Software for UAV Images Processing for Object Identification

Kateryna Merkulova, Yelyzaveta Zhabska and Ivan Ivanenko

Taras Shevchenko National University of Kyiv, Volodymyrska str. 64/13, Kyiv, 01601, Ukraine

Abstract

This paper describes the research and comparative analysis of methods for object identification in UAV images with the aim of determining the most relevant method in accordance with the described quality criteria in the context of detecting different types of vehicles for its further use during software implementation.

Three identification methods were chosen for the study, namely methods based on ResNet, MobileNet and EfficientDet models. During the research, three quality criteria for evaluating identification methods were developed and described.

As a result, none of the methods showed the best results for all three quality criteria, therefore priorities were set for each quality criterion. Having evaluated the results of the quality criteria for each of the researched identification methods, while taking into account the priorities of the quality criteria, it was concluded that the method based on the MobileNet model is the most optimal among the researched methods in the context of vehicle identification on UAV images.

Keywords [1]

Object identification, UAV, vehicles, artificial neural network, ResNet, MobileNet, EfficientDet, comparative analysis

1. Introduction

With the growing use of unmanned aerial vehicles (UAVs) in fields ranging from military surveillance to geodesy and environmental monitoring, image processing research is becoming a necessary component for the effective use of these technologies [1, 2]. One of the most significant areas of research is the identification of vehicles through the analysis of images obtained from UAVs. This article explores the methods and technologies used to process UAV imagery to accurately identify different types of vehicles.

Knowledge and understanding of modern image processing methods for the purpose of vehicle identification is becoming increasingly important for maintaining safety, efficiency and sustainable development of society. The results of these studies depend not only on the development of the latest technologies, but also on providing our world with greater safety and efficiency in various areas of life.

Solving the problem of automating the process of detecting suspicious objects on UAV images during martial law in Ukraine is an important aspect of security and control in the state. That is why the developed software is designed to identify different types of vehicles on UAV images. The discussed topic is quite relevant at the moment, as drones have become an important tool in the process of waging war. And that is why operative identification of suspicious vehicles on UAV images is a very urgent task in today's realities for our country.

2. Related works and research objective

Nowadays, there are many software applications for UAV images processing with the purpose of object identification. Applications of this type are used in various fields of human activity, starting with agriculture and ending with the military industry. Today, the most popular software tools for UAV images processing with the purpose of object identification usually include the following programs:

Dynamical System Modeling and Stability Investigation (DSMSI-2023), December 19-21, 2023, Kyiv, Ukraine
EMAIL: kate.don11@gmail.com (K. Merkulova); y.zhabska@gmail.com (Y. Zhabska); super-ivan-ivanenko@knu.ua (I. Ivanenko)
ORCID: 0000-0001-6347-5191 (K. Merkulova); 0000-0002-9917-3723 (Y. Zhabska); 0009-0008-0658-9801 (I. Ivanenko)
© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3687/Paper_3.pdf

1. Pix4Dmapper – the main functionality of the program includes automatic processing of images taken from different angles using photogrammetry and computer vision algorithms. The program supports images in JPEG, TIFF, PNG and RAW formats, which allows it to work with most cameras and UAVs on the market [3]. The program is able to determine various parameters of the territory, such as area, volume, height, slope, length, etc. Using the deep learning technology, the program can automatically recognize and classify objects in images, such as buildings, roads, trees and others.

2. DroneDeploy is a cloud-based software for processing aerial images taken by UAVs to create detailed maps and 3D models. It is a full-featured platform for flight mission planning, data collection and image processing, which allows to quickly and easily create map data and 3D models of various objects [4]. DroneDeploy's core functionality includes flight mission planning using an interactive map and automatic UAV control.

3. AgroScout is a software for processing images from a UAV, designed for diagnosing plant diseases and assessing the condition of crops in real time [5]. This technology is used in agriculture to monitor and diagnose plant diseases in the early stages, which allows to avoid the spread of diseases and preserve the harvest. AgroScout uses computer vision technology to analyze images obtained from drones.

Having analyzed the existing software solutions, it can be concluded that the analogs in question are complex software products that are used in their subject area, namely:

1. Pix4Dmapper – the functionality is focused on the analysis of the earth surface.
2. DroneDeploy – functionality is based on the creation of detailed maps and 3D models.
3. AgroScout – the functionality consists in diagnosing plant diseases and assessing the condition of crops.

