International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4054-4061 www.ijaem.net ISSN: 2395-5252

Runway Image Segmentation for UAVs using U-Net

1.Dr.V.Amirthalingam, 2.Noble Mathew Sunil, 3.Sarath S Pillai, 4.Jammula Varun

[1]*Prof., Department of Computer Science and Engineering,Vinayaka Mission's KirupanandaVariyar Engineering College, Salem, India*

[2,3,4]*UG Student,Vinayaka Mission's KirupanandaVariyar Engineering College, Salem, India*

Submitted: 15-07-2021 Revised: 29-07-2021 Accepted: 31-07-2021

ABSTRACT:Unmanned Aerial Vehicles (UAV) belong to a class of aircrafts that can be flown without a pilot on board. These are sophisticated machines that contain sensor payloads which allows for the manoeuvrability and data collection. The UAV uses wireless communication to talk to the ground station. These remotely monitored machines have found uses in various domains such as disaster management, path planning, search and rescue operation etc. This paper proposes the usage of U-Net which is a variant of deep Convolutional Neural Network to automatically segment the runway from aerial images. Input image and its corresponding binary mask are created. This is then fed into the U-Net which learns to understand which is the runway and which is not on a pixel by pixel basis. Finally, it outputs a segmented image which highlights only the runway portion in the image. Segmenting and understanding the correct runway path can help in adding autonomous capabilities to the UAV wherein without any human intervention, UAVs can take off from the airstrip do routine tasks and safely land back at the base. The proposed method achieves a training accuracy of 85% and validation accuracy was about 81.25%.

KEYWORDS:UAV, Deep learning, CNN, Image Segmentation, Aerial Images, U-Net

I. INTRODUCTION

There has been a tremendous increase in the application of the Unmanned Aerial Vehicle over the last 10 years. They find use in urban planning[1], environment mapping[2,3,4] and surveillance applications[5]. These sophisticated machines work on the basis of GPS(Global Positioning System) and the combination of GPS with INS(Inertial Navigation System). The problem with the GPS is that it's susceptible to signal loss mostly in tunnels and due to jamming. So such errors in position are balanced by INS. When both methods fail then a computer vision based method using CCD camera system is employed for vision based navigation. Apart from that they can also house built-in sensors like compasses, gyroscopes, magnetometer, LIDARs, thermal cameras and SONAR. With advancement in communication modules and battery backup, they are serving an integral part in monitoring complex and dynamic environments.

UAVs are crucial in tactical planning tasks. These are used to assist the team in disaster management[6]. These are equipped with high resolution cameras with very high precision cameras that can take pictures of the disaster struck area and pass on the information to the disaster team to formulate a plan to help the affected in that area. The ground control team will provide live updates to the team while entering the affected area and can signal the crew to move to a particular area for focussed searching. UAVs played a major role in assessing the disaster and managing it during the Hurricane Katrina and Haiti earthquake. This paper also emphasizes the use of UAV as "tower in the sky"[7], which acts as a temporary mobile tower to re-establish the communication in that affected area for a short period of time.

UAVs are classified based on the wing configurations[8] such as rotary wind, fixed wing and hybrids. Fixed wings have sturdy wings with an air foil to increase the thrust speeds. More aerodynamics and less effort enables this to cover large distances and perform long endurance flights. Due to the nature of the configurations they need an airstrip to launch & land and they also are not capable of hovering about a point. The other design with the rotary wings possess sharp manoeuvrability. The one advantage this has over the former would be the ability to do vertical take-off through aerodynamic thrusts but the fall back being they are not suited to carry heavy payloads. In the upcoming years, developments will lead to a wider use of UAV platforms in many more applications.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4054-4061 www.ijaem.net ISSN: 2395-5252

II. RELATED WORKS

This section covers works that are related to the development and application of UAVs as well as how deep learning has progressed in the area of image segmentation. The paper mentions how UAVs can be used to provide cost effective communication[9] for devices without the need to invest into infrastructure. These are ready to deploy solutions and can be flexibly configured. The concept of autopiloting[10] for small unmanned aerial vehicles is discussed here. The research focuses on using a smaller lightweight version of the original bulky UAV to introduce flight basics and how the sensors must work in unison without any human intervention. In the end they discussed the open source autopilot systems. Application of UAV in smart farming is discussed in [11]. The machines mingles with various other domains to make this happen like Artificial Intelligence, Big Data and Internet of Things. Their application includes seed sowing, timely pesticide spraying and constantly monitoring the growth.

