Proceedings of the Hamburg International Conference of Logistics (HICL) – 33

Mathias Rieder and Marius Breitmayer

Object Detection in Picking: Handling variety of a warehouse's articles

HICL

Published in: Changing Tides
Wolfgang Kersten, Carlos Jahn, Thorsten Blecker and Christian M. Ringle (Eds.)
ISBN 978-3-756541-95-9, September 2022, epubli

Object Detection in Picking: Handling variety of a warehouse's articles

Mathias Rieder[1] *and Marius Breitmayer*[2]

1 – University of Applied Sciences Ulm

2 – Ulm University

Purpose: *The automation of picking is still a challenge as a high amount of flexibility is needed to handle different articles according to their requirements. Enabling robot picking in a dynamic warehouse environment consequently requires a sophisticated object detection system capable of handling a multitude of different articles.*

Methodology: *Testing the applicability of object detection approaches for logistics research started with few objects producing promising results. In the context of warehouse environments, the applicability of such approaches to thousands of different articles is still doubted. Using different approaches in parallel may enable handling a plethora of different articles as well as the maintenance of object detection approach in case of changes to articles or assortments occur.*

Findings: *Existing object detection algorithms are reliable if configured correctly. However, research in this field mostly focuses on a limited set of objects that need to be distinguished showing the functionality of the algorithm. Applying such algorithms in the context of logistics offers great potential, but also poses additional challenges. A huge variety of articles must be distinguished during picking, increasing complexity of the system with each article. A combination of different Convolutional Neural Networks may solve the problem.*

Originality: *The suitability of existing object detection algorithms originates from research on automation of established processes in existing warehouses. A process model was already introduced enabling the transformation of laboratory trained CNNs to industrial warehouses. Experiments with CNNs according to this approach are published now.*

First received: 17. Mar 2022 Revised: 25. Aug 2022 Accepted: 25. Aug 2022

1 Introduction

Handling objects in logistics is often supported by loading equipment enabling standardization and automation of processes. Therefore, processes that require a higher amount of flexibility are still carried out manually (EHI Retail Institute, 2019). Such processes are, for example, picking in commissioning, where objects must be processed in amounts less than stored on a loading equipment or outer packaging. Every object category, e.g., cuboids, cylinders, bottles, or non-rigid objects, must be handled according to their special requirements to successfully pick and place the objects without damaging them. Consequently, enabling automated picking and placing in logistics, automation must be guided according to the flexible environment in order to identify a required object, calculate its corresponding gripping point(s), prevent collisions with other objects, storage facility, and the automation components (Wahrmann, et al., 2019). Analyzing images delivered from a vision system can be used to adapt to the environment. Detecting objects in images experienced a boost by using Convolutional Neural Networks (CNN) with suitable computing capacity within the early 2010s (Sultana, Sufian and Dutta, 2020).

This paper contributes to the question of how to implement an object detection system in logistics environments (e.g., warehouses for picking). Therefore, insights from research on object detection algorithms are used to build an object detection system facing logistics' requirements and for handling dynamics in established processes and assortments.

This paper is structured as follows. The second chapter describes related work regarding logistics, picking, and approaches to processes automation. This includes addressing object detection as a prerequisite for automated object withdrawal. Chapter 3 outlines the requirements of a picking system according to an object detection system. In chapter 4 the experimental setup concerning the defined questions is described. Results are presented in chapter 5. The paper then concludes with a discussion, conclusion, and possible future research.

2 Related Work

This chapter addresses two research areas: the process of picking in logistics scenarios as well as approaches leveraging object detection to support the automation of such process.

2.1 Logistics and Picking

A core process in warehouses is picking, which is the customer order specific composition of a subset from a total assortment of goods (VDI, 1994). Especially, this composition is often carried out manually as the number of ordered objects of each order line is smaller than the number of objects stored with a loading equipment. Consequently, this requires a specific handling according to the individual requirements of each single object. Therefore, a survey in 2016 showed that 80% of warehouses are still run manually (Bonkenburg, 2016). To assist humans in picking objects, assistance systems were introduced reducing searching time of objects by pick-by-voice systems (Dujmesic, Bajor, and Rozic, 2018) or smart glasses (Rejeb, 2021). Furthermore, by focusing on humans during the picking process, the goods-to-person principle was introduced in which goods are delivered to humans by automated storage and retrieval systems (de Koster, 2018) or mobile robots (Bozer and Aldarondo, 2018). Amazon Inc. introduced a picking challenge to find trends in robotic retrieval from shelves (Correll, et al., 2016), giving the pick-by-robot approach a boost. This challenge was carried out three times.

