Real-Time Detection and Classification for a 360°-Camera Using a YOLO Algorithm

Tetiana Lavrenko[1*], Ayman Ahmed[1], Vladimir Prokopenko[1], Thomas Walter[1], Hubert Mantz[1]

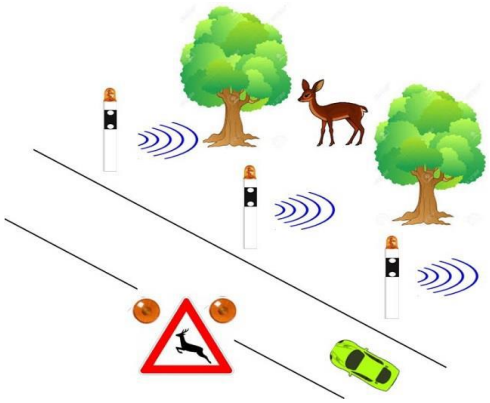
[1]Institute for Mechatronic and Medical Engineering, Ulm University of Applied Sciences, Albert-Einstein-Allee 55,
89081 Ulm, Germany; *tetiana.lavrenko@thu.de*

Abstract. The focus of this paper is the detection and classification of different objects in real time with the help of a 360°-camera. YOLO, a computer vision algorithm, is to be used to perform both the localization and classification of the objects present in the equirectangular panoramic images. The algorithm will be extended in such a way that the angles and directions with respect to the camera are assigned to the detected objects. The results of this work can contribute to enhanced road safety at the locations where many traffic accidents take place due to suddenly appearing road users.

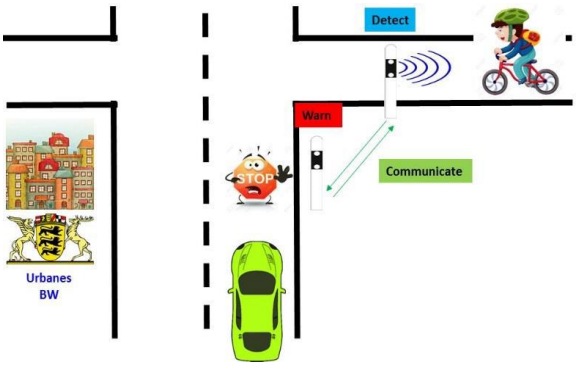
Introduction

A high level of mobility is a basis of highly developed society and economy. However, the current development of the road accidents statistics points out at the fact that vulnerable road users will have to remain a focus of a road safety work in the coming years. [1] As reported by the Federal Statistical Office [2], one in seven people killed in road traffic in 2019 was on a bicycle. If car drivers were involved in a cycling accident with personal injury, bicyclists were mainly to blame in only 23.4% of the cases. Furthermore, traffic accidents on busy roads caused by wild animals are becoming increasingly common. Every year, the Federal Statistical Office reports more than 270,000 accidents involving wild animals. According to the German Hunting Association (DJV), more than one million wild animals are killed in accidents yearly. Most collisions involve deers. [3] Thus, additional measures have to be taken to enhance safety on the roads. One of the approaches would be to inform drivers in advance about a potentially dangerous situation at the spots with increased accident risk where available warning systems fail to inform early enough about suddenly appearing animals, cyclists or pedestrians on the road.

From the aforementioned considerations, two case scenarios can be distinguished: wild animal crossings in the forestry regions and the traffic intersections with obscured side views as has been depicted in Figure 1.



(a) wild crossing



(b) intersection assistant

Figure 1: Scenarios of interest for the discussed application.

Modern computer vision algorithms in combination with 360°-cameras can contribute to the development of new road safety concepts. Nowadays 360°-cameras are experiencing enormous rise related to the increasing pop-

ularity of virtual reality products. They produce 3D-images of the surroundings, which can be stored and/or converted into 2D-projections. These new data formats of videos and images imply new challenges as well as provide numerous possibilities for computer vision and image processing. All-around monitoring systems based on spherical or panoramic recordings could become a new basis for smart road infrastructure covering both case scenarios mentioned above.