Now coming to the advancements in deep learning in segmentation, the paper[12] discusses the various deep learning techniques that are used for semantic segmentation and their application in indoor navigation, autonomous driving. It also discusses the datasets used for the purpose and the bottlenecks. Finally they mention about what all changes can be done to the current model to create a new start of the model by taking into all the fallacies of the previous models into consideration. Deep learning is instrumental in the medical field and this is proved in this paper which mentions the usage of deep learning for segmentation of deep brain regions in MRI and ultrasound. CNN employs a Hough voting mechanism that follows segmentation and localisation of anomalies. Different dimensionality data is taken for this research and compares the efficiency in each class.

Identifying brain tumours early is very important to save a person's life and that is discussed in [13], which uses deep learning based models for brain tumour segmentation. Fully Convolutional Neural Networks(FCNNs) and Conditional Random Fields(CRF)are the algorithms used to achieve this. Three segmentation models are developed using 2D image patches and slices obtained in axial, coronal and sagittal views respectively, and combine them to segment brain tumours using a voting based fusion strategy. Apart from medical imaging, image segmentation also finds use in the road, marker, sign board identification for autonomous vehicles. The paper mentions the StixelNet which is a deep Convolutional Neural Network for obstacle detection and road segmentation. Monocular dataset KITTI is used for training the model and results show how it outperforms other models that use the 3D data from the camera.

Sensor fusion by aggregating all the real time data from the sensors and improving the accuracy of the road segmentation is discussed in [14]. A novel method of fusing LIDAR point cloud and image is explained which promises good end to end segmentation is researched in the paper. Pyramid projection method is used to increase the accuracy of the multi-scale LIDAR map.

III. MATERIALS AND METHODS

This section mentions in detail about the dataset and architecture used in this proposed method to achieve the required objective.

The dataset used here is the UC Merced L and Use dataset[15]. The dataset was compiled and released by the University of California. It contains 21 classes with 100 aerial images in each class. Each image measures 256x256 pixels. The images were manually extracted from large images from the USGS National Map Urban Area Imagery collection for various urban areas around the country. The pixel resolution of this public domain imagery is 1 foot. For this work, only the Runway sub data will be used to train out a deep net model. Each image is subjected to augmentation(translation , scaling) which increases the size of each class up to 500 images. The data set is then split into - train, validation and test. Train contains 350 images, validation 100 images and finally 50 images. A sample from the dataset is shown in fig 1.

Table 1 Different classes of UC Merced Land Use dataset

Agricultural	Airplane	Baseball diamond	Beach	Buildings	Chaparral	Dense residential
Forest	Freeway	Golf course	Harbor	Intersection	Medium residential	Mobile home park
Overpass	Parking lot	RIver	Runway	Sparse residential	Storage tanks	Tennis court

Figure 1 A sample runway image from the runway subclass in UC Merced Land Use dataset

Here from the aerial image, our region of interest would be the runway strip only. Since the strip is clearly demarcated from the surrounding ground patches via a line, we can use that to differentiate the road and then segment it. In image segmentation, the pixels that are under consideration are grouped together. So essentially when comparing with object detection, image segmentation gives much more information like its shape.

There are two different methods to image segmentation, they are : Semantic segmentation and Instance segmentation. In the first method, it's kind of like a binary approach wherein every pixel belonging to the region of interest will be given the same colour while the rest will be given black. In the next method, each pixel will be assigned to a particular class but the different objects present under the same class will have different colours. The figure shows the difference between the two methods.

