Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 55-22 (2022) 310–315

Semantic segmentation based mapping systems for the safe and precise landing of flying vehicles

Harsimret Singh Dhami [*] **Dmitry Ignatyev** [**]
Antonios Tsourdos [***]

[*] *Research Student, Cranfield University, Bedford MK430AL UK (e-mail:harsimret.dhami@cranfield.ac.uk).*
[**] *Senior Research Fellow in Autonomous Systems and Control,School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK430AL UK (e-mail:D.Ignatyev@cranfield.ac.uk)*
[***] *Professor,Head of the Autonomous and Cyber-Physical Systems Centre,School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK430AL UK (e-mail:a.tsourdos@cranfield.ac.uk).*

Abstract: Unmanned Aerial Systems (UAS) are a promising technology for many areas, including transportation, agriculture, inspection, and rescue missions. However, to enable a high level of autonomy,including Beyond Visual Line of Sight (BVLOS) flight, the drones should be able to perform safe landings in unknown areas without an operator. Hence there is a need for development of safe landing methods for autonomous drones.The autonomous UAVs can often be operated more economically than the conventional manned aircraft. As technology advances, autonomous UAVs are expected to play an increasingly important role in a variety of industries and applications.In this paper we have explored a semantic segmentation-based approach for the problem of autonomous landing.

Keywords: Semantic segmentation, Computer vision, Aerial image segmentation, Unmanned aerial Vehicles (UAVs), Autonomous Landing

1. INTRODUCTION

The use of unmanned aerial vehicles (UAVs) or drones has increased significantly in recent years, especially in the commercial sector.With the rapid expansion of the drone industry, there are concerns about the safety and regulation of drones in the airspace. In the years to come, the Beyond the Visual Line of Sight (BVLOS) concept will expand the commercial use of drones compared to today's use. UAVs, or unmanned aerial vehicles, are aircraft that are operated remotely or autonomously. Autonomous UAVs are a rapidly growing technology with a range of potential applications from surveying and mapping to parcel delivery and search and rescue. One of the main advantages of autonomous UAVs is that they can be operated without a human pilot on board. This not only reduces the risk to human life, but also allows UAVs to be used in a wider range of situations and environments. In addition, autonomous UAVs can often be operated more cheaply than conventional manned aircraft. As technology advances, autonomous UAVs are expected to play an increasingly important role in a variety of industries and applications.(Liu et al. (2022))

1.1 The Landing Problem

Autonomous landing is a challenging point for a UAV, which need to be resolved as a matter of urgency. Locating a suitable landing area, sensing the position between landing platform and and the UAV, and then seeking the occasion to land are the three steps in the process of landing of an UAV. The GPS has been applied to provide position and velocity of UAV as control feedback. Since the accuracy of GPS receivers for use on UAV must be measured in meters, they are unsuitable for precision tasks such as landings.

1.2 The Proposed Solution

The aim of this work is to make the UAV identify suitable landing place in its operation path. The characteristics that define a suitable landing place are:

(1) Doesn't cause any injury to any living being in the vicinity.
(2) Any property isn't damaged.
(3) The damage to the UAV is minimized (Daniel (2007))

In Daniel (2007) based on the "size", "shape", "surface" and "slope" criteria, for finding safe landing sites these steps were discussed:

(1) Segmenting the image
(2) Identifying sites with suitable size and shape
(3) Classifying the type of surface

10.1016/j.ifacol.2023.03.052

1.3 Related works

Selecting safe landing sites is a key development of autonomous UAVs. The existing methods have the common problems of poor generalization ability and robustness.Their performance is significantly degraded, and the error cannot be self-detected and corrected. These methods have poor performance due to less diverse data.

