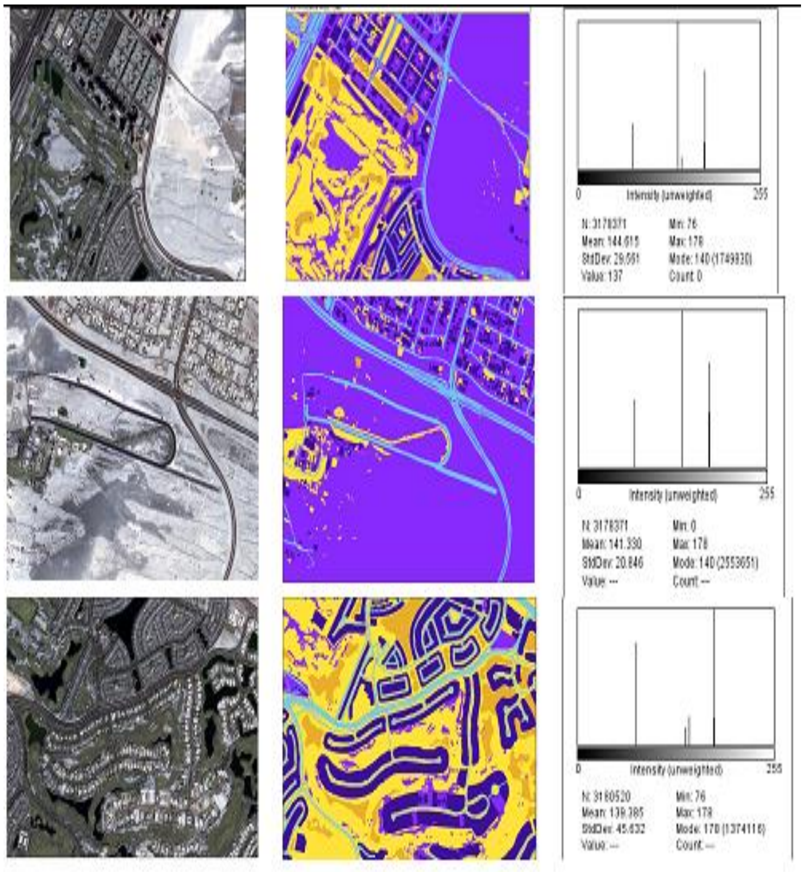
The dataset comprises of elevated symbolism of Dubai acquired by MBRSC (Mohammed Bin Rashid Space Center) satellites dataset incorporates 72 pictures assembled into 8 bigger tiles Each tile has pictures of various levels and widths, and a few pictures inside similar tiles are variable in size Annotated with pixel-wise semantic division in 6 classes. The pictures were portioned by the learners of the Roia Foundation in Syria. This semantic division dataset is devoted to the public space by Humans in the know under CC0 1.0 permit.

Images in this dataset have Class colors are in Hexadecimal, whilst the mask images are in RGB and unbalanced shown in below fig (1) make these images ready for the machine learning model as a input need processing for better feature extraction buy the proposed model in the paper class cooler are converted in to RGB in fig 1:

(a) aerial images obtained from Satellites;

(b) corresponding mask of the aerial images;

and (c) histogram of the mask, as images are annotated in six classes of Building, Land (Unpaved area), Road, Vegetation, Water, Unlabeled so in the histogram maximum of six bars are appeared and all are in different in height which shows the object in the image are unbalanced.