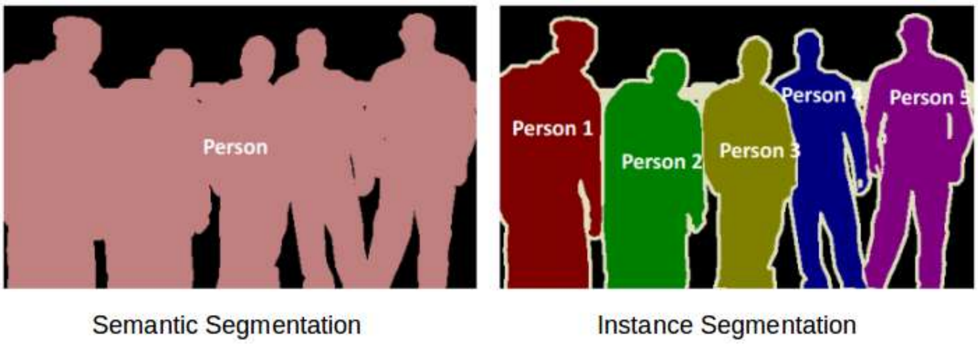

Figure 2 Difference between semantic and instance segmentation

To train the model, we have to prepare a labeled dataset. The input will be the greyscale images of the runway in from the dataset and output will be the binary masks of corresponding input images. The binary mask is a semantic segmented version of the input image and is created using a technique called Region based segmentation. In this method, we fix a threshold and pixels that are above the threshold will fall into the region of interest and will be given a value 1 which is white color while the pixels that fail to cross the threshold are considered as the background pixels which are of no interest. Such pixels will be assigned a value 0, meaning black color. This type of segmentation which relies on thresholding is called Threshold segmentation.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4054-4061 www.ijaem.net ISSN: 2395-5252

Table shows the two classes of the threshold segmentation - Global and Local. In Global thresholding, we divide the image into two regions by defining a single threshold value. In Local thresholding multiple thresholds are specified to split the image into multiple regions along with background.

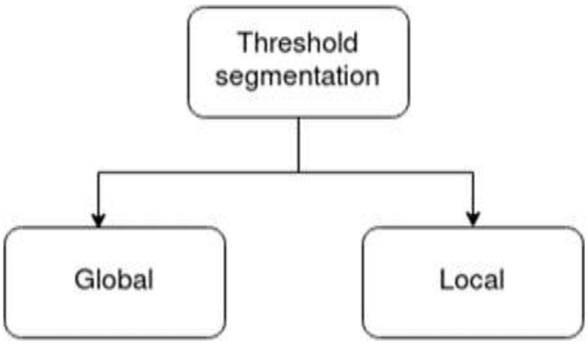

Figure 3 Different methods for threshold segmentation

Global thresholding is used to create a binary segmented mask for each input grayscale image. Pixels values greater than 128 are assigned 255 which is white and rest are assigned 0 which is black.

This is done using the OpenCV library in python and the code is shown in Figure 4 and the corresponding output of our input image from Figure 1 is shown in Figure 5.

```
for i in os.listdir(path+'/train/'):
    dim = (256, 256) #(w,h)
    image = cv2.imread(path+'/train/'+i, 0)
    resized = cv2.resize(image, dim, interpolation=cv2.INTER_AREA)
    cv2.imwrite(path+'/images_train/' + i,resized)

    resized = cv2.resize(image, dim, interpolation=cv2.INTER_AREA)
    (thresh, im_bw) = cv2.threshold(resized, 128, 255, cv2.THRESH_BINARY)
    cv2.imwrite(path+'/mask_train/' + i, im_bw)
```

Figure 4 Sample python code snippet to do threshold segmentation

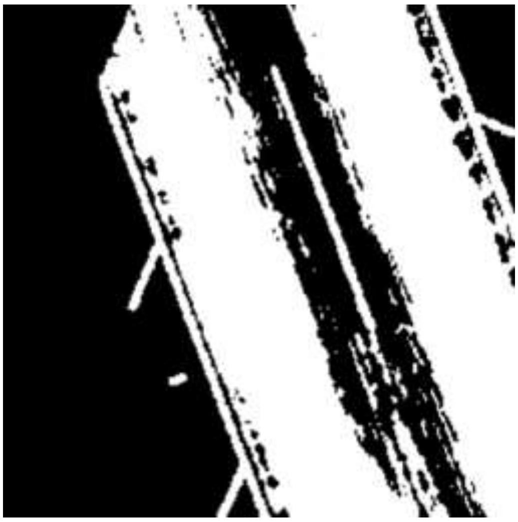

Figure 5 Output segmented image after threshold segmentation

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4054-4061 www.ijaem.net ISSN: 2395-5252

