Exploring of the MobileNet V1 and MobileNet V2 models on NVIDIA Jetson Nano microcomputer

G Edel and V Kapustin
Department of Television and Control, Tomsk State University of Control Systems and Radioelectronics, 40, prospect Lenina, 634050, Tomsk, Russia
E-mail: german.edel99@gmail.com

Abstract. This article is devoted to the study of the accuracy of MobileNet V1 and MobileNet V2 models when recognizing pedestrians at different times of the year, with changing distance and position. According to the test results, the pros and cons of the models were identified, as well as their capabilities and dependencies. Models were compared directly to each other. All studies were performed on an NVIDIA Jetson Nano microcomputer.

1. Introduction
Today it is difficult to imagine a world where it is safe on the roads. Tragedies happen every day due to human carelessness, negligence and banal stupidity. Therefore, the creation of self-driving cars is one of the most important and promising areas. Corporations are investing billions in the development of this field. The prospect of protecting citizens, giving them the opportunity to rest and work on the road, as well as getting rid of traffic jams and being able to safely deliver goods in case of emergencies is very tempting. Therefore, it is so important to conduct research in this field.

In recent decades, computer vision researchers have made tremendous efforts to enable computers to understand the world from video data. The most important task of computer vision is the automatic detection of objects. The most demanded, but at the same time difficult task is the recognition of people. Accurate detection and localization of pedestrians could significantly help road safety. It could also help take the next step in such fields of expertise as video surveillance, robotics automation and more. This topic is of interest to both researchers and industry. In video surveillance, pedestrian detection helps provide basic information for people counting, event recognition and crowd monitoring. For intelligent driver assistance systems, pedestrian detection is an important part of the system for recognizing objects in the surrounding space [1-3].

The main purpose of this work is to compare the models MobileNet V1 and MobileNet V2 running on the NVIDIA Jetson Nano microcomputer, in order to define the accuracy of pedestrian recognition under various conditions, and also to make conclusions based on the results obtained.

2. NVIDIA Jetson Nano Developer Kit
The algorithm for pedestrian recognition was implemented and run on the NVIDIA Jetson Nano microcomputer [4]. This development kit provides the computing power and functionality to leverage modern artificial intelligence (AI) in a low-cost, low-power, easy-to-use platform.

This kit is supported by the comprehensive NVIDIA® JetPack™ SDK, which includes L4T (Linux for Tegra, Linux distribution for Ubuntu Desktop with NVIDIA drivers), AI and computer vision APIs and libraries, development tools, documentation, and code samples. The complete kit consists of a Jetson
Nano module equipped with a passive heatsink that fits into the SO-DIMM slot on the carrier board (sometimes also on the base board). It contains a large number of connectors for connecting peripheral devices [4].

3. MobileNet V1 and MobileNet V2 models

Modern architectures are quite large in depth, and therefore require a significant amount of memory and output time. In addition, most CNNs need high quality graphics devices to train and render them. As the network gets deeper, it requires more memory and computation, limiting it to run on even fewer devices. A group of researchers from Google suggested MobileNet V1 and MobileNet V2. These two models are quite lightweight, and their accuracy is comparable to the larger models. These models have a lot in common, but MobileNet V2 uses newer methods adopted by ResNet [5].

The implemented algorithm was tested on pre-shot videos. The video captured the various traffic situations that the algorithm might have encountered when installed on a vehicle. The entire set of video data was filmed on the camera of a Xiaomi Redmi 8Pro mobile phone with 720p and 30FPS settings. The duration of videos is on average about 10 s, and the screen resolution is 720 × 1280 px. Then the implemented algorithm was applied to all videos. Subsequently, we have a number of videos with pedestrians detected or not. Above each detected pedestrian there is a percentage of recognition reliability. The processing power of the NVIDIA Jetson Nano microcomputer is sufficient to provide object recognition speed of about 36 ms for MobileNet V1 and 43 ms for MobileNet V2. This means that these models are capable of working in real time [5, 6].

4. Accuracy of pedestrian detection based on figure resolution and season

Investigated how distance and time of year affect recognition accuracy. The resolution of the figure steadily decreases with increasing distance. The study was carried out by filming the immediate distance from pedestrians to the camera. Noteworthy data are resolution, accuracy, and distance. In a stationary position at a distance of 5 meters, as shown in Figure 1, the human figure has dimensions of 104 × 353 px in frontal and 73 × 345 px in profile shooting in summer. In winter photos, the resolution of the figures is 98 × 325 px for frontal shooting and 94 × 323 px for profile shooting.

As can be seen from the received data MobileNet V1 model recognizes a pedestrian well at a distance of 5 m. Recognition accuracy in summer is close to 100% for both frontal and profile shooting. On winter shots, the accuracy is slightly lower than on summer shots and is 96.9% for frontal shots and 88.4% for profile shots.

Similar studies on the same videos were carried out for the MobileNet V2 model Figure 2.
The MobileNet V2 model performed similarly, also doing well in pedestrian recognition in the summer. The recognition accuracy in winter is slightly better than that of MobileNet V1 and amounted to 95.1% and 90%. It can be concluded that recognition of pedestrians in winter is more difficult for both models. Most likely this is due to the image of a pedestrian, a down jacket and a hat reduce the recognition accuracy.

After this, the distance gradually increased. Thus, Figure 3 shows the detection of a pedestrian by the MobileNet V1 model at a distance of 25 m and 30 m in different seasons. The figure resolution in summer at 25 m is $22 \times 66$ px, and at 30 m it is $17 \times 55$ px. In winter, the resolution of the figures is $20 \times 61$ px at 25 m and $15 \times 52$ px.

