Mühendislik Bilimleri ve Tasarım Dergisi
10(4), 1141 – 1154, 2022
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.1087477

Araştırma Makalesi

Research Article

U-NET BASED CAR DETECTION METHOD FOR UNMANNED AERIAL VEHICLES

Oğuzhan KATAR[1*], Erkan DUMAN[2]

[1] Firat University, Faculty of Technology, Department of Software Engineering, Elazig, Turkey
[2] Firat University, Faculty of Engineering, Department of Computer Engineering, Elazig, Turkey

Keywords

UAV,
U-Net,
Detection,
Image Processing,
Semantic Segmentation.

Abstract

With the developments in computer hardware technology, studies in the fields of computer vision and artificial intelligence has accelerated. However, the number of areas where autonomous systems are used has also increased. Among these areas are unmanned aerial vehicles, which are one of the most important parameters of today's military technology. In this study, which includes two different scenarios, we aimed to improve the vision capabilities of unmanned aerial vehicles based on artificial intelligence. Within the scope of Scenario-1, the U-Net model suitable for binary semantic segmentation method was trained with the help of images taken by unmanned aerial vehicle camera. Within the scope of Scenario-2, which is designed for moving or stationary vehicle detection, the U-Net model is trained in accordance with multi-class semantic segmentation method. In all these training processes, a publicly available dataset was used. The model trained for Scenario-1 reached mean Intersection over Union (mIoU) value of 84.3%, while the model trained for Scenario-2 reached 79.7% mIoU. In this study, approaches were shared about the use of high-resolution images in model training and testing stages. Applying such studies in the field can help improve precision and reliability in arms industry.

İNSANSIZ HAVA ARAÇLARI İÇİN U-NET TABANLI ARAÇ TESPİT YÖNTEMİ

Anahtar Kelimeler

İHA,
U-Net,
Tespit,
Görüntü İşleme,
Anlamsal Bölütleme.

Öz

Bilgisayar donanımı teknolojisindeki gelişmelerle birlikte bilgisayar görmesi ve yapay zeka alanlarındaki çalışmalar hız kazanmıştır. Bununla birlikte otonom sistemlerin kullanıldığı alanların sayısı da artmıştır. Bu alanlar arasında günümüz askeri teknolojisinin en önemli parametrelerinden biri olan insansız hava araçları yer almaktadır. İki farklı senaryoyu içeren bu çalışmamızda insansız hava araçlarının görüş yeteneklerini yapay zeka tabanlı olarak geliştirmeyi hedefledik. Senaryo-1 kapsamında ikili anlamsal bölütleme yöntemine uygun U-Net modeli sadece araç objesinin tespitini yapabilmek için insansız hava aracı kamerasıyla çekilen görüntüler yardımıyla eğitilmiştir. Hareketli veya durağan araç tespiti için tasarlanan Senaryo-2 kapsamında, U-Net modeli çok sınıflı anlamsal bölütlemeye uygun olarak eğitilmiştir. Tüm bu eğitim süreçlerinde kamuya açık veri seti kullanılmıştır. Senaryo-1 kapsamında eğitilen model %84,3 ortalama birleşim üzerinden kesişme (mIoU) değerine ulaşırken, Senaryo-2 kapsamında eğitilen model %79,7 mIoU değerine ulaşmıştır. Bu çalışmada yüksek çözünürlüklü görüntülerin model eğitiminde ve test aşamalarında kullanılabilmesi hakkında yaklaşımlar paylaşıldı. Bu tür çalışmaların sahada uygulanması, savunma sanayisinde hassaslığı ve güvenirliği iyileştirmeye yardımcı olabilir.

Alıntı / Cite

Katar, O., Duman, E., (2022). U-Net Based Car Detection Method for Unmanned Aerial Vehicles, Journal of Engineering Sciences and Design, 10(4), 1141-1154.

Yazar Kimliği / Author ID (ORCID Number)	**Makale Süreci / Article Process**	
O. Katar, 0000-0002-5628-3543	**Başvuru Tarihi / Submission Date**	14.03.2022
E. Duman, 0000-0003-2439-7244	**Kabul Tarihi / Accepted Date**	14.04.2022
	Yayım Tarihi / Published Date	30.12.2022

[*] İlgili yazar / Corresponding author: okatar@firat.edu.tr,+90-424-607-4301

1. Introduction

Unmanned aerial vehicles are complex robots that can fly autonomously or remotely with the help of ground stations (Nonami et al., 2010). In today's technology, unmanned aerial vehicles are used in various fields due to their size, cost and mobility (Boukoberine et al., 2019). Unmanned aerial vehicles take part in almost every part of our lives, such as military operations, traffic monitoring, search and rescue, and photography (Howard et al., 2018).

Thanks to the developments in computer hardware technology, the number of technologies in which autonomous systems are used has also increased. The main purpose of developing an autonomous system is to minimize the human factor (Shareef et al., 2021). Artificial intelligence-based solutions are used for this function. The main benefits of artificial intelligence-based solutions; minimizing decision-making time and margin of error. One of the technologies that needs less margin of error and decision-making time is unmanned aerial vehicle technology (Mohamed et al., 2020). For this reason, it is inevitable to use artificial intelligence supported autonomous systems in unmanned aerial vehicle technology.

By integrating the camera and other necessary hardware parts, unmanned aerial vehicles can be given the ability to visual abilities (Kuru, 2021). Today, the vision functions of unmanned aerial vehicles are used in the tasks such as monitoring the target vehicle and destroying it when necessary, during the observations made for the collection of intelligence information (Haulman, 2003). In the target detection phase, unmanned aerial vehicles usually identify the target vehicles within a predetermined rectangular area (Xu et al., 2016). Sometimes the area occupied by the target vehicles on the image does not exactly match the dimensions of the defined rectangular area. For this reason, the margin of error is accepted as the difference between the area occupied by the rectangular area and the space occupied by the vehicle object in the detection of vehicle objects. This acceptance reduces the sensitivity of the task to be performed and may negatively affects the decision-making. The ability to vision alone is not sufficient for the decision-making mechanism. In order for the computer to decide on its own, it needs to analyse the image it receives from its camera. Traditional image processing methods can be used in this analysis process, but today, advanced models based on deep learning are preferred (Dargan et al., 2020). Deep learning models can efficiently process large amounts of data, thanks to high-powered graphics processing units (Blumberg et al., 2018). A large number of samples are needed in the training of deep learning models. In addition to segmentation studies, mask images are required. Creating mask images is very costly in terms of time. For this reason, finding a public dataset is a big challenge for researchers. Many datasets have been shared in order to eliminate this difficulty for researchers (Du et al., 2018; Mueller et al., 2016). In this study, we used a public dataset, including mask images, by customizing it. Thanks to our models suitable for binary and multi-class semantic segmentation, pixel-based marking capability can be added to the unmanned aerial vehicle. In this way, the system will create the relevant environment for the interpretation of the image and taking the relevant action in return, without the need for any human intervention. With such approaches, undesirable events caused by human factors such as carelessness, demoralization, and wrong decision-making can be prevented.

The main purpose of this study is to provide autonomous unmanned aerial car with the ability to detect vehicle objects with deep learning-based segmentation methods. With its application in the field, it is aimed to increase efficiency and precision in traffic monitoring and military operations. The rest of this paper is organized as follows. Section 2 includes other studies in the literature. The details of the models, dataset and performance metrics is described in Section 3. The analysis of test results and experimental results of U-Net models are given in Section 4. Conclusion part of the study is in Section 5.