As per our calculation, in the created binary mask only our region of interest which is the road portion in the runway is segmented. Now we give the labelled dataset as input to the deep Convolutional Neural Network model U-net. U-net[16] came into light back in 2015 when they were used to process biomedical images. These were used to find out the tumor in lungs their shape and size through semantic segmentation. The U-net architecture is shown in the Figure 6 . The architecture comprises a contracting path (left side) and an expansive path(right). The contracting path comprises two convolutional layers followed by a max pooling layer. The convolution process increases the depth of the image which will give rise to the increase in number of channels from 1 to 64. The first two convolution layers are followed by a max pooling layer which reduces the size of the image to half(568x568 to 284x84). This same process is repeated three timestill we get to the flat bottom of the architecture. At this stage, there are only just two convolutional layers without any max pooling layer. At this stage, the size of the image becomes 28x28x1024.

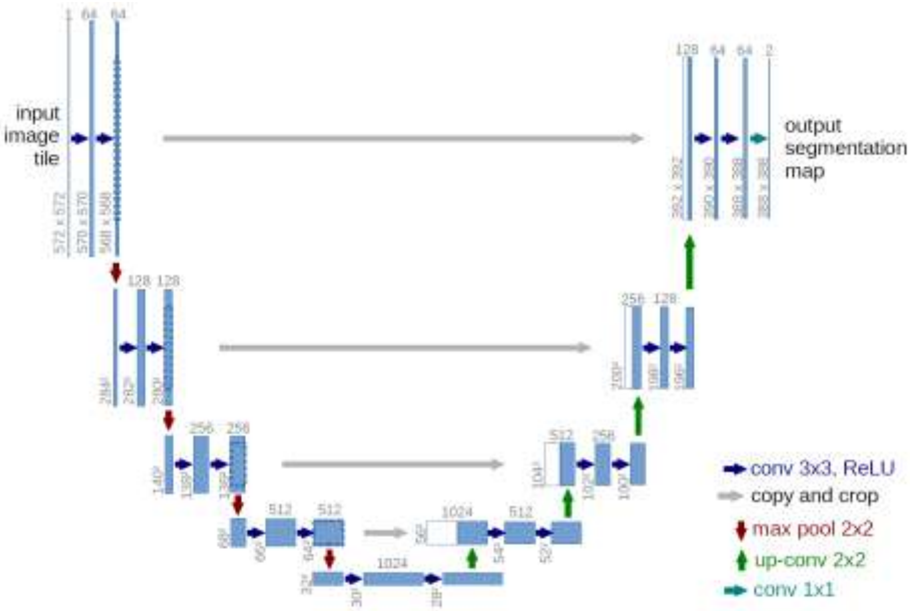

Figure 6 U-Net architecture - contracting path and expansive path

After covering the contracting path, the image then moves to the expansive path. Throughout this path the image will at each step be converted back and finally reach its original dimensions. This stage can also be called as upsampling as the image gets resized to original. This stage consists of transposition, concatenation followed by two convolutional layers. The transposed convolution is an upsampling method that expands the size of images. Soon after this the image get upsized from 28x28x1024 to 56x56x512, following which some concatenation is done to match the size of the image in the contracting path. So the size now becomes 56x56x1024. This process is repeated three more times to reach the top last layer.The energy function does pixel wise softmax computation to classify the target classes and the loss function used here is cross entropy loss function.

$$p_k(\mathbf{x}) = \exp(a_k(\mathbf{x}))/\left(\sum_{k'=1}^{K} \exp(a_{k'}(\mathbf{x}))\right)$$

The proposed architecture in the paper is shown in Figure 7. Here as we have previously discussed, we take the input data and then make two copies. One would be a grayscale version of the input images and other will be the binary mask of all the input images that has been subjected to threshold

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4054-4061 www.ijaem.net ISSN: 2395-5252

segmentation. Both use the grayscale and thresholding functionalities in OpenCV to achieve this. Then the images are split into train, test and validation sets. Normal python code is used to split the images. After this the U-net is created using keras. Each layer is created and then the loss function, metrics are mentioned in thecompile section of the code. Finally the model is trained using the fit function where in tensorflow acts as a backend and executes the code. The neural network will learn from this pixel wise segmentation and it will be capable to do segmentation on any unknown image. Finally the model weights are saved in a h5 file format so that it can be loaded for further testing purposes.