It is obvious that developing a software product with similar functionality is not the best solution, since there will be no question of novelty. That is why the developed software will be intended for a slightly different subject area, namely for the identification of different types of vehicles in images taken from UAVs. The discussed topic is quite relevant at the moment during martial law in Ukraine, as drones have become an important tool for conducting various war operations. Therefore, it is important to implement the software that can perform identification of suspicious vehicles on UAV images as quickly and accurate as possible.

3. Methods of research

Nowadays, there is no exact analytical solution to the problem of object identification in images, which complicates the development of a universal algorithm. Nevertheless, in order to endow computer systems with the possibility of so called "vision", a large number of methods and algorithms have been created and proposed [6, 7]. The purpose of this section is to consider the most popular of them, after which, on the basis of tests of their software implementations, select one or another method for the implementation of the final software product. Based on the previous experience, as well as on information from various articles, the following most popular and used methods for identifying objects in the image were selected:

4. MobileNet is a set of architectures of deep neural networks [8]. They are optimized to run on devices with limited computing resources.
5. ResNet is an abbreviation for "Residual Network" (Network with reverse connections). It was developed in order to ensure successful training of deep neural networks, avoid the problem of gradient damping and facilitate learning [9].
6. EfficientDet is a family of object detector architectures that combine efficiency and high accuracy [10]. They were developed in order to provide efficient image processing and object recognition with minimal computational costs.

So, at the moment, identification methods have been determined, which were selected for further implementation. Each of the selected methods is a certain type of artificial neural network, therefore, for their implementation, it is necessary to conduct their training on a certain set of data. Today, the issue of national security during martial law is extremely important, which is why it is extremely important to timely identify suspicious vehicles in the front-line territories and beyond. Thus, these types of objects for identification in UAV images are quite relevant now. Therefore, it was decided to use a variety of vehicles as objects for identification in the images taken from the UAV. The next task is the formation of data for training artificial neural networks. For now, there are more than a dozen sets of images with annotations to them, which would be suitable for the task of this research, have already been created. As a result of detailed analysis, a dataset called VisDrone, which contains 8408 annotated images, was selected for training [11].

VisDrone is a dataset designed for the tasks of object detection and tracking in images captured by unmanned aerial vehicles (drones). This dataset contains annotations for various object classes such as pedestrians, cars, trucks, buses, motorcycles, and others. Within the VisDrone dataset, each object is annotated with a bounding box that shows its position in the image. These annotations are used to train and evaluate object detection models. Figure 1 shows an image from the VisDrone dataset along with the bounding boxes contained in the image annotation.

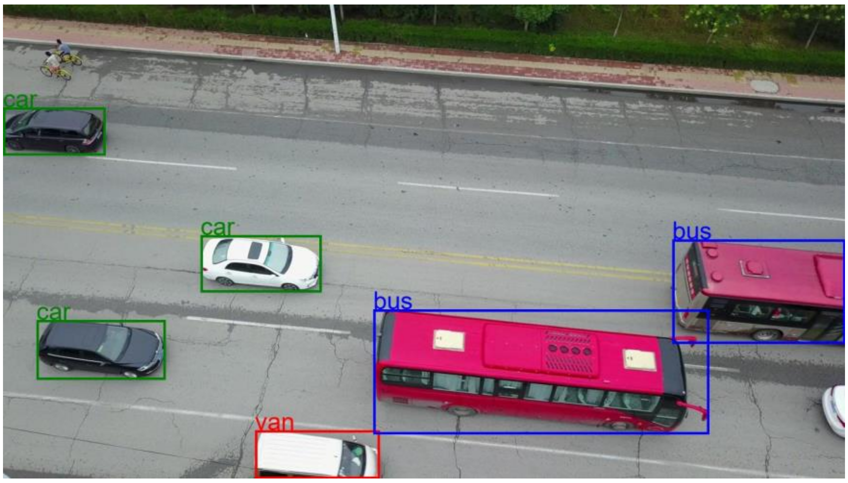

Figure 1: Image from VisDrone and its annotation

TensorFlow [12] is used as a tool for implementing selected models. TensorFlow is an open source machine learning and deep learning software developed by Google. It provides a framework for building and training a variety of artificial intelligence models, such as neural networks.

In the process of training, monitoring was also carried out, which reflects the effectiveness of model training. Monitoring displays such loss functions as Classification loss and Localization loss.