The existing methods for autonomous landing site detection can be broadly classified into two categories:

(1) Camera Based:
A UAV can locate a landing zone in a variety of methods utilising computer vision. These techniques recognise the ground environment using either monocular or binocular vision cameras. Techniques for selecting known and unknowable zones in both indoor and outdoor contexts can be done using vision-based landing zone identification methods.In Bruno and Colombini (2021) the author addresses the possibility of a robot to localize itself in an unknown environment and simultaneously build a consistent map of this environment this is called SLAM. This can be achieved using cameras and is effective in navigation in GPS denied locations.In the paper Wubben et al. (2019) UAV is equipped with a low-cost camera that can detect ArUco markers. After the marker is detected, the UAV alters its flight behaviour to land on the exact position where the marker is located.

(2) LiDAR Based:
The distance between two objects can be calculated using the remote sensing technique known as light detection and ranging (LiDAR).Pulsed laser beams are directed to the target region on the ground by a UAV equipped with LiDAR. The region of the earth it encounters reflects the light beam.The author in Chen et al. (2020) had a UAV system equipped with low-cost LiDAR and binocular camera to realize autonomous landing in by detecting the flat and safe ground area. The paper Scherer et al. (2012) also has a LiDAR based approach.In the paper they created and constructed a 3-D scanning LiDAR with two modes of operation: forward scanning for detecting obstacles during low-altitude flying and downward scanning for mapping the ground and finding landing zones from a higher height.

The proposed solution in the paper Leung et al. (2022) takes RGB image, LIDAR point cloud and robot motion information as inputs, and outputs a fused traversability cost map that is computed from both terrain types and geometrical properties.This paper provides a significant growth in autonomy level in off-road ground vehicles.They evaluated the proposed framework with synchronised sensor data captured while driving the robot in real off-road environments.

2. TYPES OF LANDING ZONES

In Shah Alam and Oluoch (2021) the author classifies the various types of landing zones into Indoor and Outdoor landing zones as shown in the Figure 1. Indoor landing zones are flat and static zones.While, the outdoor zones can be dynamic or static depending on the various factors involved. Static and dynamic outdoor landing zones can be divided further into two types the known and unknown zones. Runways that are marked, Helipads and surfaces with some distinct symbols which can be used to identify them are some examples of known static landing zones. On the other hand, the flat surfaces like roadside, field, and flat roof which are free of any unwanted materials or obstacles are some examples of unknown static landing zones. The surfaces on ship, truck or other moving vehicles that are marked and have flat surface are some examples of dynamic known landing zones.(Kaljahi et al. (2019))

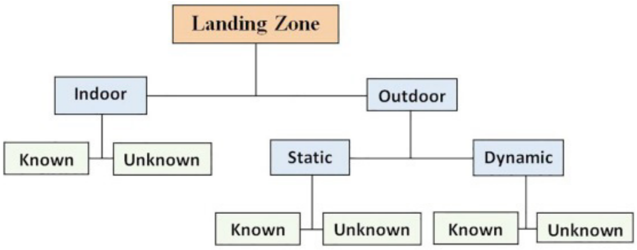

Fig. 1. Various Types of landing zones. Adopted from [Shah Alam and Oluoch (2021)]

3. SEMANTIC SEGMENTATION

There are a lot of methods for image classification like bounding boxes, semantic segmentation,etc.Bounding boxes are frequently used for object identification and image classification. A semantic segmentation technique, on the other hand, can produce pixel-perfect accuracy and expose fine-grained details about the contents of an image. This the reason why we chose sematic segmentation method over the bounding box method.The principle of semantic segmentation of an image is to assign every pixel of a picture taken by a camera with a corresponding category label (class). (Liu et al. (2022)) Semantic segmentation requires classification of all pixels within the plane of image and implementation of the image to be performed at the same time, causing a degradation in resolution in the resulting image. For the widely deployed and existing framework SegNet, it has shown high performance when detecting a small pixel target composed of small and simple features while up-sampling the compressed feature map while traversing multiple pooling layers while the complicated background is adequately controlled. However, when many complicated features need to be detected for on a complicated background, resolution degradation inevitably occurs. Thus, SegNet using a standard stacked convolutional neural network is inadequate to comprehensively learn the local features related to the whole image needed for a high-performance semantic segmentation model and all location information.(Cho and Jung (2022))

We selected a few existing architectures to train and test our dataset.