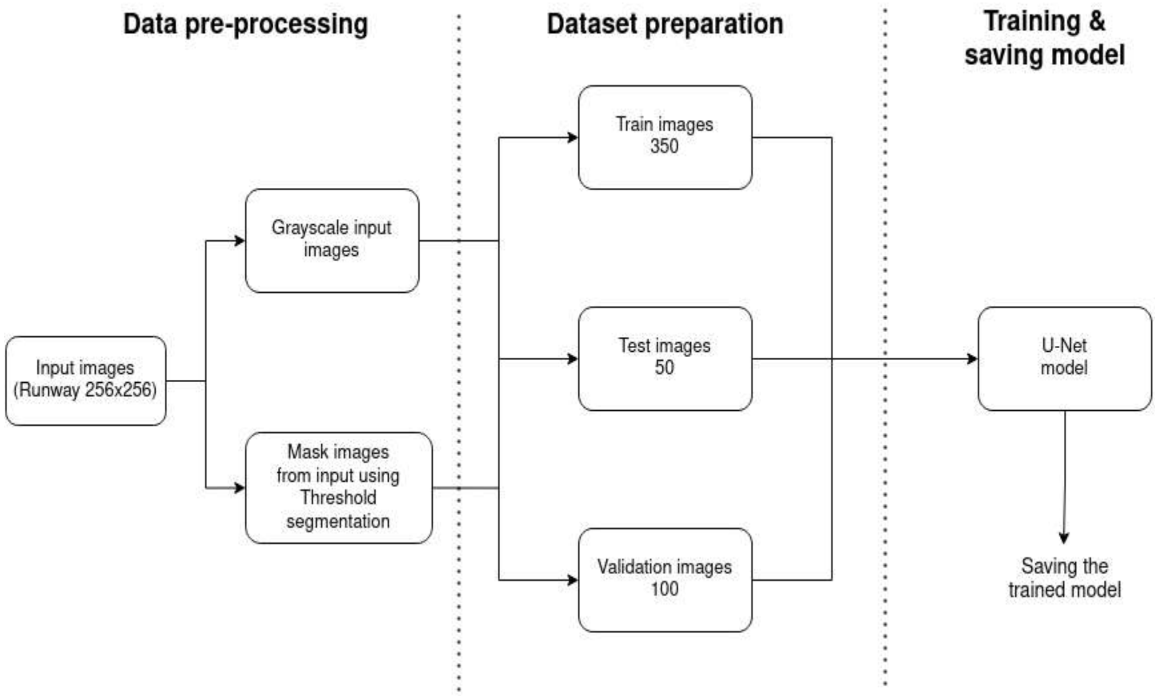

Figure 7 Proposed architecture

IV. RESULTS AND DISCUSSION

The model was trained for 30 epochs and training was completed after 3 hours. The Google Colab GPU instance was used to train the model. Figure 8 shows the training logs in the googlecolab.The model training accuracy was 85% and validation accuracy was about 81.25%. The Figure 9 shows the train vs validation loss. The Figure 10 shows the output segmented image when an unknown runway image is given as input to the trained U-Net model.

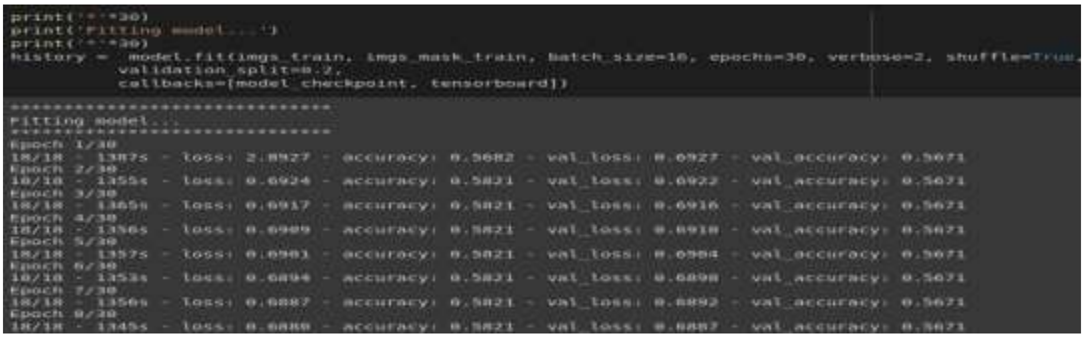

Figure 8 Model training output log from Google colab

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4054-4061 www.ijaem.net ISSN: 2395-5252