2. Related Works

In this section, some studies from computer vision in car detection are examined. The main idea behind these studies is to contribute to the computer vision systems with various methods. Zhao and Nevatia (2001) proposed a system to detect vehicle objects from aerial images. Various edge extraction algorithms were used in this study. Ammour et al. (2017) proposed a deep learning-based car detection method. This method is a two-stage method that includes the extraction and classification of candidate regions. The researchers used the mean-shift algorithm to extract the candidate regions. They used support vector machine (SVM) as a classifier with deep convolutional network. They classified the regions on the image as “car” and “no-car”. For feature extraction, the VGG16 model with an input size of 224x224 pixels (px) was used. Xu et al. (2017) trained the Faster R-CNN model for traffic monitoring with the help of low altitude unmanned aerial vehicles. Faster R-CNN achieved 98.43% correctness and 96.40% completeness. Hinz and Stilla (2006) used infrared images to detect stationary and moving vehicles. The car detection approach consists of three main parts. Predicts are made on the image using the differences in temperature levels. 96 cars were extracted from the image reserved for the test, 90 of which were real positive and the remaining 6 were false alarms.

3. Material and Method

3.1. Dataset

In this study, the public dataset called UAVid was used (Lyu et al., 2020). The relevant dataset was created with the help of cameras integrated into DJI phantom3 pro and DJI phantom4 drones. The drones were flown at a maximum flight speed of 10 m/s to record images in RGB format. Reason for keeping the flight speed so low is to prevent the possible blurring effect in the recorded images.

The dataset consists of 42 different folders, each containing 10 image frames. Randomly selected samples from these 420 high-resolution images are given in Figure 1.

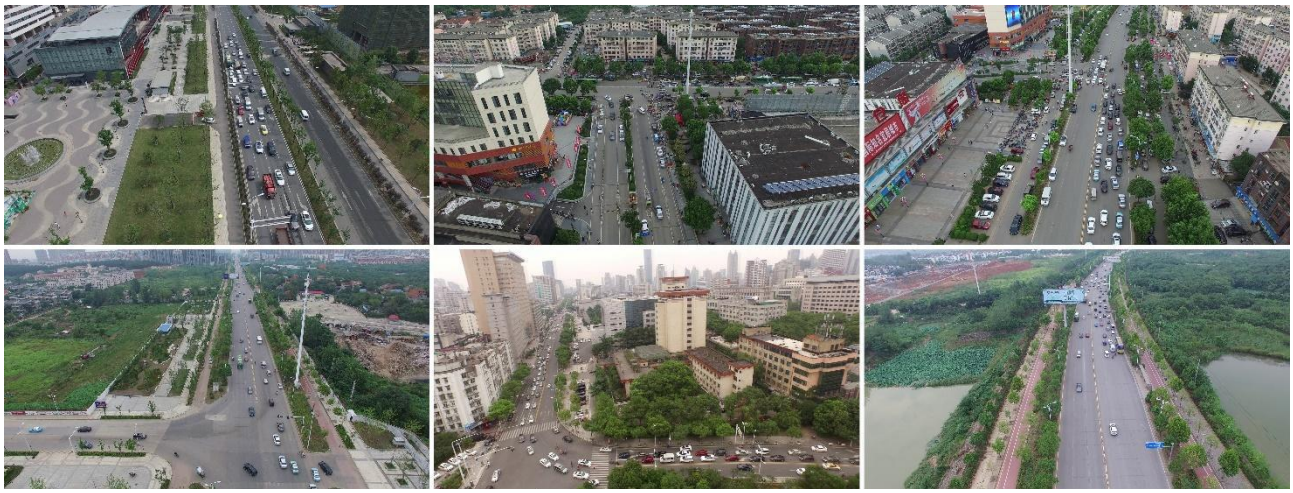


Figure 1. Randomly selected samples from dataset

Different classes were determined by the researchers to be used in segmentation studies and mask images were created. In these mask images, each class is defined by different pixel values. Samples of masks in the dataset are given in Figure 2.

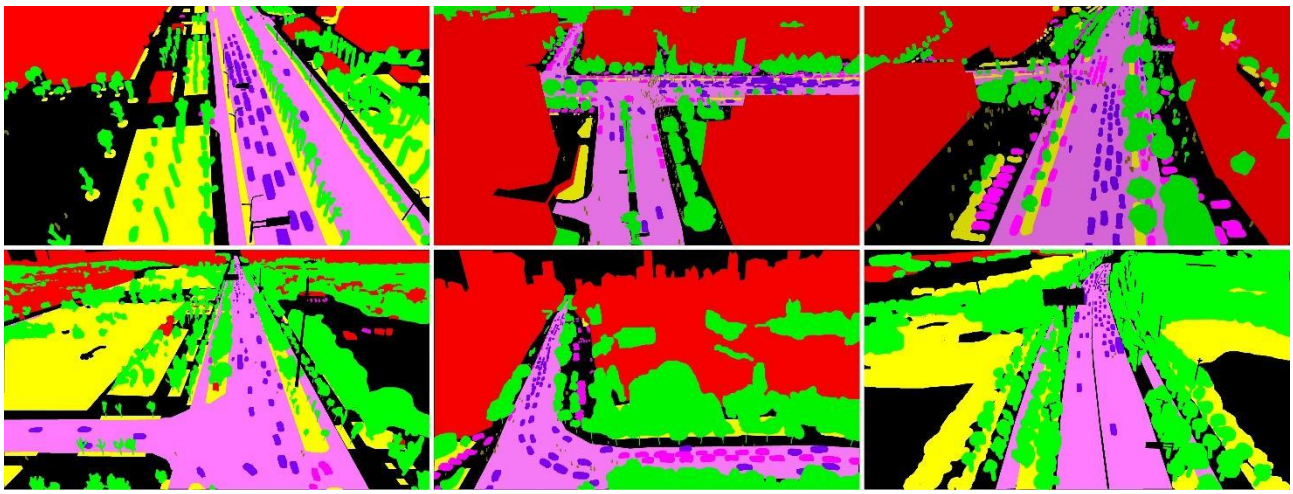


Figure 2. Samples of masks in the dataset

In the dataset, which includes 420 images in total, ground truth mask images were created for only 270 images. The remaining 150 images are recommended by the researchers to be used in the testing phase. While 170 of the 270 images created with mask images have a resolution of 3840x2160px, 100 of them have a resolution of 4096x2160px. Labelling on images with such high resolution is very costly in terms of time. For this reason, only 8 classes were determined as namely building, road, tree, low vegetation, static car, moving car, human, and clutter and mask images in RGB format were labelled. Because mask images are in RGB format, each pixel is represented by 24 bits, so different pixel values can be easily assigned to each class without conflict. Besides the advantages of the mask images in RGB format, there are also disadvantages. An example of these disadvantages is that the size of a mask image occupies in the memory is large and therefore increases the processing time in the processes in

which it will be used. Pixel values, labels and percentage of frames containing pixels with the label information are given in Table 1.

Table 1. Dataset pixel details

Label name	Pixel value (RGB)	In frames (%)
building	(128,0,0)	30.44
road	(128,64,128)	14.32
tree	(0,128,0)	25.98
low vegetation	(128,128,0)	9.46
static car	(192,0,192)	1.41
moving car	(64,0,128)	1.12
human	(64,64,0)	0.16
clutter	(0,0,0)	17.12

3.2. Scenario-1: Car Detection Based on Semantic Segmentation

Using the UAVid dataset, Scenario-1 was designed to give unmanned aerial vehicles the ability to recognize the car object. The main purpose in this scenario is to provide more sensitive approaches with artificial intelligence-based segmentation method in functions such as target detection and tracking performed by autonomous systems. In this context, before the model was created and its training started, various pre-processes were applied in the dataset. The common objectives of the pre-processes applied are to use our resources more efficiently.

3.2.1 Pre-processing for Scenario-1

The UAVid dataset created for the multi-class segmentation problem should be made suitable for binary segmentation. Mask images in RGB format were converted to 8-bit single-channel '.png' format. The result of the corresponding operation is given in Figure 3.

