Comparison of Machine Learning methods for detection and recognition of industrial objects and text labels

Piotr Hajder*, Łukasz Rauch

AGH University of Science and Technology
* phajder@agh.edu.pl

An increasingly common practice in industry is the implementation of Industrial Information Systems (IIS) supporting operation in the company. Despite their continuous improvements, human is still one of the most frequently observed causes of malfunctioning of IIS. When designing the IIS, particular attention should be paid to the reduction of human activities, especially critical to the proper functioning of the system. In this work, we present an approach based on machine learning methods and techniques in the detection of objects and text labels in industrial conditions. The results of the work were used to create a complete object and text detection module used in two steel plants. In both cases, this module is one of the components of the entire system.
In the first case, an industrial camera was set to monitor passing wagons. The main task in the system was to find the labels of the wagons, recognize them and return information about the correctness of the UIC checksum.
In the second case, it was necessary to detect the pouring ladle on the crane to determine its location in the blast furnace hall. The ladle tracking subsystem uses an object detection module as a base and is enhanced with image processing algorithms. As an output, detection module receives a set of numbers: points (x,y) defining left upper and right lower corner of a box with detected object and confidence of detection. These parameters are used to calculate exact position of the ladle on the casting hall. The tracking module uses a detection module as a base and is enhanced with image processing algorithms. Labels recognized by the text detection and recognition module were used to verify the melt.
Object detection was implemented using following machine learning models: MobileNetV2 [1], Mask R-CNN [2] and You Only Look Once (YOLO) [4]. In the case of label detection and text recognition we used Attention OCR [3]. We compared these models in order to obtain accuracy of detection and time required to detect an object. The implementation is made using TensorFlow library. As an output, detection module receives a set of numbers: points (x,y) defining left upper and right lower corner of a box with detected object and confidence of detection. Because these modules was only a part of entire IIS, the results were returned to other components via the Apache Kafka-based message bus.

Bibliography

[1] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018.

[2] K. He, G. Gkioxari, P. Dollar and R. Girshick, "Mask R-CNN," 24 January 2018. [Online]. Available: https://arxiv.org/pdf/1703.06870v3.pdf. [Accessed 25 April 2021].

[3] Z. Wojna, A. N. Gorban, D.-S. Lee, K. Murphy, Q. Yu, Y. Li and J. Ibarz, "Attention-Based Extraction of Structured Information from Street View Imagery," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto, 2017.

[4] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 6 May 2016. [Online]. Available: https://arxiv.org/pdf/1506.02640v5.pdf. [Accessed 25 April 2021].