Object Detection in Mobile Robotics

Valentyna Yakovyn, Taras Kriukov

Department of Information Technology
Vasyl Stefanyk Precarpathian National University
Ivano-Frankivsk, Ukraine

***Abstract*—This paper presents an improved object detection method based on a combination of one-stage object detector followed by an additional classification stage. The article reports an increase in the IoU metric by 8%, classification accuracy by 7%, and f1-score by 5%.The research was conducted using computational experiments and modeling techniques. The results were discussed and interpreted.**

***Keywords*—computer vision, object detection, image classification, machine learning, mobile robotics, software tools**

I. INTRODUCTION

The use of Artificial Intelligence (AI) is a great example of advances, which helps to detect various objects more accurately, which in some cases is important, for example, to identify objects on the road. The Object Detection (OD) is a computer technology whose main purpose is to recognize objects by processing images and highlight them with bounding boxes according to a certain classification. Unusual photos (image size, different viewpoints, mixing the object with the background), real-time detection rate, limited data, and class imbalance can cause significant problems for OD. Our research was triggered by the impressions of the use cases of artificial intelligence in various applications, presented at scientific conferences and challenges. The main purpose of this work is to develop a reliable object detector to be used in mobile robotics. It should detect other mobile robots and distinct them from other (non-robot) objects.

II. RELATED WORK

Today, many classical OD algorithms are used in various fields of human activity. An approach to knowledge-assisted semantic video object detection based on a multimedia ontology infrastructure is presented in [1]. Paper [2] presents an approach to activity recognition based on detecting and analyzing the sequence of objects that are being manipulated by the user. In domains such as cooking, where many activities involve similar actions, object-use information can be a valuable cue. In order for this approach to scale to many activities and objects, however, it is necessary to minimize the amount of human-labeled data that is required for modeling. Paper [2] describes a method for automatically acquiring object models from video without any explicit human supervision. It was shown in the [3] that the You Only Looks Ones (YOLO) algorithm, particularly, YOLOv4, object detection neural network based on the CSP approach, scales both up and down and is applicable to small and large networks while maintaining optimal speed and accuracy. They proposed in [3] a network scaling approach that modifies not only the depth, width, resolution, but also structure of the network. In [4], the authors proposed a novel object detection network by considering the difference between the localization and detection tasks. Algorithms such as YOLO [5], Fast R-CNN [6], HOG, and Mask R-CNN [7] are considered among the best. Each of them has their own advantages, for example YOLO is considered one of the best OD algorithms due to its speed. The processing speed of the image in real time allows processing of many frames per second and newer modifications of the algorithm can process more than hundred frames per second. Paper [8] presents various Object Detection Algorithms such as face detection, skin detection, color detection, shape detection, target detection are simulated and implemented to detect various types of objects with improved accuracy.

III. METHODOLOGY

Participating in the specialized scientific competition "Roborace" is a great way to test the efficiency and accuracy of the neural network, because due to unpredictable movement of the robot camera, we get a blurred image, which can cause problems with recognition and classification of the object. The technical challenge of the competition is to build an autonomous robot [9]-[12]. This includes the resolving of the following problems: in a shortest time the autonomous robot has to pass the set number of laps from the starting position to the finish, avoiding other robots and without going outside the track. The main data input device is the camera. Therefore, we proposed to use the additional classification stage after the OD stage. We can filter out false positives by additional image classification in the constraint frame and selecting a confidence threshold. All the models have been implemented with the open-source library TensorFlow 2. This framework is designed specifically for deep machine learning. Keras was used as a high-level, user friendly framework.