Classification loss is a loss function used during the training of a neural network for classification. It measures the distance between predicted and actual class labels and helps train the network during the training process [13, 14]. Localization loss is a loss function used for training of a neural network to localize objects in images. It measures the distance between the predicted and actual object coordinates and helps train the network during the training process.

Figures 2-4 present the graphs of the loss functions described above for each of the trained models.

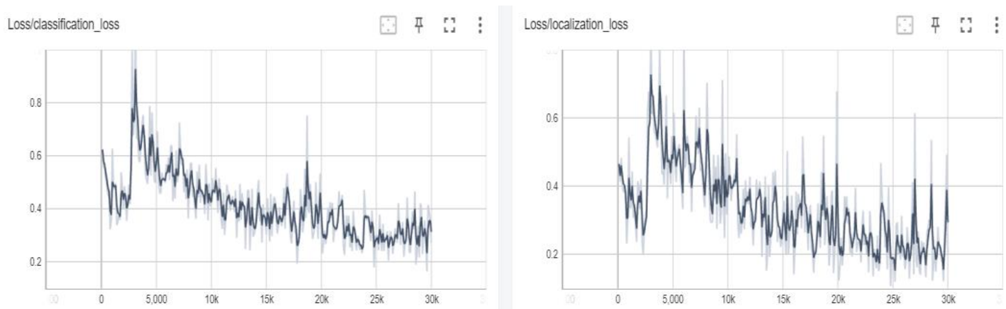

Figure 2: ResNet model training process

To compare the training results of the selected models, Table 1 is presented, which contains the values of the loss functions for each model at the end of training.

Before analyzing the obtained results, it is necessary to clarify that the lower the value of the loss function, the better the trained model performs the task. That is, the smaller the loss function, the more the model predictions correspond to the expected result.

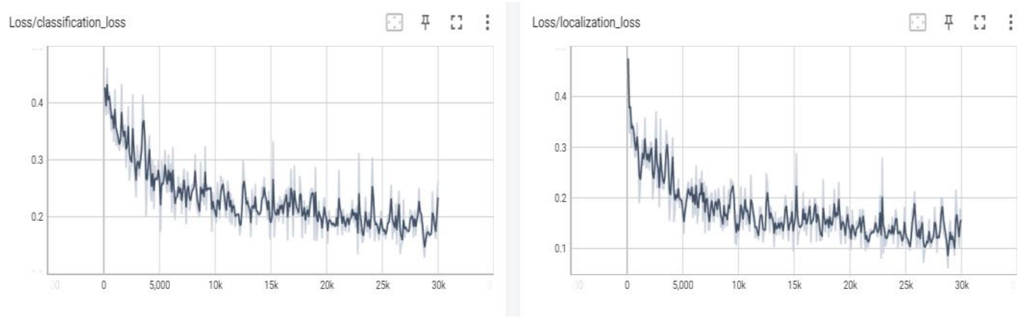

Figure 3: MobileNet model training process

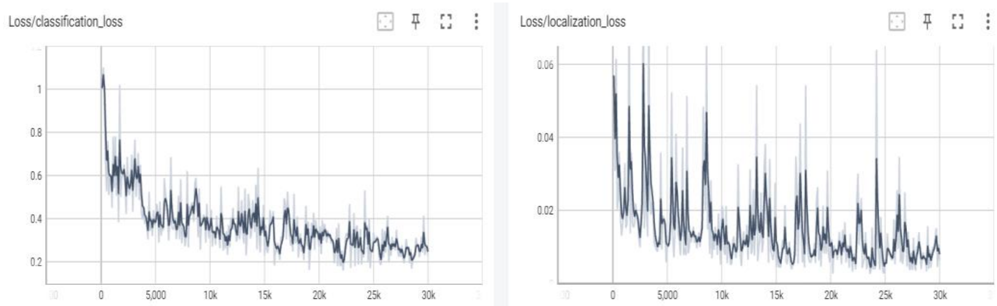

Figure 4: EfficientDet model training process

Table 1
Results of model training

	classification_loss	localization_loss
ResNet	0.2723	0.1988
MobileNet	0.2344	0.1638
EfficientDet	0.2259	0.0064

Thus, analyzing the results of model training, it can be concluded that the EfficientDet network performed best in the training process, as it has the lowest values of the loss functions. It should be noted that the value of classification_loss in all three models is almost at the same level, while localization_loss has a much smaller value in the EfficientDet model. So, the selected models were implemented and the results of their training were compared. Now it is necessary to check the performance of the implemented methods in practice. It can be done with the help of the proposed quality criteria, which will be discussed later.