3.1 DeeplabV3

The most recent version of Deeplab's image semantic segmentation models, v3+, is the most advanced available. The use of atrous spatial pyramid pooling (ASPP) operation at the encoder's end is its main contribution. (Chen et al. (2017b))

3.2 Resnet

An artificial neural network called a residual neural network (ResNet) (ANN). Additionally, Control Neural Network uses it. It is a gateless or open-gated variation of the HighwayNet, which was the first functionally complete, extremely deep feedforward neural network with hundreds of layers—much deeper than earlier neural networks. (He et al. (2015))

3.3 DenseASPP

Atrous Spatial Pyramid Pooling (ASPP)Chen et al. (2017a) was proposed to concatenate multiple atrous-convolved features using different dilation rates into a final feature representation. Although ASPP is capable of producing multi-scale features, we contend that the scale-axis feature resolution is insufficient for the autonomous driving scenario. In order to achieve this, we suggest Densely connected Atrous Spatial Pyramid Pooling (DenseASPP), which densely connects a set of atrous convolutional layers to produce multi-scale features that not only cover a wider scale range, but also do so densely, all while maintaining a relatively small model size.(Yang et al. (2018))

4. DATASET

Selecting a dataset is a crucial task before the training of Semantic segmentation models. After careful thought and evaluation of the data acquisition strategy and object-class selection for annotation we selected the UAVid Dataset.We selected the UAVid Dataset after comparison with various well-known existing semantic segmentation datasets.The CamVid dataset which is one of the most commonly used datatset. It has 701 images, each of size 960 × 720, If we compare in terms of number of Pixels this is a lot smaller than our dataset. The Cityscapes dataset which is an important dataset for urban scenes. it has 5000 images of size 2048×1024, This is much bigger than the size of UAVid dataset. However, the objects in the images of UAVid dataset are smaller than in Cityscapes dataset, providing more object variance in the same number of pixels, which compensate for the object recognition task. The size of images is quite large in the Potsdam dataset i.e., 6000 × 6000.(Lyu et al. (2020))

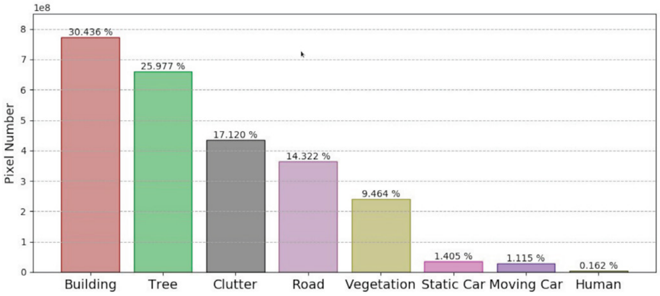

Fig. 2. The Pixel number histogram of various classes in our Dataset

4.1 Dataset size

The UAVid dataset (Lyu et al. (2020)) is a high-resolution UAV dataset which is designed for semantic segmentation tasks. The main focus of this dataset is urban scenes shot from an UAV. It has 300 images, each with a size of 4096 × 2160 or 3840 × 2160. Buildings, roads, trees, low vegetation, static cars, moving cars, humans, and background clutter are the eight classes selected for semantic segmentation task. The official data split for the UAVid dataset, i.e., 15 training scenes (It has 150 labelled images) and 5 validation scenes (50 labelled images) for training and validation, respectively. The test split consists of the left 10 scenes (100 labelled images).

4.2 Class Definition and Analysis

Labelling each object in a 4K resolution dataset which is based in an urban environment is consumes a lot of time and is expensive.The most common types of objects are labelled for in the dataset so as to ease the process.There are a total of 8 classes in the dataset.The UAVid Dataset that we used had a diverse class composition as shown in figure 2. The definition of each class is described as follows.