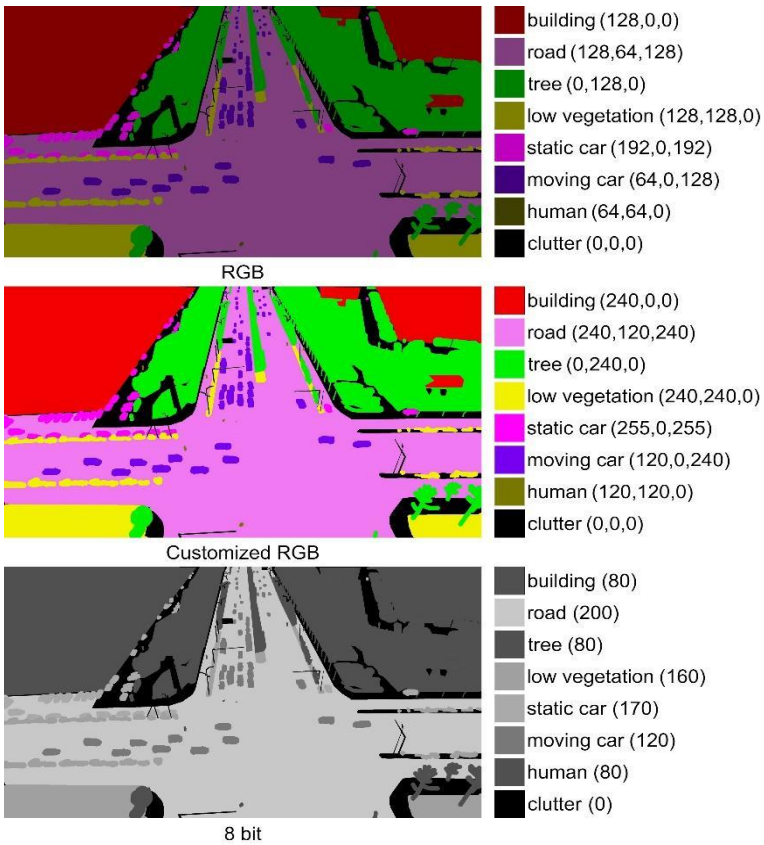


Figure 3. Pre-processing for Scenario-1

After all mask images were converted to 8-bit format, the value of "255" was assigned to only the pixels where the vehicle object was located, and the value of the other pixels was changed to "0" using the pixel-based value changing method. In this way, the necessary structure for binary segmentation was created. Result of the pixel-based value changing method is given in Figure 4.

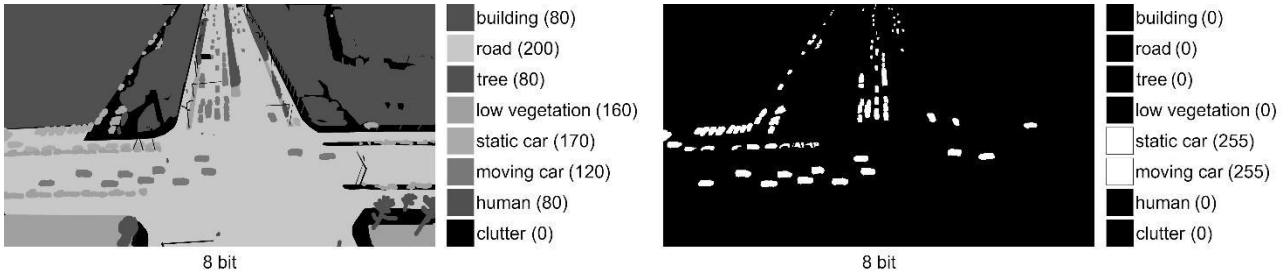


Figure 4. Pixel-based value changing method for Scenario-1

3.3. Scenario-2: Moving or Stationary Car Detection Based on Semantic Segmentation

Using the UAVid dataset, Scenario-2 was designed to give unmanned aerial vehicles the ability to recognize moving or stationary car objects. In this scenario, the main purpose is to enable the separation of moving or stationary vehicles in functions such as target detection and tracking performed by autonomous systems to be realized by artificial intelligence-based segmentation method, and to enable different functions or parameters to be used for relevant situations. Similar pre-processing steps were applied as in Scenario-1.

3.3.1 Pre-processing for Scenario-2

The UAVid dataset created for the multi-class segmentation problem; should be made suitable for multi-class segmentation, which covers only 3 classes as background, moving vehicle and stationary vehicle. Mask images in RGB format were converted to 8-bit single-channel '.png' format. The result of the corresponding operation is given in Figure 5.

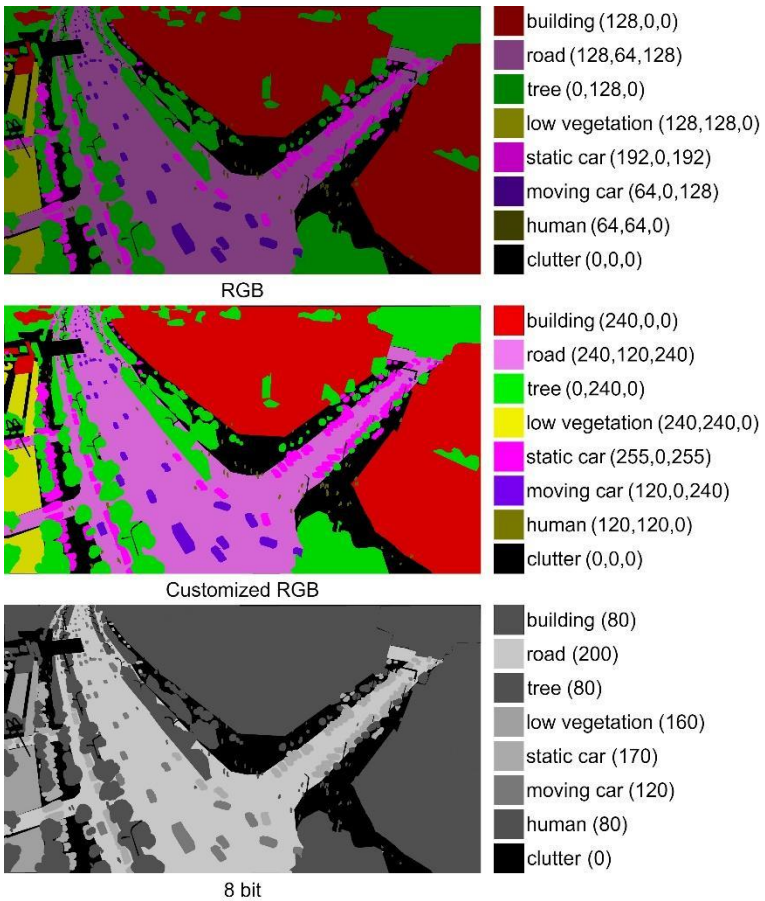


Figure 5. Pre-processing for Scenario-2

After all mask images were converted to 8-bit format, pixels with moving vehicle objects were assigned a value of "1", pixels containing stationary vehicle objects were assigned a value of "2" and all other pixels were assigned a value of "0" by pixel-based value changing method. In this way, a multi-class segmentation structure with the least number of class labels was created. Result of the pixel-based value changing method for Scenario-2 is given in Figure 6.

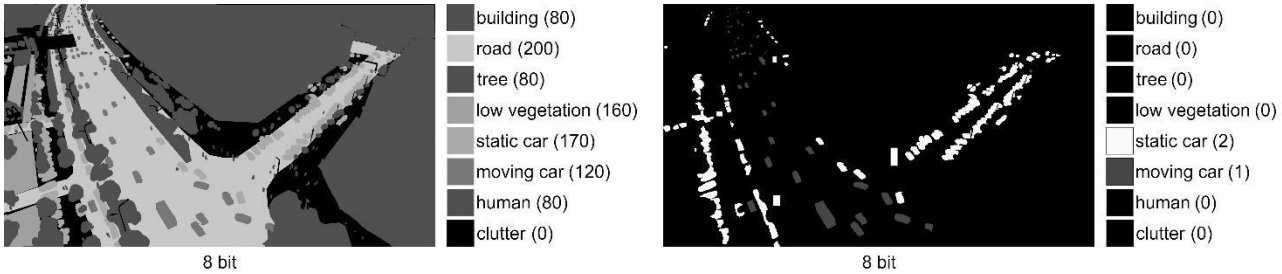


Figure 6. Pixel-based value changing method for Scenario-2

3.4. Image Cropping

As it is known, the resolution values of the dataset elements are not constant for all elements, and they are one of the values 3840x2160px or 4096x2160px. In addition, considering the hardware features we have, it is not possible to use it directly in model training without any action. In order to avoid this problem, resizing can be considered as a solution but as a result of this process, information loss in high resolution images will be quite high. Another approach is to split the image into sub-images. The resolution of the sub-images to be created is determined according to the input size value of the segmentation model that is planned to be used.