At this stage, the question arises, what should be paid attention to when choosing quality criteria? In the research process, they will be used for vehicle identification methods in UAV images, therefore, based on the previous experience and information from various papers, the following criteria will be considered the most appropriate for this type of method.

3.1.1. The ratio of the number of correctly identified objects to the number of all objects

The ratio of the number of correctly identified objects to the number of all objects is a criterion that evaluates such a quantitative characteristic of the model as the ability to identify given objects. In order not to specify the name of this metric every time, let's highlight for it, for example, the symbol *R*. The expression 1 demonstrates the formula for calculating the quality criterion *R* for the identification method, using a sample of images of size *N*:

$$R = \frac{\sum_{i=1}^{N} m_i}{\sum_{i=1}^{N} k_i}, \tag{1}$$

where *R* is the quality criterion of the identification method, namely the ratio of the number of correctly identified objects to the number of all objects, *N* is the number of all images, m_i is the number of objects that the method was able to correctly identify on the *i*-th image, k_i is the actual number of objects in the *i*-th image.

In which case is it considered that the method correctly identified the object in the image? The quality assessment metric called Intersection over Union, which will be discussed in more detail later in the paper, will help answer the first question. The value of the IoU quality assessment metric can vary from 0 to 1. It is generally accepted that $IoU > 0.5$ is a good prediction of the object detector, otherwise the prediction is not good. Thus, when counting correctly identified objects, it is needed to calculate the value of the IoU quality assessment metric for each of them. If $IoU > 0.5$, the object is included in the calculation, otherwise it is ignored. Expression 2 demonstrates the formula for calculating the number of correctly identified objects in the *i*-th image:

$$m_i = \sum_{j=1}^{k_i} \begin{cases} 1, & IoU_j > 0.5, \\ 0, & IoU_j \leq 0.5, \end{cases} \tag{2}$$

where m_i is the number of correctly localized objects in the *i*-th image, k_i is the actual number of objects in the *i*-th image, IoU_j is the value of the IoU quality assessment metric calculated for the *j*-th object in the *i*-th image.

Thus, when calculating the quality criterion *R*, formula 2 will be substituted into formula 1.

The last question, that remains open within the scope of this point, is what should be the number of all images in order to calculate the value of the quality criterion with a given error ε. Expression 3 demonstrates the formula for finding the number *N* for a given error ε:

$$\varepsilon = |f(n + step) - f(n)|, \tag{3}$$

where ε is the specified error for calculating *f*, *f(N)* is the value of the metric *f* for some object identifier, calculated using a sample of images of size *N*, *n* is the current value of the number of images for calculating the value of *f*, *step* is a fixed step that increases the value *n* for each subsequent iteration.

Figure 5 contains a block diagram of the algorithm for finding the value of *N* for a given error ε.

There are probably no separate rules or advice for choosing the permissible error value ε, again everything depends on the intuition of the developer himself. Of course, it's not worth to take too large values for ε, such as 10^{-2}. Based on the experience of previous research of various computational methods, it is recommended to take the value of ε as 10^{-5}. Usually this accuracy is completely sufficient for most numerical methods.

3.1.2. Intersection over Union

Intersection over Union is a criterion that evaluates the accuracy with which the model localizes objects in the image. Expression number 4 demonstrates the formula for calculating the IoU metric:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}, \tag{4}$$

where *IoU* is the Intersection over Union quality assessment metric, *Area of Overlap* is the area of intersection of the predicted bounding frames with the actual bounding frames, *Area of Union* is the area of the union of the predicted bounding frames with the actual bounding frames.

From formula 4 it can be seen that the possible values for the IoU metric are in the numerical range from 0 to 1, taking into account the extreme points. An $IoU > 0.5$ is generally considered to be a good predictor of an object detector. The next formula will allow to calculate the average value of the IoU quality criterion for a sample of N images:

$$IoU_c = \frac{\sum_{i=1}^{N} IoU_i}{N}, \tag{5}$$

where IoU_c is the average value of the Intersection over Union quality assessment metric, IoU_i is the value of the IoU quality assessment metric calculated for the *i*-th image and *N* is the number of all images.