In this paper, an experimental setup for real-time detection and localization for a 360°-camera using a YOLO algorithm will be presented. In the context of improved road safety, the computer vision algorithm will be extended with the possibility to estimate an angle of arrival of detected objects assigning them additionally relative geographical directions with respect to the camera's position.

1 Experimental

The discussed setup consists of a GARMIN Virb 360 camera, an Nvidia board for GPU-enhanced artificial intelligence (AI) projects and a YOLO-algorithm based on [4].

The Virb 360 camera is able to record all-around images and videos, which can be also viewed with virtual reality glasses for the "in the middle of it" experience [5]. It has two optical sensors, front and rear, with resolution of 12 Mpx and the objective with the extra wide opening angle of 200°. The maximum resolution of the camera is 5.7k, but for real-time recording and streaming, it is downgraded to 4k.

The Nvidia Jetson Nano developer kit is a small and powerful computer board able to run multiple neural networks in parallel such as image classification, object recognition, segmentation, etc [5]. The Jetson Nano delivers 472 GFLOPs to enable modern AI algorithms to run quickly. This computational capacity can also process data from numerous high-resolution sensors simultaneously with full analysis capabilities. The high performance is facilitated by GPUs – Nvidia graphics processor with the Maxwell architecture and 128 Nvidia CUDA computing units.

Deciding upon which computer vision algorithm to choose for the specified task, the following aspects have been considered. Self-explanatory that a corresponding detection speed of a chosen algorithm has to satisfy real-time requirements as a street situation may change rapidly. Moreover, a multi-object detection capability represents another challenge by itself as well as concerning both a speed and accuracy of the detection. Furthermore, understanding a complete image is of importance, as this should increase the overall accuracy of the object classification and detection.

YOLO is a computer vision algorithm able to detect and localize objects in real time. It is based on the single neural network called Darknet described in [6]. The advantages of this algorithm are the following: it is fast and easy to set up; open source, therefore available for everyone; can be used with other frameworks and libraries such as OpenCV, TensorFlow, PyTorch, etc.; and is highly accurate.

In the discussed experiments, YOLOv3 has been used. The comparison of the algorithm to other detectors has been clearly demonstrated with performance metrics in [6]. The feature extaction network of YOLOv3 employs a hybrid approach between the network used in YOLOv2 [7] and the one of the residual network family ResNet [8]. The phenomenon of residual networks lies in short circuiting shallow to deep layers, also known as shortcuts, resulting in deep neural networks without performance degradation problem. The main conclusion of testing in [6] is that the classifier of YOLOv3 demonstrates the highest measured floating-point operations per second (1457 vs. 1090 billion FLOP/s, the best result of the ResNet family). On the application level, it means that the network uses the GPU resources more optimally compared to the ResNet, thereby increasing own efficiency and speed. Intersection over Union (IOU) is a parameter, which describes the precision of the predicted bounding boxes of the detector network. The detection metrics of mean average precision (mAP) at IOU of 0.5 describing the accuracy vs. speed tradeoff confirms that YOLOv3 also has significant benefits compared to other detection systems demonstrating shortest inference time during testing [6].

The YOLOv3 classification network is trained on the COCO dataset (Common Objects in COntext) that covers 80 different object categories. Among them are cars, trucks, pedestrians, bicyclists that make it possible to implement this detection algorithm in the urban environment as it is.

2 Results

As has been mentioned above, the presented setup aims

at improving road safety in such way that different vulnerable road users, such as pedestrians, bicyclists, wild animals, etc., are detected and localized in the space by an additionally calculated angle/direction of arrival with respect to the camera location. This information can be analyzed together with the location of an approaching car, truck or any heavy vehicle with a further decision to send or not a warning signal to the drivers.