The figure shows how the accuracy decreases significantly with increasing distance. On summer shots, already at a distance of 25 m, it is equal to 72.2%, and at 30 m MobileNet V1 did not detect a pedestrian. On winter frames, the recognition accuracy at 25 m is 89.5%. This is higher than in the frames in the summer, and the pedestrian, albeit bad, was recognized at a distance of 30 m. The same studies were carried out for the MobileNet V2 model, Figure 4.
Figure 4. Detection of pedestrians with increasing distance in summer up to 25 m (a), 30 m (b) and in winter up to 25 m (c), 30 m (d).

The recognition accuracy for 25 m is 71.9%. At a distance of 30 m, the MobileNet V2 model was unable to detect a pedestrian. Winter results are almost the same, but at 25 m, the recognition accuracy is 66.5%.

The MobileNet V1 model does a better job of recognizing pedestrians in winter, it showed much better accuracy and was able to recognize a person at a distance of 30 m.

5. Accuracy of detecting pedestrians crossing the road depending on the season

According to statistics, a huge number of road accidents with pedestrian collisions occur when crossing the road in the wrong place. As a consequence, the algorithm was tested for the ability to recognize a running person. Figure 6 shows a running pedestrian at a distance of 5 m, 10 m in summer and winter, detected by the MobileNet V1 model.

Figure 5. Recognition of a running pedestrian by the MobileNet V1 model in summer at a distance of 5 m (a) 10 m (b) and in winter (c) 10 m (d).

The test results are shown in the graphs in Figures 6 and 7.
From the time-lapse accuracy graphs, it can be seen that the MobileNet V1 model is much worse at recognizing a running pedestrian in winter both at 5 m and at 10 m. The average recognition accuracy for recognition at a distance of 5 m in summer is 94.7%, and 80.8% in winter. At a distance of 10 m, the average accuracy of summer frames is 63.4%, and winter ones - 57.8%.

Similar studies were carried out for the MobileNet V2 model Figure 8.

The result of such testing was uninterrupted recognition on summer frames with a very high accuracy at a distance of 5 m. The frame-by-frame accuracy of a person in motion at a distance of 5 m is shown in Figure 9.
However, as can be seen from the figure, the recognition accuracy in winter frames is significantly lower than in summer ones. For winter shots at a distance of 5 m, the average accuracy is 88.3% versus 94.9% for summer ones. With an increase in the distance to 10 m, the average accuracy drops rapidly in frames of both seasons. Thus, winter shots have an average accuracy of 66.6%, and summer ones - 60.3%. As you can see, the increase in distance significantly affected both accuracy and stability with the overall dynamics of the graph. The recognition accuracy has noticeably decreased, in several places it even turned out to be below the threshold value equal to 50%. It can be concluded that distance plays its role in recognizing a person crossing the road.

6. Conclusion
At the end of the study of the implemented algorithms, one can come to the following conclusions. The resolution of the figure image significantly affects the recognition accuracy. With an increase in the distance to 30 m, the algorithm was no longer able to recognize a pedestrian in the summer. The MobileNet V2 model showed a similar result. The older model showed slightly better accuracy results at several distances at once. In winter, at a distance of 30 m, only MobileNet V1 recognized a pedestrian. However, this recognition has a poor level of accuracy. Undoubtedly, what kind of clothes a person is wearing and what background has a huge impact on the accuracy of recognition. Winter clothing is new to models and reduces accuracy. With an increase in the distance, its influence will decrease, and the effect of the background will increase.

Both models do a pretty good job of recognizing a running pedestrian in the summer at a distance of 5 m. However, in winter shootings, the recognition becomes less effective. The difference is especially big for MobileNet V1, where the average accuracy of summer and winter differs by almost 14%. MobileNet V2 has a less noticeable difference, the average accuracy of winter frames is less by 6.6%. With an increase in the distance to 10 m, the average accuracy of winter frames turned out to be even higher than summer ones by 6.3%. While the average recognition accuracy of MobileNet V1 for winter frames at a distance of 10 m was a modest 57.8%. MobileNet V2 does a much better job at recognizing pedestrians running across in winter, and in summer frames the accuracy of the models is approximately equal.

Summing up, we can say that the MobileNet V1 model is still relevant for practical use. It is practically in no way inferior to its new version, especially in performance. The model is very promising for further improvements.

Acknowledgement
The research was carried out at the expense of the Russian Science Foundation grant No. 21-79-10200 at TUSUR.

References
[1] Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z 2016 Rethinking the Inception Architecture for Computer IEEE Conf. on Computer Vision and Pattern Recognition. (Las Vegas: Nevada/ USA) pp 2818–26
[2] Liu T and Stathaki T 2017 Enhanced pedestrian detection using deep learning based semantic image segmentation 22nd International Conf. on Digital Signal Processing. (Birmingham: West Midlands/ UK), pp. 1–5
[3] Chaldina E S, Movchan A K, Kapustin V V and Kuryachiy M I 2020 Multi-Area Range Measurement Method Using Active-Pulse Television Measuring Systems IEEE 21st International Conf. of Young Specialists on Micro/Nanotechnologies and Electron Devices (Novosibirsk: NSTU/ Russia) pp 293-297
[4] NVIDIA Corporation. Jetson Nano Developer Kit https://developer.nvidia.com/embedded/jetson-nano-developer-kit
[5] Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L 2018 MobileNetV2: Inverted
Residuals and Linear Bottlenecks *IEEE Conf. on Computer Vision and Pattern Recognition*. (Salt Lake City: Utah/ USA) pp 4510–20

[6] Walk S, Majer N, Schindler K and Schile B 2010 New features and insights for pedestrian detection *IEEE Conf. on Computer Vision and Pattern Recognition*. (San Francisco: California/ USA) pp. 1030–37