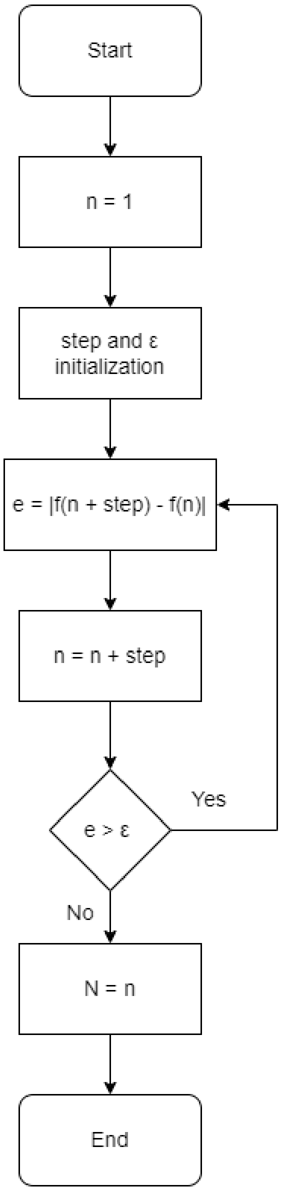

Figure 5: Block diagram of the algorithm for finding *N*-value for a given *ε*-value

3.1.3. Average object localization time

Average object localization time is a criterion that is designed to demonstrate the speed of the researched method. The following expression demonstrates the calculating of the average value of the T criterion.

$$T = \frac{\sum_{i=1}^{N} T_i}{N}, \tag{6}$$

where T is the average object identification time in the image, T_i is the average object identification time in the i-th image and N is the number of all images.

Thus, all three criteria cover the most significant characteristics of the object identification method, namely identification ability, localization accuracy, and recognition speed. It should also be noted that each of the mentioned metrics will be calculated not for a single image, but for some sample of size N, that is an averaged value is used.

4. Research results

The results of the calculation of all three quality criteria for each object identification method are summarized in Table 2.

Table 2
Quality criterions for object identification methods

	R	IoU_c	T
ResNet	0,82798	0,8767	0.00282
MobileNet	0,92156	0,8656	0.00245
EfficientDet	0,62243	0,8349	0.00087

Now, based on the obtained results, comparative analysis of the selected identification methods can be conducted. For this, their qualitative and quantitative comparison was carried out in the context of the previously described quality criteria.

Let's start with such a quality criterion as Intersection over Union. The physical meaning of this metric is a numerical representation of the accuracy with which the identification method predicts the location of the object in the image. Figure 6 demonstrates three graphs for each of the identification methods, which show the dependence of the average value of the IoU quality criterion on the number of images required for its calculation.