First to create a sub-image, the reference image to be cropped is determined. 3840x2160px image or 4096x2160px is cropped in a square format, starting from the first pixel, with edges on the X-axis of 256 pixels and the Y-axis of 256 pixels. This process is repeated until it covers the size values of the image. In Figure 7., the cropping process and numbering of rows and columns are given.

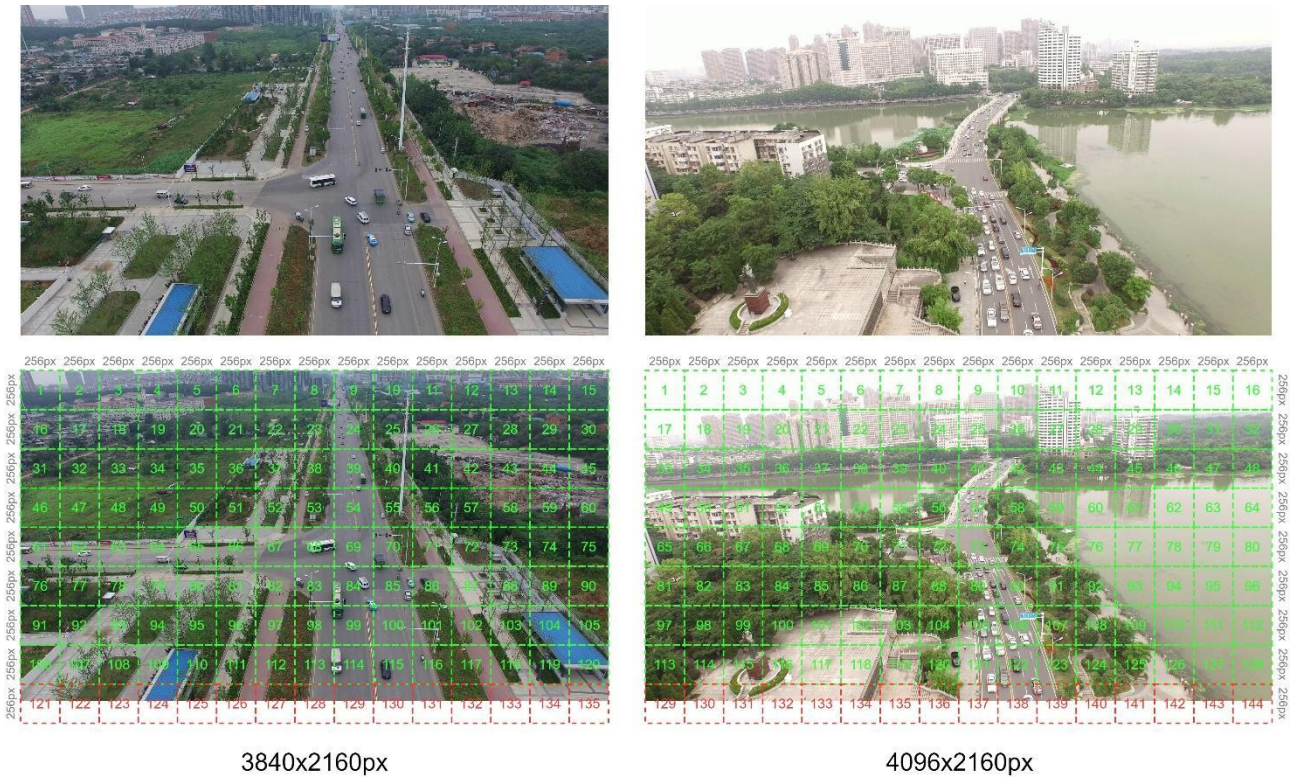


Figure 7. Image cropping method

When the sub-image division process is completed, two different sub-image types occur, which can complete the 256x256px size and cannot. In Figure 7., sub-images that can complete 256x256px are indicated in green, while images that cannot complete are indicated in red. The divided sub-images indicated in red may not be included in the model training by ignoring them but in order not to lose the information in these areas and to benefit from this

information in model training, it was ensured that the dimensions were changed to 256x256px by padding methods. The padding method applied for 3840x2160px images and its result are given in Figure 8.

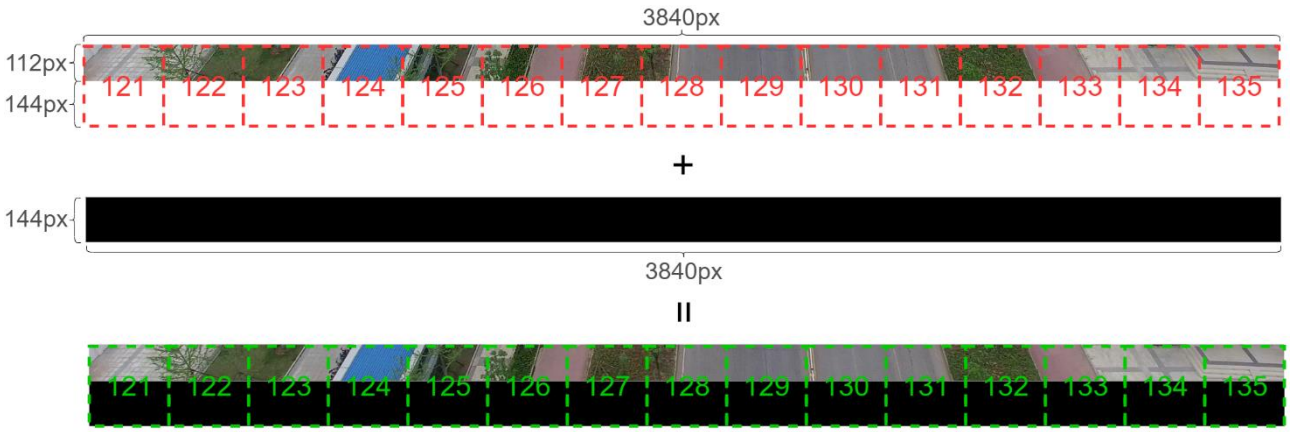


Figure 8. Padding methods for 3840x2160px images

The same procedures were applied for images with 4096x2160px size. The only difference is that the size of the added piece is 4096x144px. The result of the padding method applied for 4096x2160px images is given in Figure 9.

Figure 9. Padding methods for 4096x2160px images

When all steps were completed, an image of 3840x2160px was divided into 135 sub-images of 256x256px, and an image of 4096x2160px was divided into 144 sub-images of 256x256px. Since the same steps are applied to mask images, the specified numerical amounts are also valid for mask images. Examples of the padding method applied for mask images are given in Figure 10.