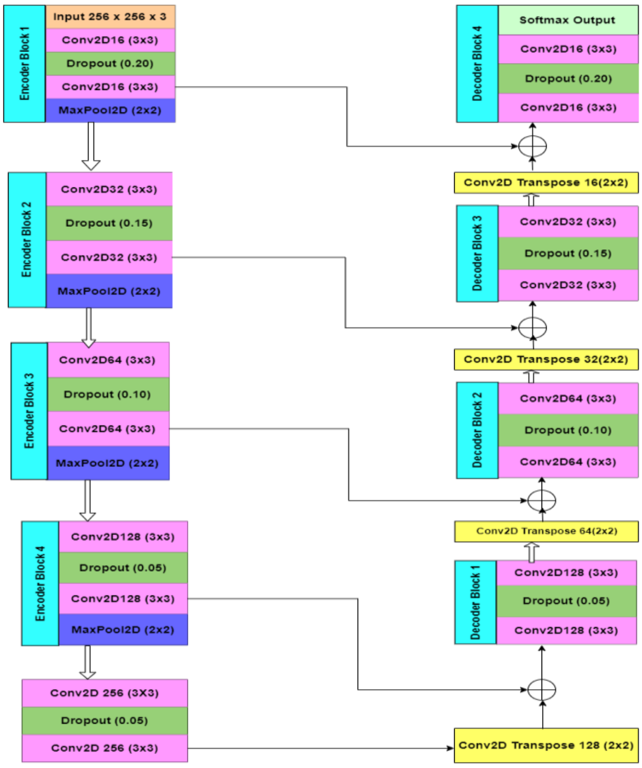


(a) **(b)** **(c)**

Fig 1. (a) Aerial images obtained from Satellites (b) corresponding mask of the aerial images and (c) histogram of the mask

B. Experiment Setup

In this paper U-Net model use as backbone model with some hyper parameters like optimization, Max-pooling, relu activation function, Decreasing dropout rate in Encoder block 1 to Encoder block 5. It is 0.20, 0.15, 0.10, 0.05 and 0.01 and reverse in decoder block it result the better segmentation. since images are variable in size and in limited numbers, we crop them in to patches of 256x256 in smooth manner from the original images as a result a greater number of patches are generate for the training of model. This process is also applied on mask image. U-Net model input is image patches of size 256x256 of batch size 8. For 50 epoch. Model is implemented on Google Colab in python programming with the use of standard Keras and Tensor flow environment in Google Colab with Nvidia K80/T4 GPU with ram size of 12GB/16GB. U-Net perform downwards operations(Encoder) and upwards operations (Decoder). For the encoder operation, input of size 8x256x256x3 and batch for the input is 8to the propose architecture, which contains two times of convolution two dimension of operations and size of 3 x 3 filter of 16, followed by decreasing dropout rate and this process repeats in each node, after this we apply max pooling of 2X2 with stride2, this process repeated for each encoding block.



Source: Realized by authors

Fig 2. Propose U-NET Architecture

In decoder block, reverse of encoding task is applied for producing the same output size as input the concatenation operation is performed as shown in fig 2 for the feature size has been expended.

The growing way has the up testing (deciphering) activity with the past layer in which each step has half of the component channels in convolution two layered layers that are 256 down to 16 in every hub. It followed every one of the means by up inspecting. The last soft max activation function is used because it is multi class segmentation.

It is repeated for all batches of cropped images during training. The model streamlines weight cycle by emphasis. The architecture of propose U-Net model is shown in Fig-2. For the assessment of model execution for the semantic segmentation Intersection over Union (IoU) Jaccard index is used which measure the comparability between limited number of objects of contrast class. In the data set object classes are unbalanced combination Dice loss and Focal loss is use as loss function in propose model.

Jaccard index (J) A is set of actual class and B is Predicted class.

$$J(A, B) = |A \cap B| / |A \cup B|$$

$$IoU = \frac{\sum_{j=1}^{k} n_{ij}}{\sum_{j=1}^{k} (n_{ij} + n_{ij} + n_{ij})} \qquad (1)$$

3.1.1. Dice-Loss (DL)

$$DL(Y, \hat{P}) = 1 - \frac{2y\hat{p}+1}{y+\hat{p}+1} \qquad (2)$$

In equation (2) 1 is added to denominator and numerator to stay away from the mistake in edge situation when y and $\hat{P}$ are zero. Where Y, $\hat{P}$ is actual and predicted respectively