(1) Building: houses, skyscrapers and also including buildings under construction. walls and fences are not included.
(2) Road: road or bridge surface that cars can operate on legally. The places where vehivles can be parked are excluded.
(3) Tree: trees that have canopies and main trunks.
(4) Low vegetation: grass, small shrubs, and bushes.
(5) Static car: cars that are not moving, buses that are not moving , trucks, automobiles. two wheeler vehicles are not included.
(6) Moving car: cars that are moving, including moving buses, trucks, automobiles, and tractors. two wheeler vehicles are excluded.
(7) Human: pedestrians, bikers, and other humans
(8) Background clutter: all objects not belonging to any of the classes above.(Lyu et al. (2020))

5. EXPERIMENT

5.1 System

The task of training a semantic segmentation model requires a lot of computational power. So, for the task we used Delta HPC (High Performance Computing), it is a supercomputer at Cranfield University. The computer has two "front-end" login nodes. These are known as delta-login-1 and deltalogin-1. They each contain two Intel E5-2620 v4 (Broadwell) CPUs giving 16 CPU cores and have a total of 256 GB of shared memory.

The compute nodes are housed in standard rack mount 6U chassis, with 12 nodes per 6U chassis. There are 10 chassis for compute nodes spread over five racks. This gives a total of 120 compute nodes. 118 nodes are for general use and each of these nodes has two Intel E5-2620 v4 (Broadwell) CPUs giving 16 CPU cores and 128GB of shared memory. Taken together these give a total of 1888 available cores.

There are two GPU nodes housed within one of the standard rack mounts 6U chassis, each node has two Intel E5-2620 v4 (Broadwell) CPUs giving 16 CPU cores, and 128GB of shared memory. One node has 4 Tesla K80 GPU cards, and the other node has 4 V100 cards.

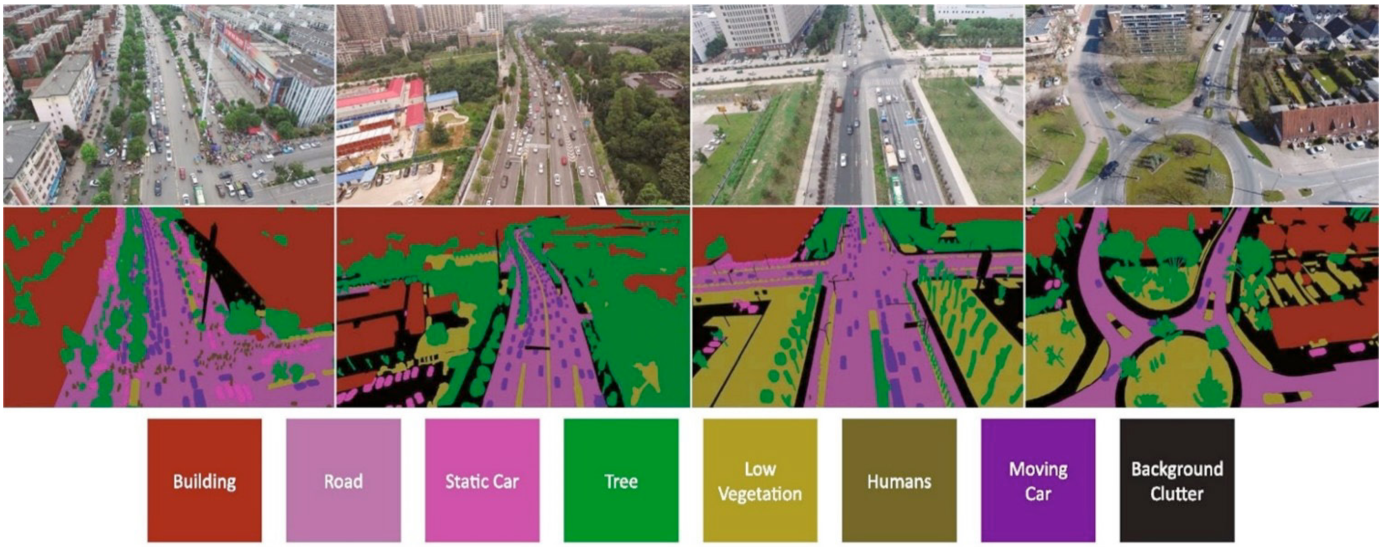

Fig. 3. Various classes in the UAVid Dataset (Lyu et al. (2020))

5.2 Training

The machine learning libraries that we used for our paper are listed below.