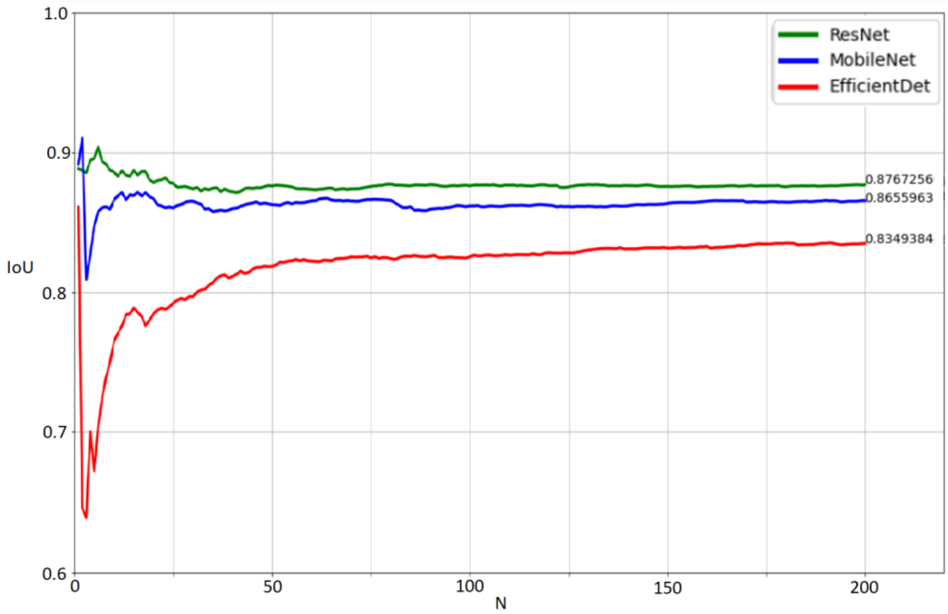

Figure 6: Graphs of the IoU(N) function for three identification methods

Qualitative comparison of detection methods according to the IoU metric: starting from about 50 images, the values of the IoU metric for all three identification methods actually do not change, that is, they acquire constant values. For the ResNet model the final value is $IoU_1 \approx 0.87673$, for the MobileNet model $IoU_2 \approx 0.86559$ and for EfficientDet $IoU_3 \approx 0.83493$. In general, the IoU metric can take values between 0 and 1, including extreme values. A detection method for which $IoU > 0.5$ is considered good. Therefore, based on this fact, it can be stated that all three methods for which the calculations were performed are quite good detectors. But, of course, one of them showed a slightly better result than the others, it is a model based on ResNet.

Quantitative comparison of identification methods by the IoU metric can be calculated as follows:

$$P_{IoU} = \frac{0,87673 - 0,86559}{0,86559} \cdot 100\% \approx 1,3\%, \tag{7}$$

$$P_{IoU} = \frac{0{,}87673 - 0{,}83493}{0{,}83493} \cdot 100\% \approx 5\%. \tag{8}$$

As a result of the quantitative comparison of the two detection methods based on the IoU metric, it can be concluded that the ResNet model predicts the placement of given vehicles on the image by 1.3% more accurately than the MobileNet model, and by 5% more accurately than the EfficientDet model.

Next, for the comparative analysis of the methods, we will take the metric R, that is, the ratio of the number of correctly located vehicles to the number of all vehicles. The physical meaning of this metric is a numerical representation of the ability of the method to correctly identify the given vehicles. Figure 7 shows contains graphs for each of the identification methods, which show the dependence of the numerical value of the metric R on the number of images required for its calculation.

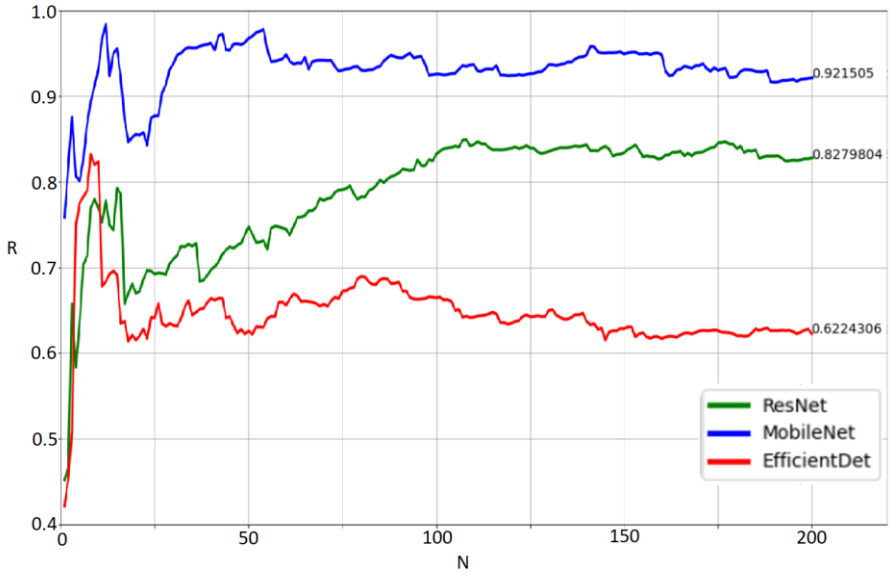

Figure 7: Graphs of the R(N) function for three identification methods

Qualitative comparison of detection methods according to the R metric: starting from about 125 images, the values of the R metric for the three identification methods do not actually change, that is, they acquire constant values. In Figure 7 it can be seen that in the case of the method based on MobileNet the final value $R_1 \approx 0.92156$, for the method based on ResNet $R_2 \approx 0.82798$, for EfficientDet $R_3 \approx 0.62243$. Summarizing the qualitative comparison of the methods, it can be definitely concluded that the identification method based on the MobileNet model is the best among the studied methods according to the quality criterion R. At the next stage, it is necessary to determine how much the MobileNet model is better than the other studied methods in the context of the identification of given vehicles by the R metric.