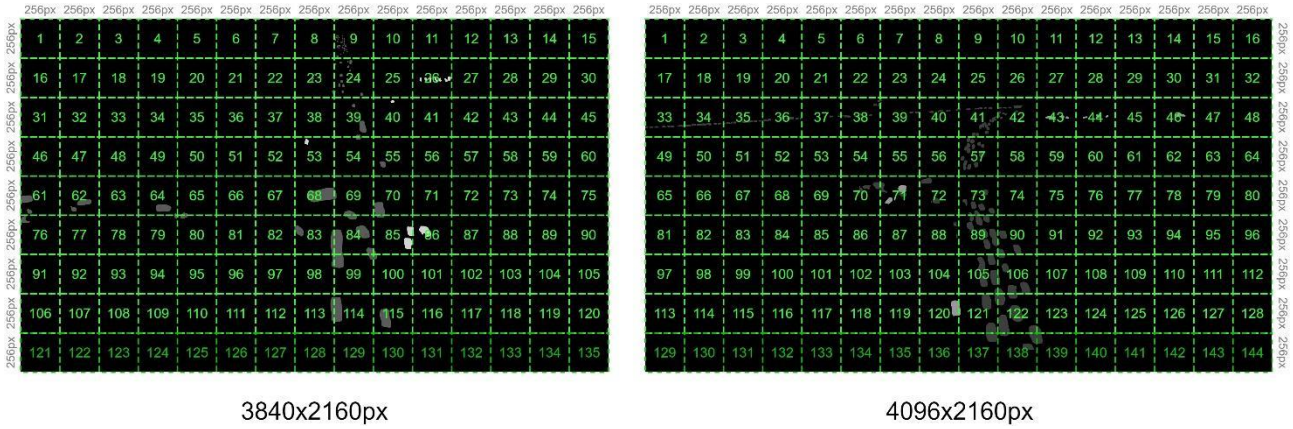


Figure 10. Image cropping and padding methods for masks

There are frames in which there is no meaningful information among the 256x256px size mask images. Samples defined as no meaningful information correspond to frames only with a pixel value of '0' and where the car object is not included. These frames can be detected by image-based pixel distribution calculation and may not be included in model training. If the specified process is realized, training time will be reduce, but it is not preferred because it will reduce the number of samples used in model training. After the image cropping process is completed, the samples counts are given in Table 2.

Table 2. Dataset samples counts

Resolution	İmages (256x256px)	Masks (256x256px)
3840x2160px	22950	22950
4096x2160px	14400	14400

3.5. Data Splitting

Considering the unity of the images and mask images in the dataset, it was randomly divided into 70% train, 20% validation and 10% test. Data splitting is shown in Figure 11.

Figure 11. Data splitting

3.6. U-Net Model

U-Net is a kind of artificial neural network that contains a series of convolutional layers and non-convolutional layers. U-Net, which can be used in any semantic segmentation application today, is basically designed for biomedical images. U-Net gets its name from its architecture similar to the letter U, as can be seen in Figure 12.

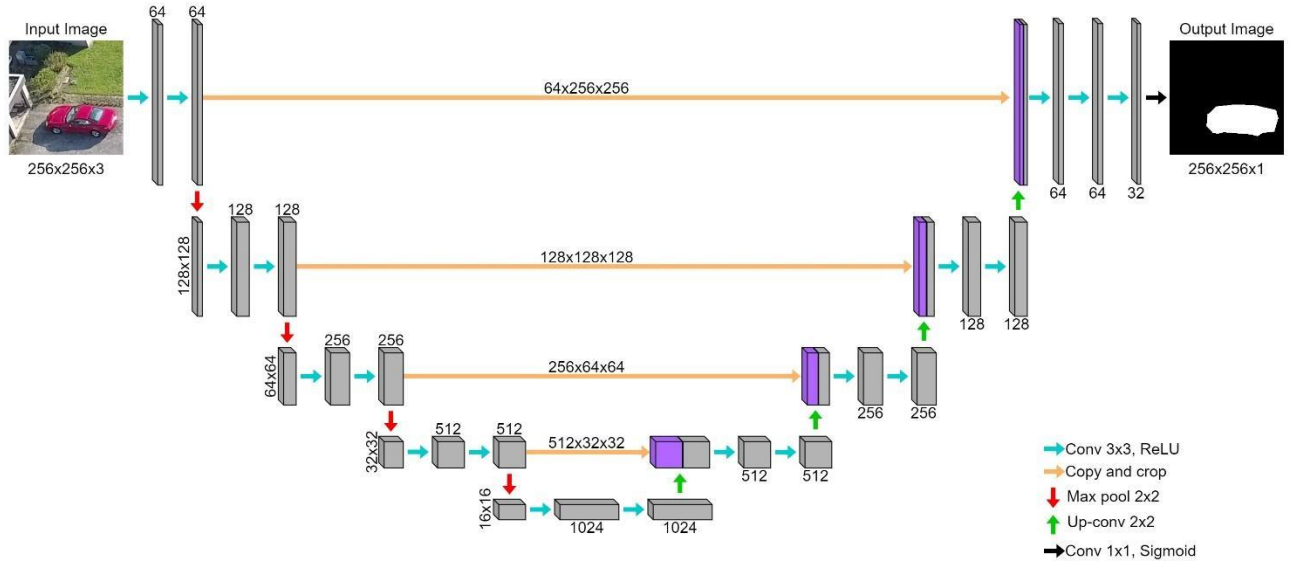


Figure 12. U-Net architecture

U-Net network architecture consists of a contracting path and an expansive path. The contracting path is called encoder network. The encoder network acts as a feature extractor and learns an abstract representation of the input image through a sequence of encoder blocks. The expansive path is called decoder network. Decoder network is used to take the abstract representation and create a semantic segmentation mask.

3.6.1 Encoder Network

Each encoder block consists of two 3×3 convolutions where each convolution is followed by a Rectified Linear Unit (ReLU) activation function. ReLU introduces non-linearity to the network, which helps to better generalize the training data. Then comes the 2×2 maximum pooling, in which the height and width of the feature maps are reduced by half. This reduces the calculation cost by reducing the number of trainable parameters.

3.6.2 Decoder Network

The decoder block starts with a 2×2 transposed convolution. It is then combined with the corresponding feature map from the encoder block. These links provide features from previous layers that are sometimes lost due to the depth of the network. Two 3×3 convolutions are then used, where each convolution is followed by a ReLU activation function. At the output of the final decoder, sigmoid is used for Scenario-1, while softmax activation function is used for Scenario-2.

3.7. Implementation Details

Our networks are implemented in Keras with a single Nvidia GPU Quadro P1000. Network designed for Scenario-1 is trained by binary cross entropy loss and is optimized using the Adam optimizer. Network designed for Scenario-2 is trained by categorical focal loss and is optimized using the Adam optimizer. For the both networks number of epochs is 100, batch size is 16 and learning rate is 0.001.

3.8. Evaluation Metrics

Predictions in artificial intelligence studies; It is evaluated in 4 categories: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These categories are also used in segmentation studies. In such studies, the most important success criterion is the similarity ratio, which is revealed by comparing the ground truth mask and the mask image estimated by the model. The methods used to calculate model prediction success in segmentation studies are based on pixel-based comparison. These methods include metrics such as pixel accuracy and jaccard index (codes aveabile at github.com/OguzhanKATAR23).

3.8.1 Pixel Accuracy

Pixel Accuracy (PA) is the ratio of the overlapping and correctly predicted pixel values to the total number of pixels as a result of comparing the ground truth mask with the predicted mask. PA is calculated by the mathematical equation given in (1).

$$PA = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

In Figure 13., the PA value components for a randomly selected test image and the mask image predicted by the model are given.

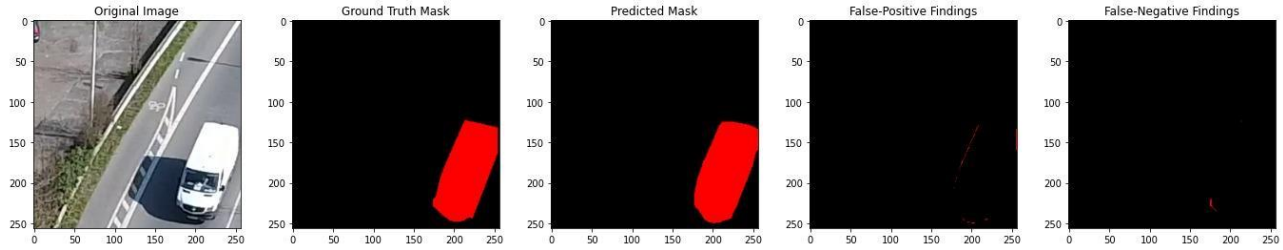


Figure 13. Pixel accuracy value components

PA seems useful due to its easy computation and complexity of performance metrics, but it can produce deceptive results. High resolution images may cause deceptive results as stated.