Libraries

(1) TensorFlow: A complete open-source machine learning platform is called TensorFlow. Researchers can advance the state-of-the-art in ML thanks to its extensive, adaptable ecosystem of tools, libraries, and community resources, while developers can simply create and deploy ML-powered apps.

(2) Matplotlib: Python's Matplotlib toolkit provides a complete tool for building static, animated, and interactive visualisations. Matplotlib makes difficult things possible and simple things easy.

(3) OpenCV: A set of programming tools called OpenCV is primarily focused on real-time computer vision.

5.3 Evaluation Metrics

The following evaluation metrics standard for classification methods were used.

(1) Intersection Over Union(IoU):
The Intersection over Union (IoU) metric is essentially a way to measure the amount of overlap between our prediction output and the target mask.The IoU metric measures the number of pixels common between the target and prediction masks divided by the total number of pixels present across both masks.

$$IoU = \frac{T_P}{T_P + F_P + F_N} \tag{1}$$

$T_P = True\ Positives$
$T_N = True\ Negatives$
$F_P = False\ Positives$
$F_N = False\ Negatives$

(2) Avg. Accuracy:
One parameter for assessing classification models is accuracy. The percentage of predictions that our model correctly predicted is known as accuracy.The average of accuracy across the image is considered.

$$\frac{T_P + T_N}{T_P + F_N + T_N + F_P} \tag{2}$$

(3) Precision:
The proportion of accurately categorised positive samples (True Positive) to the total number of positively classified samples is known as precision.

$$Precision = \frac{T_P}{T_P + F_P} \tag{3}$$

(4) F1 Score:
By calculating the harmonic mean of a classifier's precision and recall, the F1-score integrates both into a single metric. It mainly used to compare the effectiveness of two classifiers.

$$F1\ Score = 2(\frac{Precision \times Recall}{Precision + Recall}) \tag{4}$$

$$Recall = \frac{T_P}{T_P + F_N} \tag{5}$$

5.4 Training the Model

We trained our models in the high-performance computer. It took around 9-10 hours to train our model to 200 epochs. We trained two different models. The specifications of the models are elaborated further in the paper.

(1) Model 1:
The 1st model that we trained was a DeeplabV3 model with Resnet 50 frontend. It provided us the following metrics. As the images in the dataset were quite big (4096 × 2160 or 3840 × 2160) we cropped them into 1080 × 1920 and then began the training.We selected 100 epochs with a batch size of 10.

(2) Model 2:
The 2nd model that we trained was a DenseASPP model with Resnet 50 frontend. It provided us the

following metrics. As the images in the dataset were quite big (4096 × 2160 or 3840 × 2160) we cropped them into 2560 × 1440 and then began the training. We selected 200 epochs with a batch size of 10.