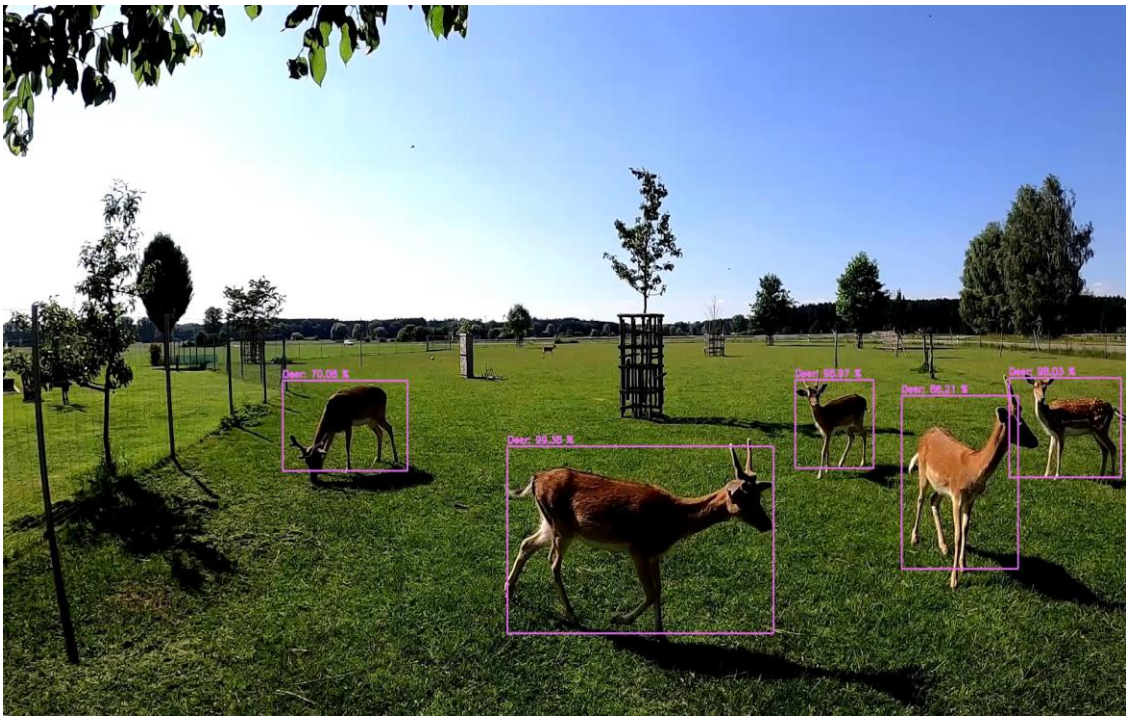


Figure 2: Transfer learning using a YOLO algorithm for detecting deers.

2.1 Animal Detection with YOLO Using Transfer Learning

Considering a scenario with a hotspot of wild animal crossings depicted in Figure 1a, the YOLO algorithm can be retrained for required animal classes in case if the algorithm has not been trained for them initially. It is the case for detecting a deer, as this animal category is not present in the COCO dataset. Using pre-trained networks for further applications saves training time and requires lesser amount of training data. This approach is widely used and well known as transfer learning.

In this work, Darknet was retrained using the Google Colab. It provides the user with sufficiently powerful hardware to perform complex calculations. The training execution is done with commands from the C and Python programming languages. The neural network was trained on 3500 images of deers. The images come from the database "Open Images Dataset V6" under the link https://storage.googleapis.com/openimages/web/download.html. The achieved overall accuracy of detection for the category deer was 97.77%.

The result of transfer learning applied to the YOLO algorithm can be seen in Figure 2. The deer have been detected with high confidence scores and successfully tracked during the whole measurement time. The possibility of detecting multiple objects in one frame is an important advantage of the algorithm. The presented video has been recorded using a normal camera (GoPro Hero 8 [9]), therefore the angle assignment has not been implemented in this example.

2.2 Angle Assignment

Using a 360°-camera provides numerous possibilities for further image and data analysis. Transforming a spherical image into a cylindrical or equirectangular projection re-

tains information about the surroundings in every viewing direction. Such projection ultimately can be treated as a normal image, which has a particular relationship between an image pixel and a spatial direction. Figure 3 schematically describes the approach to estimate the angle of arrival of an object (in the discussed figure a person) with respect to the camera position. The reference point (0° north) is assumed to be in the middle of an