3.8.2 Jaccard Index

The Jaccard index, also called Intersection over Union (IoU), is an approach used to measure the percentage of overlap between the ground truth mask and the predicted mask. It is calculated by dividing the number of intersecting pixels in the compared masks by the number of union pixels. Its mathematical equation is given in (2).

$$IoU(A,B) = A \cap B / A \cup B \quad (2)$$

In Figure 14., the IoU value components for a randomly selected test image and the mask image predicted by the model are given.

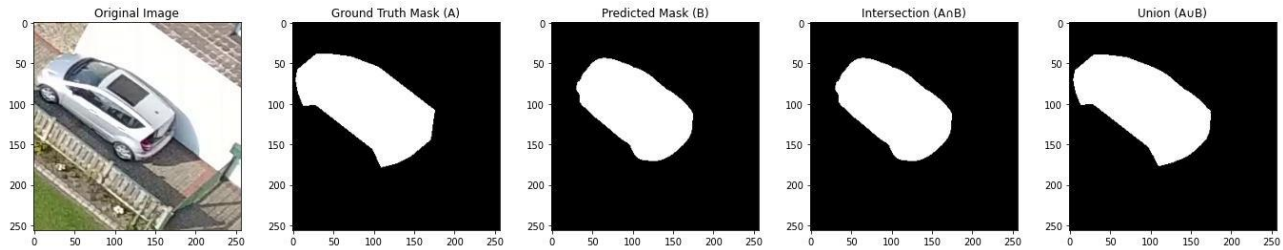
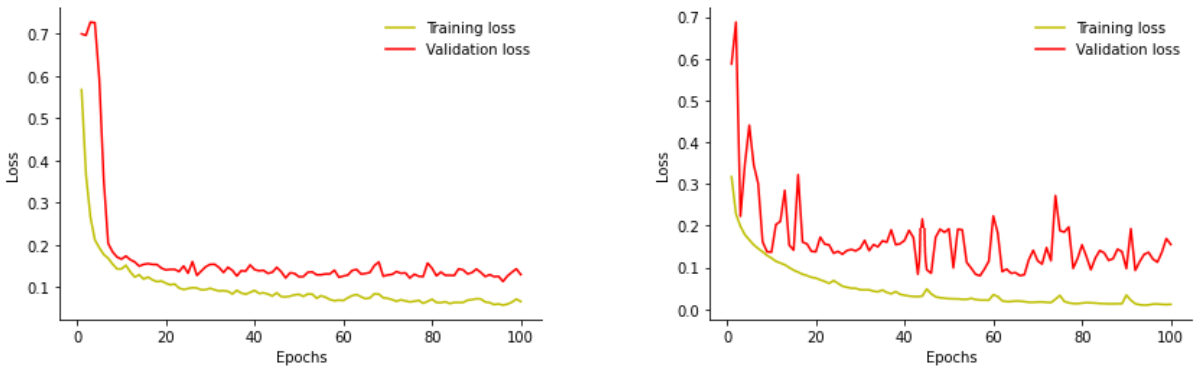


Figure 14. Intersection over Union value components

4. Experimental Results and Discussion

In this study, we developed a fully automatic method to provide vision ability to unmanned aerial vehicles. The model within the scope of Scenario-1, which was designed to mark only the pixels where the vehicle objects are located from the images obtained thanks to the cameras integrated into the unmanned aerial vehicles, was trained for 100 epochs with the help of the customized public dataset. As a result of this training, the mean IoU reached 84.3% and the related loss graph is given in Figure 15(a).



a) Model-1 b) Model-2

Figure 15. Loss graphs of Model-1 and Model-2

Before starting the model training within the scope of Scenario-1, the images that were allocated with 10% share during the test phase were randomly selected and used. The mask images predicted by the model using these images are given in Figure 16.

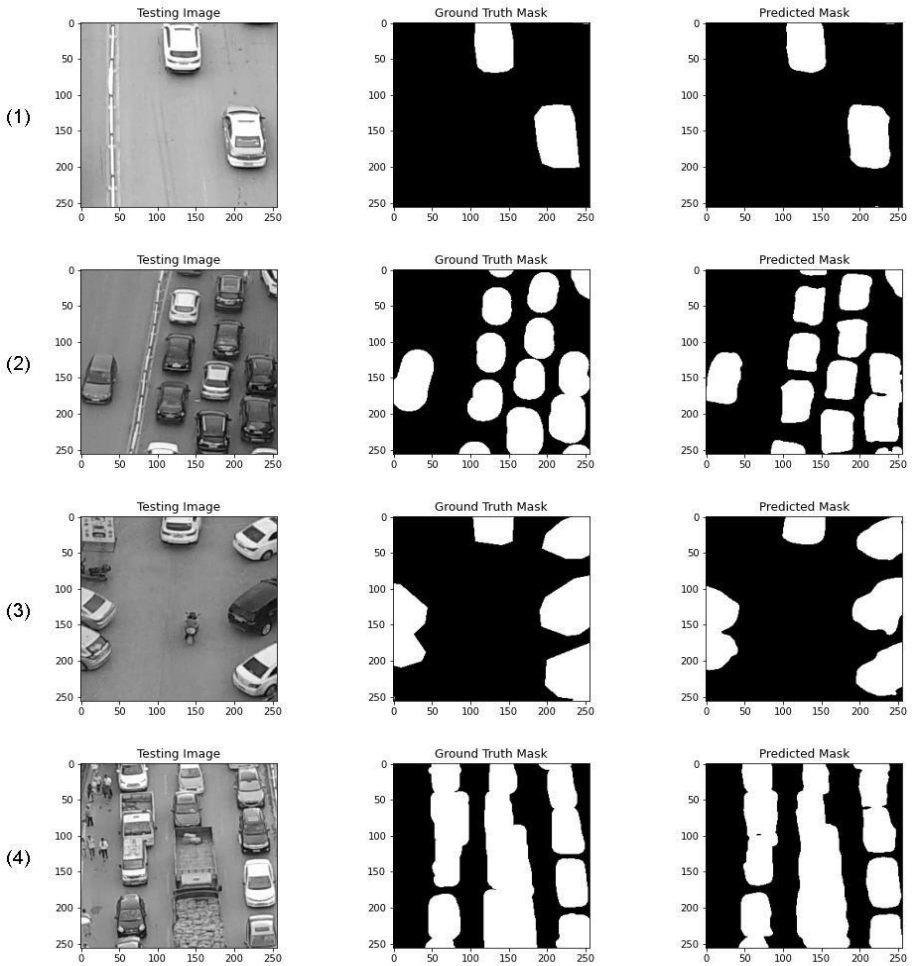


Figure 16. Test phase of Model-1

Table 3. contains the FP, FN, TP and TN values used in the calculation of the PA performance metric for each image given with the image sequence number in Figure 16.

Table 3. PA calculation for test phase of Model-1

Image Sequence Number	False-Positive (px)	False-Negative (px)	True-Positive (px)	True-Negative (px)	Pixel Accuracy (%)
1	90	85	7514	57847	99.73
2	920	1543	22745	40328	96.24
3	217	1117	15062	49140	97.96
4	341	955	30822	33418	98.02

Table 4. contains the intersection and union pixels values used in IoU performance metric for each image given with the image sequence number in Figure 16.

Table 4. IoU calculation for test phase of Model-1

Image Sequence Number	Intersection (px)	Union (px)	IoU (%)
1	7514	7689	97.72
2	22745	25208	90.22
3	15062	16396	91.86
4	30822	32118	95.96

In addition to the ability to detect the vehicle object and mark the location of the relevant vehicle objects on the image, Scenario-2 has been designed to give the unmanned aerial vehicles the ability to distinguish whether the detected vehicle object is moving or stationary. The model is trained for 100 epochs with the help of the customized public dataset. As a result of this training, the mean IoU reached 79.7% and the related loss graph is given in Figure 15(b).