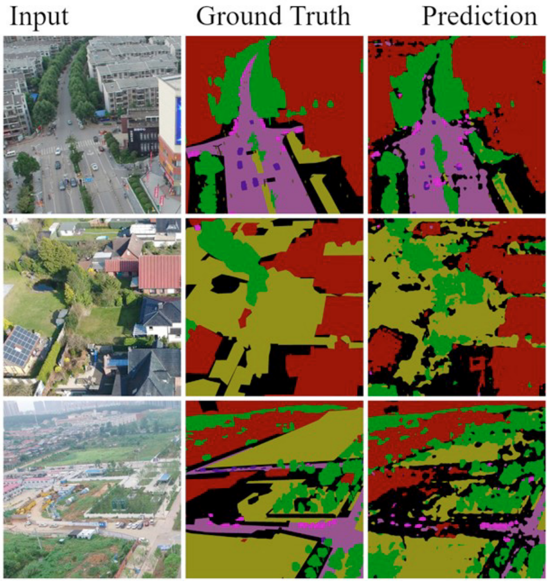

Fig. 4. Predictions of Model 1

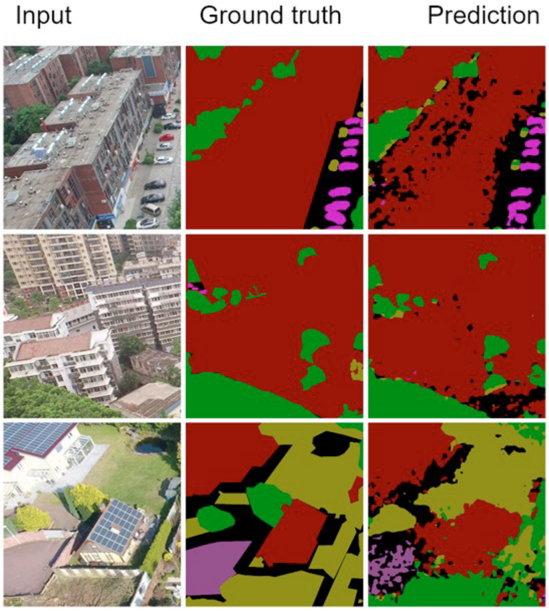

Fig. 5. Predictions of Model 2

6. DISCUSSIONS

Model	Avg. Accuracy	Precision	F1 Score	Mean IOU
Model 1	0.62979	0.69487	0.60854	0.31632
Model 2	0.73284	0.77983	0.73284	0.4154

Table 1. Evaluation results

We presented a semantic segmentation approach for identifying the landing areas.The models we trained provided us with promising results. The figure 4 Shows the predictions of the model 1 and the figure 5 shows the predictions of model 2 In these figure we can see that on the left there is Input Image,in the center the ground truth and on the right is the prediction of our models. We can see that in the prediction there is a lot of noise(black colour). However,this can be further improved by training.The noise in model 2 less than model 1.As in model 2 we had taken 200 epochs while they were 100 in model 1 and also taken a bigger input image of size 2560 × 1440 rather than 1920 × 1080 in the model 1. These results can also be improved with further training.

Class	Model 1	Model 2
Background clutter	0.50139	0.58976
Building	0.70461	0.82515
Road	0.63151	0.63797
Tree	0.69069	0.70734
Low vegetation	0.57367	0.61167
Moving car	0.77507	0.78286
Static car	0.34426	0.42309
Human	0.61922	0.63094

Table 2. Class-wise accuracy of our models

The numerical metrics for determining the accuracy and evaluation of our models are shown in table 1. One can see from the table,that the model 2 provides better results according to all the evaluation parameters.The table 2 provides the class-wise accuracy of our models .The Class “Building” has the highest pixel wise composition in the Dataset hence it has the highest accuracy.The accuracy can be further improved with increasing the number of iterations i.e. number of epochs. As our Dataset has 4K Resolution (4096 × 2160 or 3840 × 2160). For ease of training, we cropped the input into 1920 × 1080 and 2560 × 1440 then passed it into our model.It helped us increase our dataset and improve the accuracy.

7. CONCLUSION AND FUTURE SCOPE

BVLOS flight is a Holy Grail for the UAV industry. A decision making algorithm for drone safe landing is an enabling technology. In our research, we proposed to use a semantic segmentation for analysis of landing sites. We compared two different models, namely, a DeeplabV3 model with Resnet 50 frontend and a DenseASPP model with Resnet 50 frontend. Our results manifest that both models give promising results, however, the latter exhibiting better performance. In the future we plan to integrate our model into a real time semantic segmentation module and deploy it on an UAV onboard a NVIDIA Jetson Nano.We will try to increase the accuracy of our models by collecting additional data and increasing the dataset size and diversity. Including images with new types of areas, e.g. forests, farms or land fields, as well as images, collected at different lighting and weather conditions might improve the robustness of the algorithm. Combining semantic segmentation with various other deep-learning techniques might also help to improve prediction accuracy at the expense of additional computational resources. Building on this paper we plan to develop a vision-based navigation system for UAVs in GPS denied locations. We Plan to implement this as real time video semantic segmentation model.