These technologies help handling the assortment which ranges, for example at Amazon for German warehouses, from 100,000 to 2,000,000 different articles, depending on their product categories (Schwindhammer, 2022).

2.2 Object Detection

For object detection in 2D-images, a variety of algorithms already exists (Sultana, Sufian and Dutta, 2020). The most used algorithms based on CNNs being Mask Regions with CNN features (Mask R-CNN) (He, et al., 2017), You Only Look Once (YOLO) (Redmon, et al., 2016) and Single-Shot Detector (SSD) (Liu, et al., 2016) including their subsequent developments (Pal, et al., 2021).

Different metrics and data sets were introduced for comparing algorithms for object detection (Padilla, Netto, and da Silva, 2020). Yang, et al. (2020) identified that most data sets provide only few classes for object detection, e.g., COCO data set includes 80 classes (Lin, et al., 2014), ImageNet 200 classes (Russakovsky, et al., 2015) and Open Images Dataset distinguishes between 19,794 classes, but only 600 are annotated with bounding boxes (Kuznetsova, et al., 2020) In the context of industrial settings, however, these numbers of classes are not sufficient as warehouses assortments can consist of thousands of articles.

In general, different challenges for object detection algorithms exist, including handling occlusion (Saleh, Szénási and Vámossy, 2021), the imbalance problem (Oksuz, et al., 2020), and the central or decentral allocation of computation capacities (Ren, et al., 2018). Additional challenges are posed by the context of object detection in logistics scenarios: Pathaka, Pandeya and Rautaraya (2018) stated that there is a lack of data sets for object detection in general. Bormann, et al. (2019), and Thiel, Hinckeldeyn and Kreutzfeldt (2018) confirm the need for training data, particularly in the context of logistics applications. Li, et al. (2018) observed that “there is no public data set of logistics warehouse” and consequently Mayershofer, et al. (2020) introduced Logistics Objects in Context (LOCO) data set for warehouse surroundings like pallets or forklift. In 2015, a special data set for object detection in a warehouse environment was published by Rennie, et al. (2015), focusing on a setup such as Amazon’s picking challenge. Li, et al. (2019) discussed the complex task of detecting pallets in logistics, particularly illumination conditions and object dimensions. Mok, et al. (2021) also focused on detecting pallets, confirming the complexity of object detection in flexible environments such as logistics. Poss (2019) stated, that continuous changes in logistics, e.g., of containers, are problematic for object detection performance.

Object detection results are categorized into True Positives (TP) (correct prediction: correct object class and location), False Positives (FP) (false prediction: false object or incorrect located), False Negatives (FN) (no prediction but image contains searched object) and True Negatives (TN) (no prediction and no known object in the image)

(Padilla, Netto, and da Silva, 2020). Such categorization is achieved using the Intersection over Union (IoU) comparing the area of overlap of the prediction with the expected result with the union of both. Figure 1 displays the approach of IoU and its calculation. According to related approaches, IoU > 0.5 leads to TP categorization.

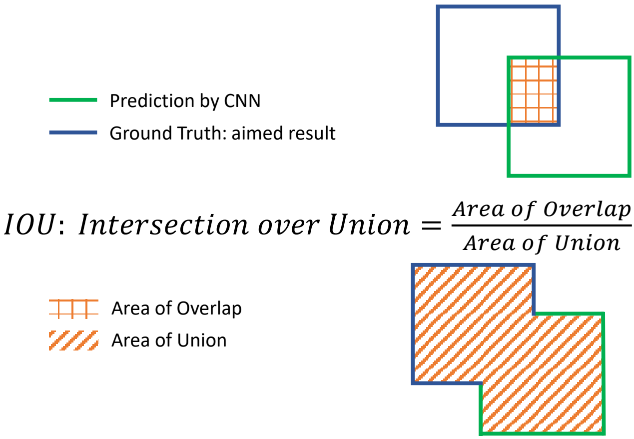

Figure 1: Intersection over Union (modified from Kaggle, 2022)

Categorizing a set of images into TP, FP, TN and FN enables calculating scores for Precision, Recall and F1-score metrics (Hui, 2018):

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Plotting Precision and Recall can be done using a curve. Calculating the Area under the Curve (AuC) gives the Average Precision (AP) (Hui, 2018) also called mean Average

Precision (mAP) in the context of Common Objects in Context data set (COCO) (Lin, et al., 2014).

3 Object Detection in Picking System

In addition to the described discrepancy between number of articles stored in a warehouse and the possibilities to distinguish objects using existing CNN approaches, the topic of changes in a warehouse's assortment has not been addressed yet. The packaging and design of articles, especially in commerce, is changed regularly based on marketing activities or product packaging redesign. Moreover, the assortment within a warehouse is very dynamic, concerning seasonal impact or product lifecycles.