3.1.2. Focal Loss (FL)

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \qquad (3)$$

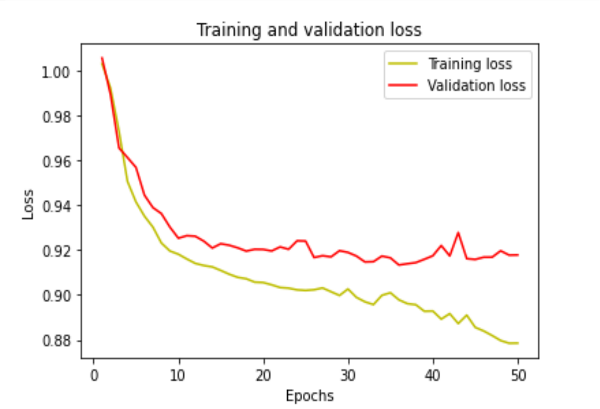
Loss function total loss TL is defined as sum of (2) and (3)

TL = Dice Loss + Focal Loss

We optimize the model by Adam optimizer. We trained the networks for 50 epochs for images and mask. Every age was prepared with a group size of 8 picture patches for all class objects. Each clump haphazardly trimmed a fix from a unique picture of 256 x 256.

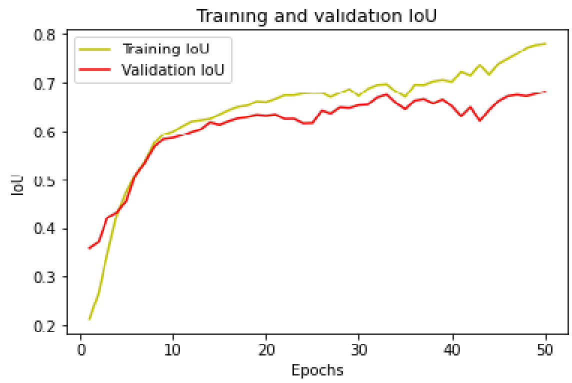
4. RESULTS AND DISCUSSIONS

In the propose model total of 1,941,190 parameters were trained by the input of images in x train and mask in y train as in dataset total of 8 tile of image folder is available and each folder have 9 images of aerial view and its corresponding mask is present, we have chosen tile no 8 in which 9 large images and 9 mask is present for training the experiment. With these limited data we have achieved the good result which is shown in give below fig 3 and Mean IoU is 0.5938 is achieved for the segmentation task IoU more than 0.50 is good and can be consider for the segmentation prediction result is shown in the fig (4)



Source: Realised by authors

(a)



Source: Realised by authors

(b)

Fig 3. **(a)** Training loss Vs validation loss **(b)** Training IoU vs Validation IoU of Propose U-NET Architecture

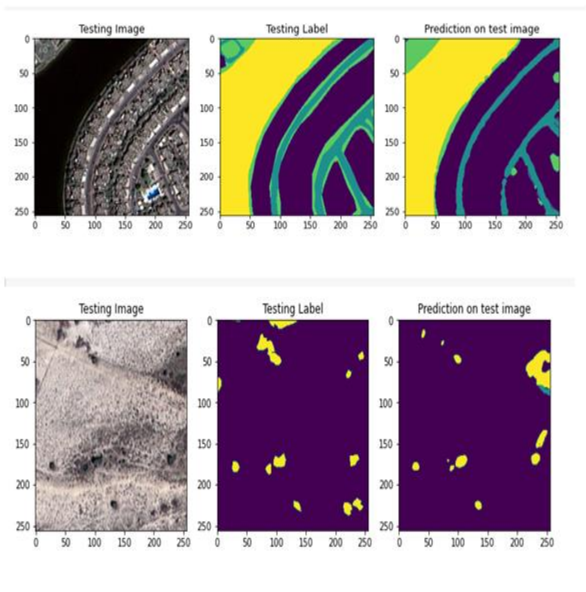


Fig 4. (a) Testing image (b)Testing Label (c) Prediction on Test

5. CONCLUSION

We are segmented the aerial images using U-net model with hyper parameters semantic segmentation result for this dataset is fair with small set of training data of only 9 images patches we were not applied any data augmentation technique we process the original images with minimum information loss during the pre-processing of data which result early and good feature extraction during the training, we introduce the different dropout rate for each encoder and decoder block which shows fair Jaccard score we have shown u-net can perform the semantic segmentation task on aerial images on small dataset in this paper images were cropped not resized to minimize the so that efficient and good features were extracted by the model. This experiment can utilize for disaster management and other monitoring activity where human cannot easily reach, in future more data with batch size of 128 can be used for better training the model.