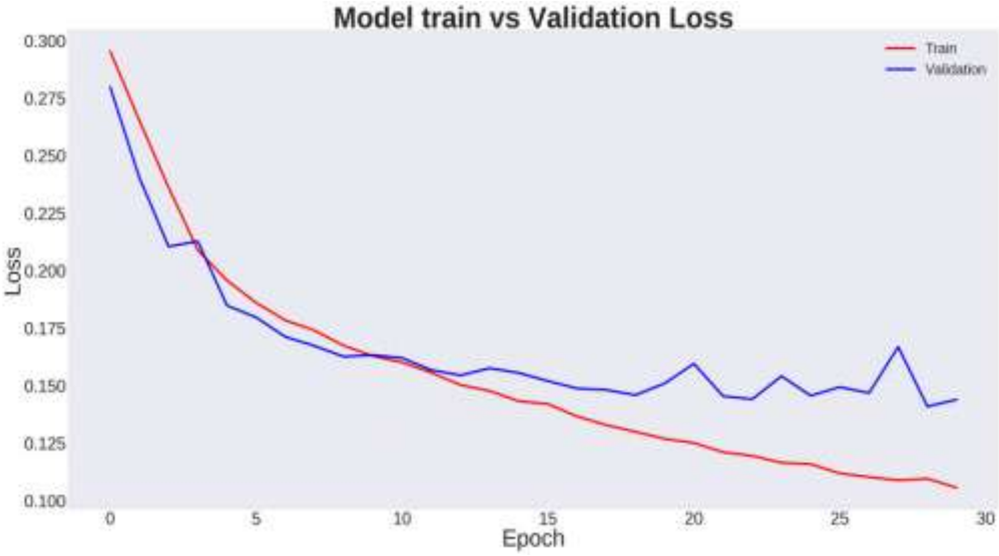

Figure 9 The model train vs validation loss plot

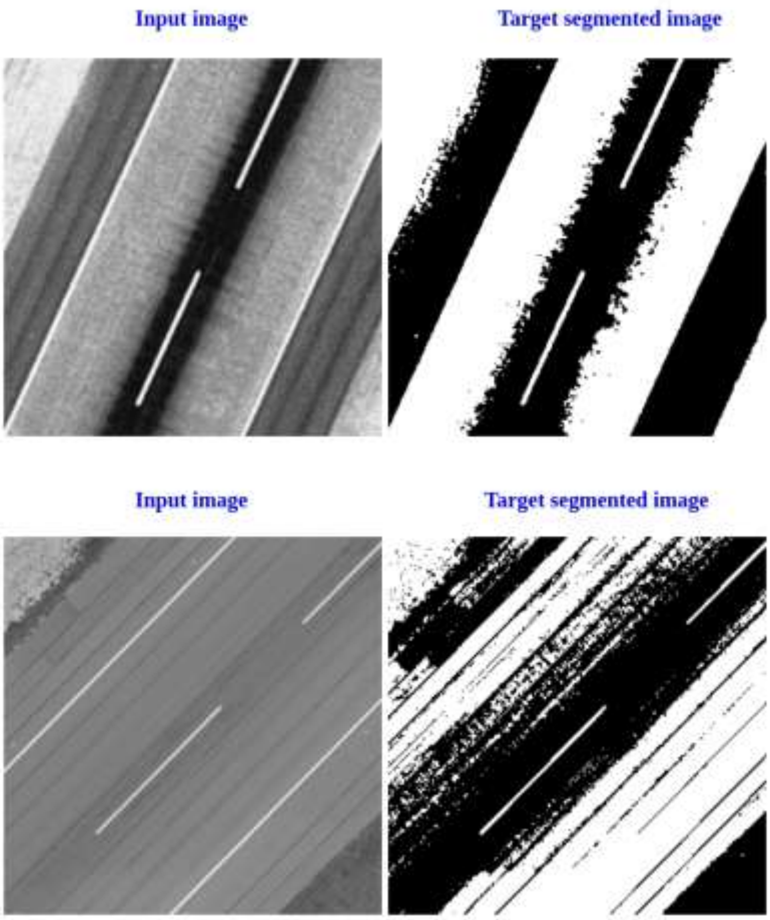

Figure 10 The input image and output segment comparison

International Journal of Advances in Engineering and Management (IJAEM)
Volume 3, Issue 7 July 2021, pp: 4054-4061 www.ijaem.net ISSN: 2395-5252