Most publications dealing with object detection, however, neglect such facts. Thus, the dynamic assortments and big number of articles in logistics environments remains unconsidered when designing an object detection system.

In this paper, this issue is tackled by using multiple CNNs to distinguish between all articles. In a nutshell, for every article or article group respectively a CNN is designed. Besides the "positive" images, containing the searched object, "negative" samples must be applied, containing images of all other relevant articles to avoid confusion. Figure 2 gives an idea of the lifecycle of a CNN used in a warehouse for picking. Especially re-training is important to adapt to changes to guarantee a sufficient object detection and picking performance.

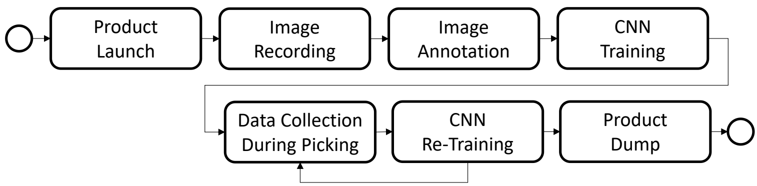

Figure 2: Outline of CNN lifecycle

A setup using multiple CNNs for articles or article groups instead of one CNN for the whole warehouse's assortment bears following advantages:

- Avoidance of framework violation: For YOLO, e.g., the number of articles must be defined before trainings starts (Bochowskiy, 2022). Adding articles later may lead to problems in CNN configuration.
- Re-Training for relevant articles: In case changes occur, only relevant CNNs must be re-trained. These can be defined by applying a confusion matrix to show articles that could be mixed up during object detection. This simplifies maintenance of CNNs during their lifecycle.

Comparing the effect during CNN re-training experiments are defined in Chapter 4.

4 Experiments

A custom data set was designed for first experiments showing effort and effects of the setup described in Chapter 3.

4.1 Data Collection and Preparation

Images were recorded with a Picture Recording Machine (cf. Figure 3), hence, enabling automated recording with a custom definition of number of images at a possible object rotation of 360° and camera movement of 90° each in steps of 1°-movement.

Next, recorded images were annotated using YOLO Mark (Bochowskiy, 2020), and object detection was done using YOLOv4 (Bochowskiy, Wang and Liao, 2020), where 2,000 training iterations for each article of the set is recommended (Bochowskiy, 2022). The training was run on a working station equipped with a Nvidia GeForce RTV 3090. During training the images are augmented. In other words, changes to the images are being applied for training purpose increasing the robustness of trained CNNs with respect to changes in images, lighting, or surroundings. For YOLOv4 MixUp, CutMix, Mosaic, Bluring data augmentation, and label smoothing regularization methods are applied (Bochowskiy, Wang and Liao, 2020).

4.2 Data Set

The data set contains 16 different ceramic cups and is used for initial testing with, showing effort and effects of the setup described in Chapter 3.

For each object, pictures were recorded in 9°-steps on the turning table and 5°-steps with camera movement, resulting in 760 images per object class. Example images for classes one to three are depicted in Figure 4 (surrounding cut off to focus the objects). On the left-hand side with a view of about 45° and with 0° camera view (recording starts from top view) on the right-hand side to emphasis the challenge of object detection dependent on perspective to the object. Figure 5 displays all sixteen articles.

Figure 3: Picture Recording Machine

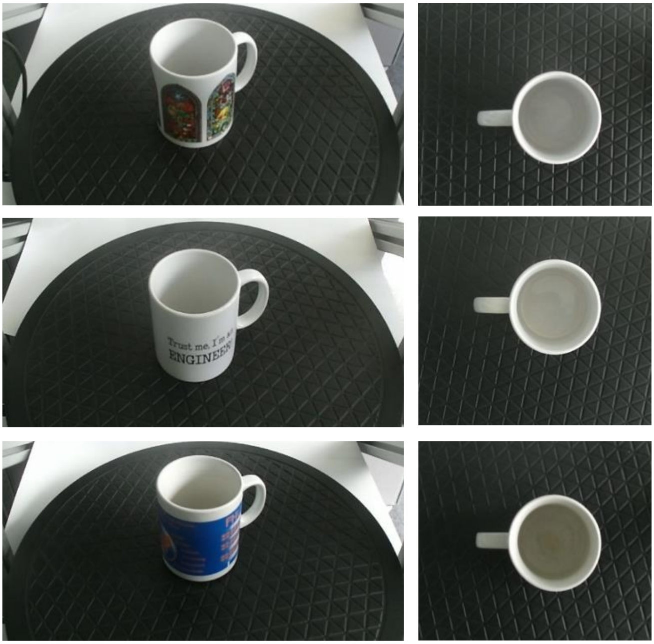

Figure 4: Pictures of ceramic cups (article one - three, from top)