Before starting the model training within the scope of Scenario-2, the images that were allocated with 10% share during the test phase were randomly selected and used. In this scenario, which is created with the multi-class segmentation method, the difficulty is higher than in Scenario-1. For this reason, misclassification may occur in the predicted mask images. The misclassified masks is given in Figure 17.

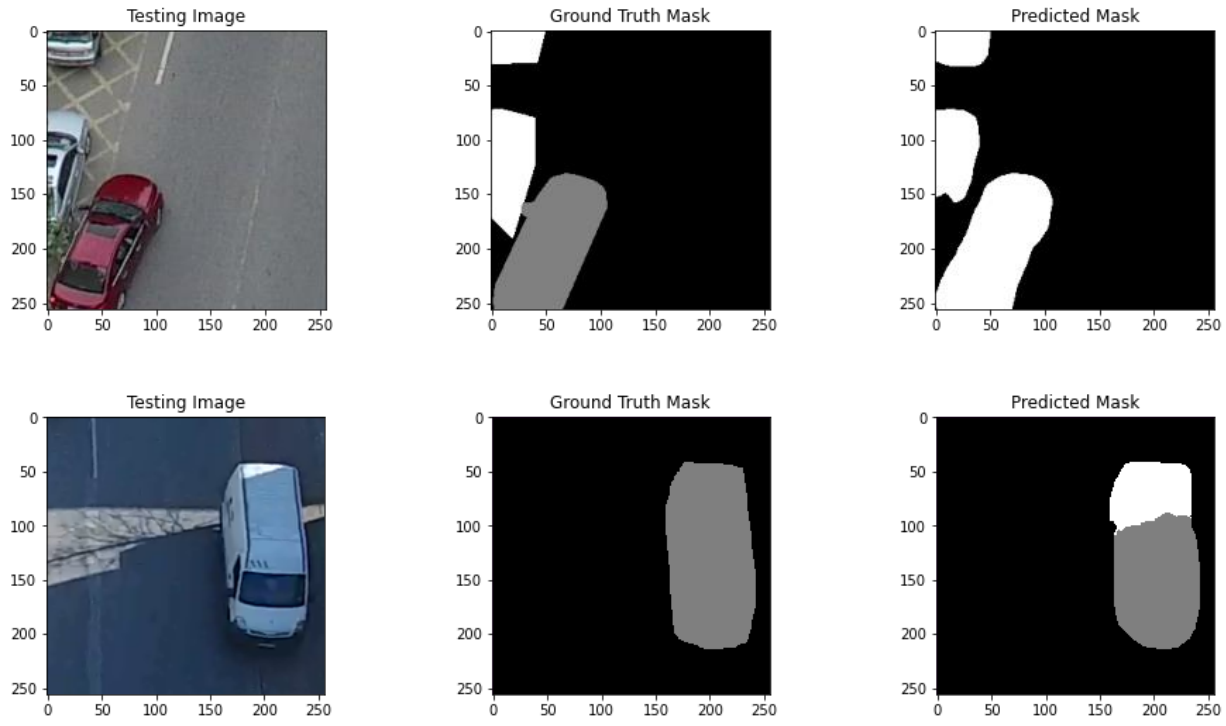


Figure 17. Misclassified mask samples

The images reserved for the test phase are given as input to the model. The masks predicted at the model output and ground truth masks are given in Figure 18.

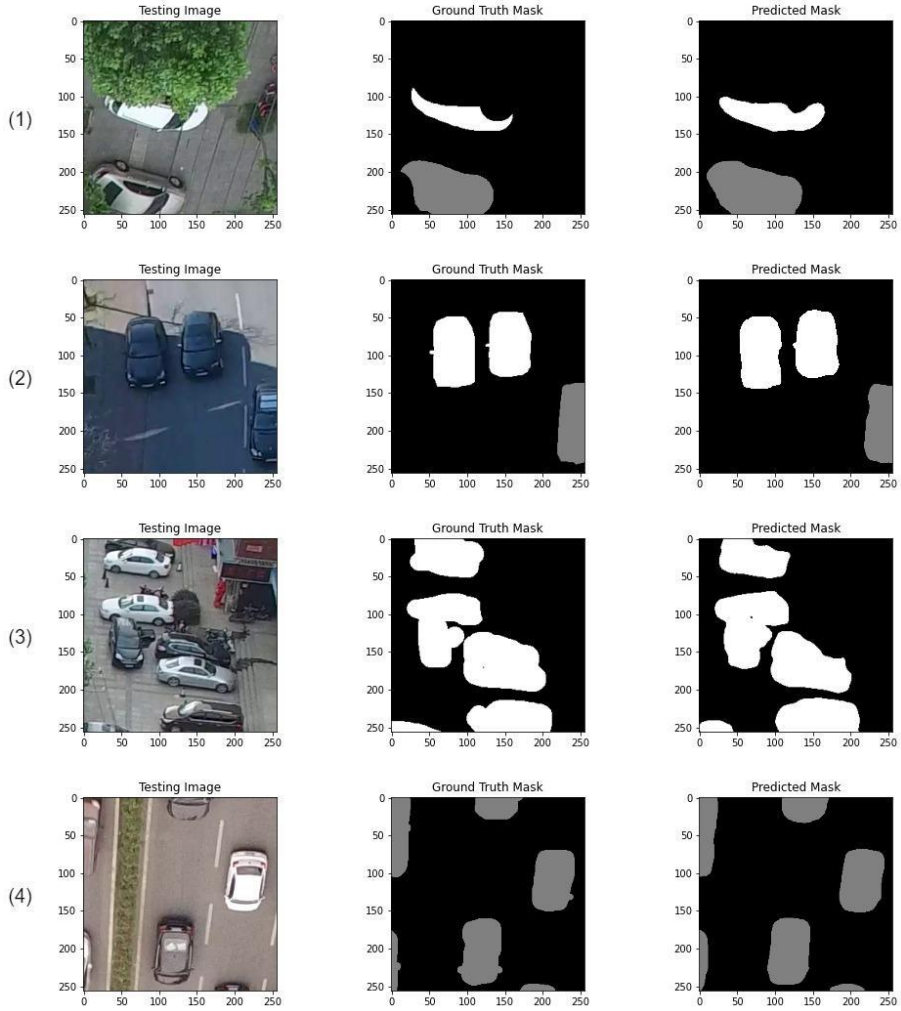


Figure 18. Test phase of Model-2

Table 5. contains the FP, FN, TP and TN values used in the calculation of the PA performance metric for each image given with the image sequence number in Figure 18.

Table 5. PA calculation for test phase of Model-2

Image Sequence Number	False-Positive (px)	False-Negative (px)	True-Positive (px)	True-Negative (px)	Pixel Accuracy (%)
1	1290	68	9216	54962	97.92
2	47	276	12377	52836	99.50
3	461	1321	20560	43194	97.28
4	319	185	13067	51965	99.23

Table 6. contains the intersection and union pixels values used in IoU performance metric for each image given with the image sequence number in Figure 18.

Table 6. IoU calculation for test phase of Model-2

Image Sequence Number	Intersection (px)	Union (px)	IoU (%)
1	9216	10574	87.15
2	12377	12700	97.45
3	20560	22342	92.02
4	13067	13571	96.28

When mask image prediction for high resolution images is requested by our trained models; firstly, the high-resolution image should be divided into sub-images of 256x256px sequentially. After the first step is completed, the images should be given as input to the model with the order that occurs when the images are divided into sub-images. This process should continue until the 256x256px sub-image in the last row. The predicted mask images for each sub-image should be combined using the same sequence number and the pixels added by the padding method should be dropped if there are any. Mask prediction method for high resolution images is given in Figure 19.