Quantitative comparison of identification methods by the R metric can be demonstrated with the following:

$$P_R = \frac{0{,}92156 - 0{,}82798}{0{,}82798} \cdot 100\% \approx 11{,}3\%, \tag{9}$$

$$P_R = \frac{0{,}92156 - 0{,}62243}{0{,}62243} \cdot 100\% \approx 48\%. \tag{10}$$

As a result of the quantitative comparison of the detection methods by the R metric, it can be concluded that the method based on the MobileNet model has 11.3% more ability to identify vehicles in the image than the method based on ResNet, and 48% more than the method based on EfficientDet.

The last quality criterion for comparing the identification methods is the average identification time of one vehicle in the image. Figure 8 shows three graphs for each of the identification methods, which show the dependence of the numerical value of the metric *T* on the number of images *N* required for its calculation.

Qualitative comparison of detection methods by the metric T: starting from about 125 images, the values of the metric T for all three identification methods practically do not change, that is, they acquire constant values. For the method based on the ResNet model, the final value is T1 ≈ 0.00282 s, for the method based on MobileNet T2 ≈ 0.00245 s, for EfficientDet T3 ≈ 0.00087 s. In the case of the first two metrics, which were considered earlier, the method in which the corresponding metric has a larger numerical value was considered better. Instead, for this metric, the method in which the numerical value is smaller will be better. This is intuitive because the shorter the average identification time of one vehicle in the image, the more efficient the identification method is.

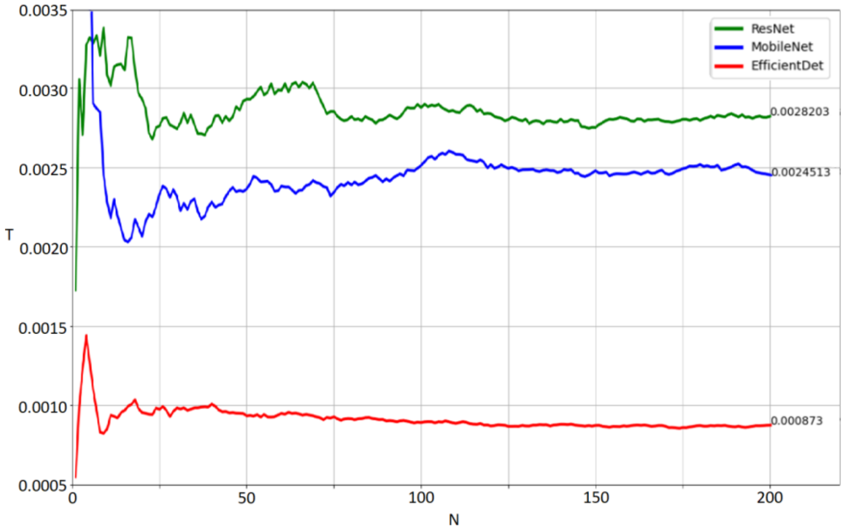

Figure 8: Graphs of the function T(N) for three identification methods

After a qualitative comparison of the methods, it can be unequivocally stated that the method based on EfficientDet is the best among the investigated methods according to the quality criterion T. Now it is necessary to determine how much this method is better than the others in numerical equivalent.

For their quantitative comparison, it will be easier to find how many times one of them is more than the other. This can be calculated as follows:

$$\frac{0{,}00282}{0{,}00087} \approx 3{,}24; \quad \frac{0{,}00245}{0{,}00087} \approx 2{,}91. \tag{11}$$

Therefore, as a result of the quantitative comparison of the three identification methods based on the T metric, it can be concluded that the identification method based on EfficientDet on average identifies one vehicle in the image 3 times faster than the other two methods.

5. Conclusion

After conducting a comparative analysis of the studied identification methods, it is currently difficult to determine which of them is the best in the context of vehicle identification in UAV images, since each of them showed the high result in at least one quality criterion.

Thus, based on the obtained results, it can be said that each of the considered identification methods has its own advantages and disadvantages. However, it is necessary to clearly determine which of the given methods is the most optimal according to the given criteria in the context of vehicle identification on UAV images. Since none of the methods showed the best results for all three quality criteria, it is now necessary to prioritize each quality criterion.