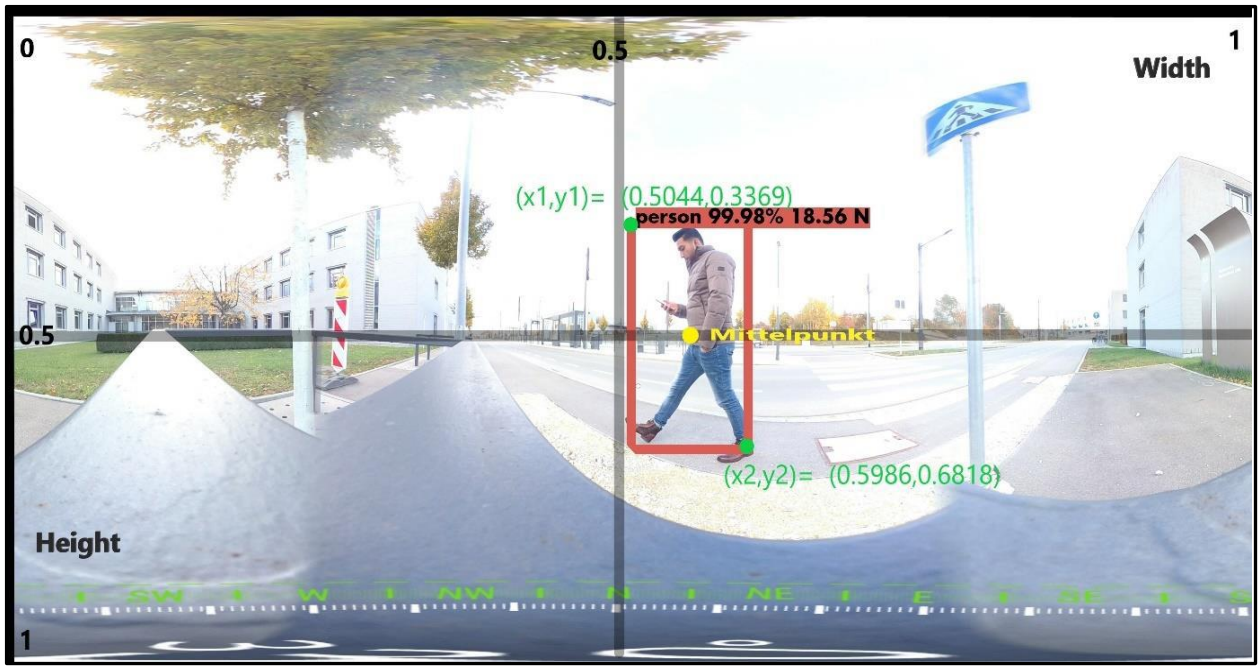


Figure 3: Example of the object detection extended with an angle of arrival.

image (reading 0.5), which is normalized to one (1) in both vertical (height) and horizontal (width) directions. The coordinates of the bounding box predicted by the YOLO algorithm are used afterwards to estimate the object position at every time point resulting in a real-time object tracking. The spatial location of the detected object is equal to the midpoint of the bounding box calculated from its upper left (x1, y1) and bottom right (x2, y2) coordinates. The approach can be validated by overlaying an image with the readings of the digital compass in-built in the camera as can be seen in the bottom part of Figure 3.

2.3 Intersection Assistant

Figure 4 demonstrates the measurement results in the urban surroundings corresponding to the scenario in Figure 1b. The camera was located at the corner of a pedestrian path and a street. The algorithm could successfully detect and localize two pedestrians and two cars that were moving at different angles to each other. The full detection information has been saved into a .txt-file. This file can be used for further data analysis as it contains a current video frame number, a detected category, an angle of arrival as well as an exact timestamp of the detection. At a closer look, one can see that a female pedestrian and a dark car had perpendicular movement trajectories. At an intersection where the driver's view could be obscured due to a construction site or some greenery, this could lead to a potentially dangerous situation. From these considerations, a warning system consisting from a 360°-camera to monitor the surroundings and a computer vision algorithm able to analysis a situation in a global context could contribute to an improved safety at the roads both in the urban environment as well as in the forestry areas.

Figure 4: Object detection using the YOLO algorithm extended with an angle of arrival.