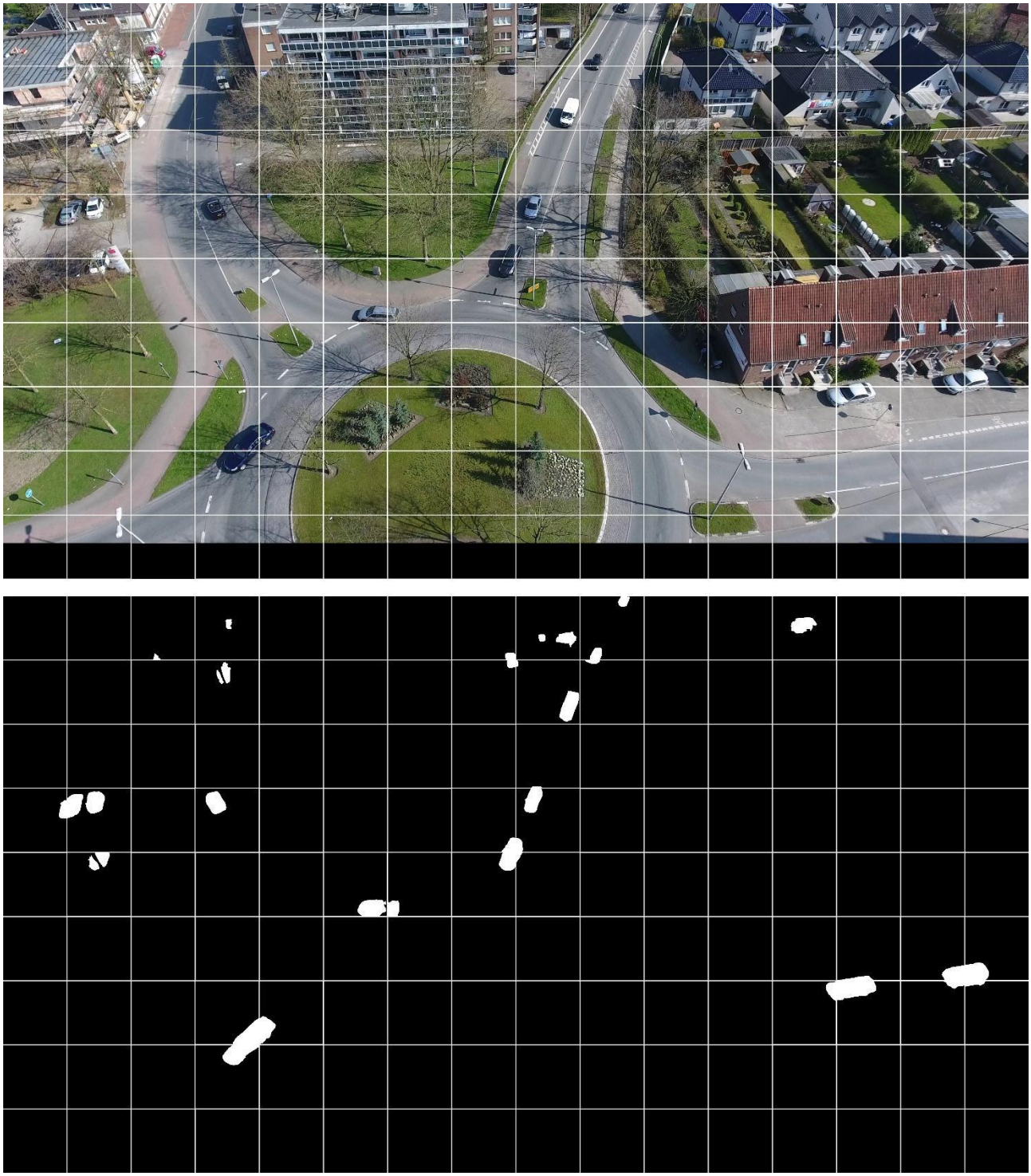


Figure 19. Mask prediction method for high resolution images

5. Conclusion

With the developments in artificial intelligence, the increase in the usage areas of autonomous systems has increased the tendency to add capabilities to unmanned aerial vehicles. In this study, an artificial intelligence-

based system was proposed to improve the vision features of unmanned aerial vehicles. A dataset available to researchers was customized for the scenarios included in this study. In this study, it is aimed to detect vehicle objects autonomously by the computer and to mark the relevant areas on a pixel basis with high IoU values. Indication of detected vehicle objects by pixel-based marking increases the sensitivity of detection. In addition to sensitivity, it is necessary to increase the number of dataset samples and the scope of samples in order to increase the prediction success of our models. For unmanned aerial vehicles that are planned to perform vital tasks such as military operations, the margin of error should be very low. Therefore, it will not be enough to increase the number and scope of dataset samples, but also the deep learning models used should be improved. The proposed system can be easily integrated into the technology used in the field of unmanned aerial vehicles today. Designing unmanned aerial vehicles with such autonomous capabilities in order to prevent possible losses by minimizing the human factor will create the future of the arms industry. Revealing similar studies can help researchers working on autonomous systems and computer vision.

Conflict of Interest

No conflict of interest was declared by the authors.

References

Nonami, K., Kendoul, F., Suzuki, S., Wang, W., Nakazawa, D., 2010. Autonomous flying robots: unmanned aerial vehicles and micro aerial vehicles. Springer Science & Business Media.

Boukoberine, M. N., Zhou, Z., Benbouzid, M., 2019. A critical review on unmanned aerial vehicles power supply and energy management: Solutions, strategies, and prospects. Applied Energy, 255, 113823.

Howard, J., Murashov, V., Branche, C. M., 2018. Unmanned aerial vehicles in construction and worker safety. American journal of industrial medicine, 61(1), 3-10.

Shareef, M. A., Kumar, V., Dwivedi, Y. K., Kumar, U., Akram, M. S., Raman, R., 2021. A new health care system enabled by machine intelligence: Elderly people's trust or losing self control. Technological Forecasting and Social Change, 162, 120334.

Mohamed, N., Al-Jaroodi, J., Jawhar, I., Idries, A., Mohammed, F., 2020. Unmanned aerial vehicles applications in future smart cities. Technological Forecasting and Social Change, 153, 119293.

Kuru, K., 2021. Planning the future of smart cities with swarms of fully autonomous unmanned aerial vehicles using a novel framework. IEEE Access, 9, 6571-6595.

Haulman, D. L., 2003. US unmanned aerial vehicles in combat, 1991-2003. AIR FORCE HISTORICAL RESEARCH AGENCY MAXWELL AFB AL.

Xu, Y., Yu, G., Wu, X., Wang, Y., Ma, Y., 2016. An enhanced Viola-Jones vehicle detection method from unmanned aerial vehicles imagery. IEEE Transactions on Intelligent Transportation Systems, 18(7), 1845-1856.

Dargan, S., Kumar, M., Ayyagari, M. R., Kumar, G., 2020. A survey of deep learning and its applications: a new paradigm to machine learning. Archives of Computational Methods in Engineering, 27(4), 1071-1092.

Blumberg, S. B., Tanno, R., Kokkinos, I., Alexander, D. C., 2018. Deeper image quality transfer: Training low-memory neural networks for 3d images. International Conference on Medical Image Computing and Computer-Assisted Intervention, 118-125.

Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Tian, Q., 2018. The unmanned aerial vehicle benchmark: Object detection and tracking. European conference on computer vision, 370-386.

Mueller, M., Smith, N., Ghanem, B., 2016. A benchmark and simulator for uav tracking. European conference on computer vision, 445-461.

Zhao, T., Nevatia, R., 2003. Car detection in low resolution aerial images. Image and vision computing, 21(8), 693-703.

Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., Zuair, M., 2017. Deep learning approach for car detection in UAV imagery. Remote Sensing, 9(4), 312.

Xu, Y., Yu, G., Wang, Y., Wu, X., Ma, Y., 2017. Car detection from low-altitude UAV imagery with the faster R-CNN. Journal of Advanced Transportation.

Hinz, S., Stilla, U., 2006. Car detection in aerial thermal images by local and global evidence accumulation. Pattern Recognition Letters, 27(4), 308-315.

Lyu, Y., Vosselman, G., Xia, G. S., Yilmaz, A., Yang, M. Y., 2020. UAVid: A semantic segmentation dataset for UAV imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 165, 108-119.