V. CONCLUSIONS

The U-net which was developed with focus on biomedical application performs surprisingly well in other domains. In this paper, the deep learning based method was able to understand the intricate details of how well the pixels were classified into two different regions. In future work, instead of threshold based segmentation other advanced methods including automatic thresholding as well as using Instance segmentation to detect different parts of the runway roads will be used to check the output accuracy.

REFERENCES

[1]. Gevaert, C.; Sliuzas, R.; Persello, C.; Vosselman, G. Evaluating the Societal Impact of Using Drones to SupportUrban Upgrading Projects.ISPRS Int. J. Geo-Inf.2018,7, 91

[2]. Lottes, P.; Khanna, R.; Pfeifer, J.; Siegwart, R.; Stachniss, C. UAV-based crop and weed classification for smartfarming. InProceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA); IEEE:Singapore, 2017; pp. 3024–3031.

[3]. Feng, Q.; Liu, J.; Gong, J. Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing andRandom Forest Classifier—A Case of Yuyao, China.Water2015,7, 1437–1455

[4]. Manfreda, S.; McCabe, M.; Miller, P.; Lucas, R.; Pajuelo Madrigal, V.; Mallinis, G.; Ben Dor, E.; Helman, D.;Estes, L.; Ciraolo, G.; et al. On the Use of Unmanned Aerial Systems for Environmental Monitoring.Remote Sens.2018,10, 641.

[5]. Woo, J., Son, K., Li, T., Kim, G. S., &Kweon, I. S. (2007, May). Vision-based UAV Navigation in Mountain Area. In MVA (pp. 236-239).

[6]. Narayanan, Ram Gopal Lakshmi, and Oliver C. Ibe. "Joint network for disaster relief and search and rescue network operations." Wireless Public Safety Networks 1. Elsevier, 2015. 163-193.

[7]. Bernard, M., Kondak, K., Maza, I., &Ollero, A. (2011). Autonomous transportation and deployment with aerial robots for search and rescue missions. Journal of Field Robotics, 28(6), 914-931.

[8]. Kazim, M., Azar, A. T., Koubaa, A., Ibrahim, Z. F., Zaidi, A., & Zhang, L. (2021). Event-driven programming-based path planning and navigation of UAVs around a complex urban environment. In Unmanned Aerial Systems (pp. 531-565). Academic Press.

[9]. Fotouhi, A., Qiang, H., Ding, M., Hassan, M., Giordano, L. G., Garcia-Rodriguez, A., & Yuan, J. (2019). Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges. IEEE Communications Surveys & Tutorials, 21(4), 3417-3442.

[10]. Chao, HaiYang, YongCan Cao, and YangQuan Chen. "Autopilots for small unmanned aerial vehicles: a survey." International Journal of Control, Automation and Systems 8.1 (2010): 36-44.

[11]. Lottes, P., Khanna, R., Pfeifer, J., Siegwart, R., &Stachniss, C. (2017, May). UAV-based crop and weed classification for smart farming. In 2017 IEEE International Conference on Robotics and Automation (ICRA) (pp. 3024-3031). IEEE.

[12]. Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857.

[13]. Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., ... &Glocker, B. (2017, September). Ensembles of multiple models and architectures for robust brain tumour segmentation. In International MICCAI brainlesion workshop (pp. 450-462). Springer, Cham.

[14]. Zhang, R., Candra, S. A., Vetter, K., &Zakhor, A. (2015, May). Sensor fusion for semantic segmentation of urban scenes. In 2015 IEEE international conference on robotics and automation (ICRA) (pp. 1850-1857). IEEE.

[15]. Yi Yang and Shawn Newsam, "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification," ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010.

[16]. Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.