REFERENCES

Bruno, H.M.S. and Colombini, E.L. (2021). Lift-slam: A deep-learning feature-based monocular visual slam method. *Neurocomputing*, 455, 97–110. doi:10.1016/j.neucom.2021.05.027.

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A.L. (2017a). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *arXiv:1606.00915 [cs]*. URL `http://arxiv.org/abs/1606.00915`.

Chen, L.C., Papandreou, G., Schroff, F., and Adam, H. (2017b). Rethinking atrous convolution for semantic image segmentation. *arXiv:1706.05587 [cs]*. URL `http://arxiv.org/abs/1706.05587`.

Chen, L., Yuan, X., Xiao, Y., Zhang, Y., and Zhu, J. (2020). Robust autonomous landing of uav in non-cooperative environments based on dynamic time camera-lidar fusion. *arXiv:2011.13761 [cs]*. URL `https://arxiv.org/abs/2011.13761`.

Cho, S. and Jung, Y. (2022). Semantic segmentation-based vision-enabled safe landing position estimation framework. *AIAA SCITECH 2022 Forum*. doi:10.2514/6.2022-1475.

Daniel, L.F. (2007). Landing site selection for uav forced landings using machine vision.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *arXiv:1512.03385 [cs]*. URL `http://arxiv.org/abs/1512.03385`.

Kaljahi, M.A., Shivakumara, P., Idris, M.Y.I., Anisi, M.H., Lu, T., Blumenstein, M., and Noor, N.M. (2019). An automatic zone detection system for safe landing of uavs. *Expert Systems with Applications*, 122, 319–333. doi:10.1016/j.eswa.2019.01.024.

Leung, T.H.Y., Ignatyev, D., and Zolotas, A. (2022). Hybrid terrain traversability analysis in off-road environments. In *2022 8th International Conference on Automation, Robotics and Applications (ICARA)*, 50–56. doi:10.1109/ICARA55094.2022.9738557.

Liu, F., Shan, J., Xiong, B., and Fang, Z. (2022). A real-time and multi-sensor-based landing area recognition system for uavs. *Drones*, 6, 118. doi:10.3390/drones6050118.

Lyu, Y., Vosselman, G., Xia, G.S., Yilmaz, A., and Yang, M.Y. (2020). Uavid: A semantic segmentation dataset for uav imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 165, 108–119. doi:10.1016/j.isprsjprs.2020.05.009.

Scherer, S., Chamberlain, L., and Singh, S. (2012). Autonomous landing at unprepared sites by a full-scale helicopter. *Robotics and Autonomous Systems*, 60, 1545–1562. doi:10.1016/j.robot.2012.09.004.

Shah Alam, M. and Oluoch, J. (2021). A survey of safe landing zone detection techniques for autonomous unmanned aerial vehicles (uavs). *Expert Systems with Applications*, 179, 115091. doi:10.1016/j.eswa.2021.115091.

Wubben, J., Fabra, F., Calafate, C.T., Krzeszowski, T., Marquez-Barja, J.M., Cano, J.C., and Manzoni, P. (2019). Accurate landing of unmanned aerial vehicles using ground pattern recognition. *Electronics*, 8, 1532. doi:10.3390/electronics8121532.

Yang, M., Yu, K., Zhang, C., Li, Z., and Yang, K. (2018). Denseaspp for semantic segmentation in street scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

2023-04-17

Semantic segmentation based mapping systems for the safe and precise landing of flying vehicles

Dhami, Harsimret

Elsevier

Dhami HS, Ignatyev D, Tsourdos A. (2022) Semantic segmentation based mapping systems for the safe and precise landing of flying vehicles, IFAC-PapersOnLine, Volume 55, Issue 22, 2022, pp. 310-315

https://doi.org/10.1016/j.ifacol.2023.03.052