6. ACKNOWLEDGEMENTS

We thank MBRSC (Mohammed Bin Rashid Space Center) and Kaggle for giving an informational collection to this exploration.

7. DECLARATIONS OF CONFLICT OF INTERESTS

There is no conflict of interest between the authors.

8. REFERENCES

[1] Chiu, M. T., Xu, X., Wei, Y., Huang, Z., Schwing, A. G., Brunner, R.,& Shi, H., 2020. “Agriculture-vision: A large aerial image database for agricultural pattern analysis,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2825–2835, doi: 10.1109/CVPR42600.2020.00290.

[2] H. Xiang and L. Tian, 2010. “Method for automatic georeferencing aerial remote sensing (RS) images from an unmanned aerial vehicle (UAV) platform,” *Biosyst. Eng.*, vol. 108, no. 2, pp. 104–113, 2011, doi: 10.1016/j.biosystemseng.2010.11.003.

[3] S. Pandey, R. Miri, G. R. Sinha, and R. Raja, 2022. “AFD filter and E2N2 classifier for improving visualization of crop image and crop classification in remote sensing image,” *Int. J. Remote Sens.*, pp. 1–26, Jan. 2022, doi: 10.1080/01431161.2021.2000062.

[4] H. Zheng *et al.*, 2020. "Early season detection of rice plants using RGB, NIR-G-B and multispectral images from unmanned aerial vehicle (UAV),” *Comput. Electron. Agric.*, vol. 169, no. March 2019, p. 105223, 2020, doi: 10.1016/j.compag.2020.105223.

[5] S. Yahyanejad and B. Rinner, 2015. “A fast and mobile system for registration of low-altitude visual and thermal aerial images using multiple small-scale UAVs,” *ISPRS J. Photogramm. Remote Sens.*, vol. 104, pp. 189–202, 2015, doi: 10.1016/j.isprsjprs.2014.07.015.

[6] X. Zhang, F. Zhang, Y. Qi, L. Deng, X. Wang, and S. Yang, 2019.“New research methods for vegetation information extraction based on visible light remote sensing images from an unmanned aerial vehicle (UAV),” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 78, no. January, pp. 215–226, doi: 10.1016/j.jag.2019.01.001.

[7] E. Kot, Z. Krawczyk, K. Siwek, L. Królicki, And P. Czwarnowski, 2021. “Deep learning-based framework for tumour detection and semantic segmentation,” *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 69, no. 3, pp. 1–7, 2021, doi: 10.24425/bpasts.2021.136750.

[8] Z. Krawczyk and J. Starzyński, 2021. “Segmentation of bone structures with the use of deep learning techniques,” *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 69, (3), pp. 1–8, doi: 10.24425/bpasts.2021.136751.

[9] T. Les, 2021. “U-Net based frames partitioning and volumetric analysis for kidney detection in tomographic images,” *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 69, no. 3, pp. 1–9, 2021, doi: 10.24425/bpasts.2021.137051.

[10] M. Mafanya, P. Tsele, J. Botai, P. Manyama, B. Swart, and T. Monate, 2021. “Evaluating pixel and object based image classification techniques for mapping plant invasions from UAV derived aerial imagery: Harrisia pomanensis as a case study,” *ISPRS J. Photogramm. Remote Sens.*, vol. 129, pp. 1–11, 2017, doi: 10.1016/j.isprsjprs.2017.04.009.

[11] Y. Hua, L.Mou, and X.X. Zhu, 2020. “Relation network for multilabel aerial image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, (7), pp. 4558–4572, doi: 10.1109/TGRS.2019.2963364.