Based on the expert experience of the authors, it can be noted that the quality criterion T, which characterizes the speed of the method, is the least significant (third priority) in this context, since each of the methods showed results of the order of 10^{-3} seconds. That is, it will be difficult to notice a significant difference between them in terms of speed in practice. Instead, other quality criteria that characterize other aspects of the studied methods are more significant, as they can be felt in practice. Quality criterion R, which characterizes the ability of the method to identify vehicles, is the most significant (first priority). As it will be clearly seen that one method managed to identify 10 vehicles and the other only 7 in the same image. Accordingly, only the second priority remains for the IoU quality criterion. Having evaluated the results of the quality criteria for each of the researched identification methods, while taking into account the priorities of the quality criteria, it was concluded that the method based on the MobileNet model is the most optimal among the researched methods in the context of vehicle identification on UAV images. Because it showed the best results for the quality criterion R, which has the highest priority, while breaking away from the other two methods by a margin (by 11.3% - ResNet; by 48% - EfficientDet). The next most important is the IoU quality criterion, according to which the method based on the MobileNet model showed the second result, lagging behind the ResNet by 1.3%. In other words, MobileNet and ResNet are actually equal to each other according to this criterion. According to the third least significant quality criterion, the MobileNet-based method showed again the second result, which is three times worse than the first result, which was shown by the EfficientDet-based method. Again, it may seem that this is quite a noticeable difference, but if we take into account the fact that the obtained speed values are of the order of 10^{-3} seconds, then in practice this difference will not be noticeable. Although EfficientDet is the fastest in the context of vehicle identification in UAV images, on the other hand it performed the worst in two other more significant quality criteria. Therefore, based on the obtained results, the method based on the EfficientDet model is the worst in the context of vehicle identification in UAV images according to the given quality criteria, taking into account their priority.

6. References

[1] A. N and U. S V, “Drone Object Detection Using Deep Learning Algorithms,” 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, India, 2021, pp. 1187-1192, doi: 10.1109/ICIRCA51532.2021.9544983.

[2] G. Cheng, Z. Shao, Q. Cheng and H. Luo, “Small Object Detection on Drone-view Images from Multiple UAVs Combination,” 2023 30th International Conference on Geoinformatics, London, United Kingdom, 2023, pp. 1-7, doi: 10.1109/Geoinformatics60313.2023.10247874.

[3] Pix4Dmapper. URL: https://www.pix4d.com/product/pix4dmapper-photogrammetry-software/

[4] DroneDeploy. URL: https://www.dronedeploy.com/

[5] AgroScout. URL: https://agro-scout.com/

[6] V. Martsenyuk, O. Bychkov, K. Merkulova and Y. Zhabska, “Exploring Image Unified Space for Improving Information Technology for Person Identification,” in IEEE Access, vol. 11, pp. 76347-76358, 2023, doi: 10.1109/ACCESS.2023.3297488.

[7] S. Kumar and C. Kumar, “Drone/UAVs Surveillance feeds Video Object Recognition Using Transfer Learning,” 2022 International Conference on Futuristic Technologies (INCOFT), Belgaum, India, 2022, pp. 1-4, doi: 10.1109/INCOFT55651.2022.10094328.

[8] A. G. Howard, et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, CoRR, 2017, abs/1704.04861. URL: http://arxiv.org/abs/1704.04861.

[9] K. He, X. Zhang, S. Ren and J. Sun, “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[10] M. Tan and A. Yu, “EfficientDet: Towards Scalable and Efficient Object Detection”, April 15, 2020. URL: https://ai.googleblog.com/2020/04/efficientdet-towards-scalable-and.html

[11] P. Zhu, et al., “Detection and Tracking Meet Drones Challenge” in IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 44, no. 11, pp. 7380-7399, 2022. doi: 10.1109/TPAMI.2021.3119563.

[12] TensorFlow. URL: https://www.tensorflow.org/

[13] Ivanytska, L. Zubyk, Y. Zubyk and D. Ivanov, “The advertising prediction model based on machine learning technologies”, CEUR Workshop Proceedings of the 8th International Conference "Information Technology and Interactions" (IT&I-2021), Kyiv, Ukraine, December 1-3, 2021. CEUR-WS, vol. 3179, pp. 35-44. https://ceur-ws.org/Vol-3179/Paper_4.pdf

[14] Kutyrev, A., Kiktev, N., Kalivoshko, O., Rakhmedov, R. Recognition and Classification Apple Fruits Based on a Convolutional Neural Network Model. CEUR Workshop Proceedings of the 9th International Conference "Information Technology and Implementation" (IT&I-2022), Kyiv, Ukraine, November 30 - December 2, 2022. CEUR-WS, vol. 3347, pp. 90-101. https://ceur-ws.org/Vol-3347/Paper_8.pdf