3 Discussion and Conclusion

Safe and sustainable mobility in forested areas as well as in urban surroundings can be ensured by developing smart road infrastructure to reduce the number of car accidents in connection with vulnerable road users. The ideal system would be a setup that only warns drivers when there is a high risk of a potentially dangerous situation to take place. As could be seen from the results presented in this paper a combination of a 360°-camera and a YOLO algorithm can solve this task.

The advantage of the system is its ability to detect and classify multiple objects in real time. By analyzing consecutive frames, a movement direction of an object towards or away from the camera can be estimated. However, the disadvantage is a lack of information on the distance between detected objects and the camera. The respective distances can be provided by integrating a radar sensor in the system, which depending on its modulation type can measure the distance, velocity and direction of objects in front of it. The integration of a radar sensor is planned as a next step in this project. A further advantage of having a radar sensor is that this system should work reliably irrespective of environmental conditions, that is, at night and in the presence of fog when an optical camera is of no use. Moreover, the detection results of two sensors can be compared, and thereby the accuracy of a warning signal can be enhanced. However, the challenge is that the radar measurement data is difficult to interpret in order to classify different road users. Animal walking, similarly to a human movement, is a complex process that comprises different movement patterns of single body parts. This implies a complex analysis of single movement components resulting in the unique radar features that distinguish one road user from another. Therefore, by combining an optical camera and a radar sensor, the difficulty of interpreting the radar measurements can be overcome. The camera recordings can serve as the ground truth for the further analysis of the unique radar features which can be used afterwards as an input for a classification algorithm using a supervised learning. The measurement results of the radar-based detection system will be published elsewhere.

The application for such a real-time detection system combining both optical and radar sensors can be wide-ranging in the context of a smart and sustainable city. One can start from a basic idea to collect mobility data in public urban surroundings and make it available on a data platform for agile urban planning. From another side, the system can be used more specifically for traffic counting

of particular road users irrespective it is urban surroundings or forestry rural regions. Moreover, the installation and control of adaptive lightning along the streets can also be supported by the real-time detection system. Light duration can be adjusted automatically depending whether a fast moving bicycle or a strolling pedestrian has been detected passing by. Within the framework of the InnoSüd project the bicycle counting system in the city Ulm is to be installed based on a real-time detection system combining both optical and radar sensors.

References

[1] adac.de, 10.12.2020. [Online]. Available: https://www.adac.de/news/bilanz-verkehrstote/#:~:text=Der%20bisherige%20Tiefststand%20lag%20nach,Minus%20von%2013%2C2%20Prozent.. [Visited on 11. 02. 2021].

[2] destatis.de, 19. 08. 2020. [Online]. Available: https://www.destatis.de/DE/Presse/Pressemitteilungen/2020/08/PD20_N049_46241.html. [Visited on 11. 02. 2021].

[3] adac.de, 23. 04. 2020. [Online]. Available: https://www.adac.de/verkehr/verkehrssicherheit/tiere/wildunfaelle/#:~:text=Jedes%20Jahr%20meldet%20das%20Statistische,Wildtiere%20bei%20Unf%C3%A4llen%20ums%20Leben.. [Visited on 11. 02. 2021].

[4] AlexeyAB, github.com, [Online]. Available: https://github.com/AlexeyAB/darknet. [Visited on 11. 02. 2021].

[5] Nvidia, [Online]. Available: https://developer.nvidia.com/embedded/jetson-nano-developer-kit. [Visited on 11. 02. 2021].

[6] A. F. Joseph Redmon, „YOLOv3: An Incremental Improvement,“ arXiv:1804.02767, 2018.

[7] J. Redmon und A. Farhadi, „YOLO9000: Better, Faster, Stronger,“ arXiv, 2016.

[8] K. He, X. Zhang, R. Shaoqinq und J. Sun, „Deep Residual Learning for Image Recognition,“ arXiv, 2015.

[9] gopro.com, [Online]. Available: https://gopro.com/de/lu/shop/hero8-black/tech-specs?pid=CHDHX-801-master. [Visited on 12. 02. 2021].

[10] VIRB® 360 Owner’s Manual,“ [Online]. Available: http://static.garmin.com/pumac/Virb_360_OM-EN.pdf. [Visited on 11. 02. 2021].