[12] Z. Li, X. Liu, Y. Zhao, B. Liu, Z. Huang, and R. Hong, 2021. “A lightweight multi-scale aggregated model for detecting aerial images captured by UAVs,” *J. Vis. Commun. Image Represent.*, vol. 77, no. January, p. 103058, 2021, doi: 10.1016/j.jvcir.2021.103058.

[13] K. Yue, L. Yang, R. Li, W. Hu, F. Zhang, and W. Li, 2019. “TreeUNet: Adaptive Tree convolutional neural networks for subdecimeter aerial image segmentation,” *ISPRS J. Photogramm. Remote Sens.*, vol. 156, no. April 2018, pp. 1–13, 2019, doi: 10.1016/j.isprsjprs.2019.07.007.

[14] Kyrkou, Christos, and Theocharis Theocharides, 2019. "Deep-Learning-Based Aerial Image Classification for Emergency Response Applications Using Unmanned Aerial Vehicles." *In CVPR Workshops*, 2019, pp. 517-525, doi: 10.48550/arXiv.1906.08716.

[15] T. C. Su, 2017. “A study of a matching pixel by pixel (MPP) algorithm to establish an empirical model of water quality mapping, as based on unmanned aerial vehicle (UAV) images,” *Int. J. Appl. Earth Obs. Geoinf.*, vol. 58, pp. 213–224, doi: 10.1016/j.jag.2017.02.011.

[16] Raja, Rohit, Chetan Swarup, Abhishek Kumar, Kamred Udham Singh, Teekam Singh, Dinesh Gupta, Neeraj Varshney, and Swati Jain. 2022. "Detection of Behavioral Patterns Employing a Hybrid Approach of Computational Techniques," *Computers, Materials and Continua*, vol. 72, no. 1, pp. 2015–2031, 2022, doi: 10.32604/cmc.2022.022904.

[17] Zheng, Juepeng, Haohuan Fu, Weijia Li, Wenzhao Wu, Le Yu, Shuai Yuan, Wai Yuk William Tao, Tan Kian Pang, and Kasturi Devi Kanniah 2021. "Growing status observation for oil palm trees using Unmanned Aerial Vehicle (UAV) images," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, no. August 2020, pp. 95–121, 2021, doi: 10.1016/j.isprsjprs.2021.01.008.

[18] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, 2017.“A Review on Deep Learning Techniques Applied to Semantic Segmentation,” pp. 1–23, 2017, [Online]. Available: http://arxiv.org/abs/1704.06857

[19] W. Weng and X. Zhu, 2021. “INet: Convolutional Networks for Biomedical Image Segmentation,” *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.

[20] Xia, Gui-Song, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang, 2018. "DOTA: A large-scale dataset for object detection in aerial images," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3974–3983, doi: 10.1109/CVPR.2018.00418.

[21] Waqas Zamir, Syed, Aditya Arora, Akshita Gupta, Salman Khan, Guolei Sun, Fahad Shahbaz Khan, Fan Zhu, Ling Shao, Gui-Song Xia, and Xiang Bai 2019.. "isaid: A large-scale dataset for instance segmentation in aerial images," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 28-37, Available on line from: http://arxiv.org/abs/1905.12886

[22] Xia, Gui-Song, Jingwen Hu, Fan Hu, Baoguang Shi, Xiang Bai, Yanfei Zhong, Liangpei Zhang, and Xiaoqiang Lu, 2017. "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, 2017, doi: 10.1109/TGRS.2017.2685945.

[23] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, 2015.“DeepSat - A learning framework for satellite imagery,” *GIS Proc. ACM Int. Symp. Adv. Geogr. Inf. Syst.*, vol. 03-06-Nove, pp. 1–22, 2015, doi: 10.1145/2820783.2820816.

[24] S. B. Rosende, S. Ghisler, J. Fernández-Andrés, and J. Sánchez-Soriano, 2022. “Dataset: Traffic Images Captured from UAVs for Use in Training Machine Vision Algorithms for Traffic Management,” *Data*, vol. 7, no. 5, 2022, doi: 10.3390/data7050053.