IV. RESULTS

The result of our work is the coordinate detection with an IoU metric increased by 8% in comparison to detection without extra classification. Also, Fig. 1 and Fig. 2 present the classification metrics.

```
[8]: cm = confusion_matrix(y_true, y_pred)
     print(cm)
     print(classification_report(y_true, y_pred, target_names=target_names))

[[472 140]
 [ 80 235]]
              precision    recall  f1-score   support

 Not a Robot       0.86      0.77      0.81       612
       Robot       0.63      0.75      0.68       315

    accuracy                           0.76       927
   macro avg       0.74      0.76      0.75       927
weighted avg       0.78      0.76      0.77       927
```

Fig. 1. Classification report for YOLO detection

```
[40]: cm = confusion_matrix(y_true, y_pred)
      print(cm)
      print(classification_report(y_true, y_pred, target_names=target_names))

[[524  88]
 [ 77 238]]
              precision    recall  f1-score   support

 Not a Robot       0.87      0.86      0.86       612
       Robot       0.73      0.76      0.74       315

    accuracy                           0.82       927
   macro avg       0.80      0.81      0.80       927
weighted avg       0.82      0.82      0.82       927
```

Fig. 2. Classification report for YOLO detection with extra classification stage

V. DISCUSSION

We achieved this result by using a classifier head after OD, which helped reduce the number of false positives. It should be noted that our neural network does not always accurately recognize the selected objects; the reason for this is that at the stage of recognition the network may receive a very indistinct image and the system may not correctly identify the object in the image. Also, the use of our method is limited by the computational hardware of a mobile robot and power and capacity of a power supply.

VI. Future research

In the future, the direction of research will be aimed at improving the detection quality and reducing the training time and computational complexity of the model.

VII. Conclusion

In this paper we proposed the enhanced method for OD for mobile robotics. It is based on the stack of OD and classification heads implemented using artificial neural networks. It provides an increase in the IoU metric by 8%, classification accuracy by 7%, f1-score by 5%.

VIII. Acknowledgment

We are grateful to the researchers of the Department of Information Technology of Vasyl Stefanyk Precarpathian National University for the professional advice provided to us in the research process.

IX. Disclosures

The authors have nothing to declare.

References

[1] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V.-K. Papastathis, and M. G. Strintzis, "Knowledge-assisted semantic video object detection," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 10, pp. 1210-1224, Oct. 2005, doi: 10.1109/TCSVT.2005.854238

[2] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. M. Rehg, "A Scalable Approach to Activity Recognition based on Object Use," *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1-8, doi: 10.1109/ICCV.2007.4408865

[3] C -Y. Wang, A. Bochkovskiy, and H -Y. M. Liao, "Scaled-YOLOv4: Scaling Cross Stage Partial Network," *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 13024-13033, doi: 10.1109/CVPR46437.2021.01283

[4] J. U. Kim and Y. Man Ro, "Attentive Layer Separation for Object Classification and Object Localization in Object Detection," *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3995-3999, doi: 10.1109/ICIP.2019.8803439

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91

[6] R. Girshick, "Fast R-CNN," *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322

[8] A. Raghunandan, Mohana, P. Raghav, and H. V. R. Aradhya, "Object Detection Algorithms for Video Surveillance Applications," *2018 International Conference on Communication and Signal Processing (ICCSP)*, 2018, pp. 0563-0568, doi: 10.1109/ICCSP.2018.8524461

[9] М. Козленко, "Система компенсації гравітаційного впливу на лінійну швидкість автономного мобільного колісного робота," *I Науковий семінар Кіберфізичні системи досягнення та виклики*, 2015, с. 149-157

[10] М. Козленко, "Модуль інерціальної навігації автономного мобільного робота та аналіз характеристик його сигналів," *Вісник Хмельницького національного університету Технічні науки*, № 6, с. 160-164, 2014

[11] М. Козленко, "Характер та симетрія функцій правдоподібності при демодуляції шумоподібних сигналів в мобільній робототехніці," *Вісник Хмельницького національного університету Технічні науки*, № 6, с. 135-139, 2013

[12] M. Kozlenko, "The interference immunity of the telemetric information data exchange with autonomous mobile robots," *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, no. 1, pp. 107–113, 2015