The pictures of the data set are allocated randomly to either training (60%), testing (20%) or validation (20%) subsets. Training and testing subsets are used during training for adjustment of CNN parameters. The validation subset is used for experiments. The separation is done to avoid a CNN to "know" validation images from training. As the distribution to training, testing in validation subsets is done for the whole setup the numbers may differing between the classes.

Figure 5: Pictures of articles one to sixteen, starting in upper left

4.3 Setup

This section describes the setup of the experiments conducted. Figure 6 supports the understanding of follow up sections by describing used CNNs and their configurations.

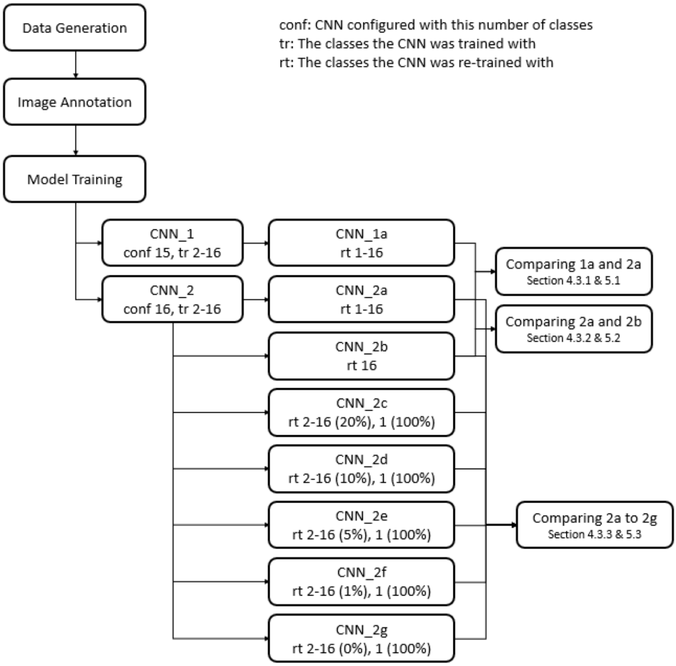

Figure 6: Pipeline of experiments

4.3.1 Extension of number of articles

When training CNNs, first the number of classes (objects to distinguish) must be defined. In case other articles are added at a later stage, the configuration of the CNN must be adapted accordingly. To test the effect of re-training, a YOLOv4 CNN was configured and trained using fifteen classes with object classes two to sixteen (CNN_1). Later, article one was added to the training set for re-training (CNN_1a).

The alternative test is the configuration with sixteen articles but only handing over samples of article two to sixteen (CNN_2) and using all sixteen articles for re-training (CNN_2a).

4.3.2 Use of negative samples

Further tests evaluating the impact of re-training onto object detection performance were conducted: CNN_2 was used to show "unlearning" of a CNN by re-training with images of all classes (CNN_2a) and images of article one only (CNN_2b). The object detection performance was then compared according to TP and FP.

4.3.3 Amount of negative samples

When equipping each article with a CNN begs the questions which images to use for training as training requires images of other articles to avoid erroneous object detection. Considering the number of articles in a warehouse, an additional follow-up question regarding the number of images required to train for one article arises.

Using the result from previous sections, CNN_2 was used as basis and CNN_2a as benchmark. For re-training articles of all sixteen classes were used, differing in the amount of negative samples: CNN_2c with 20%, CNN_2d with 10 %, CNN_2e with 5% and CNN_2f with 1% of training and testing samples as well as CNN_2g without training and testing images of classes two to sixteen.

5 Results

This section presents the results of experiments introduced in Chapter 4. Figures 7-10 display the first 2,000 iterations of training, as biggest changes of loss and mAP occur in this training phase. Training loss is displayed in black color. Additionally, Figures 7-10 indicate the mAP in red color located on the upper right as continuous line, starting with iteration 1,000. In most cases mAP is very low for previous iterations and the mAP calculation starts from iteration 1,000 to safe computation power (Bochowskiy, 2022).

5.1 Extension of number of articles

This section shows the comparison of adding an article to a CNN when configuration must be changed for re-training (increasing the number of classes) (cf. Figure 7) compared to a configuration with the final number of classes at the beginning of the training (cf. Figure 8).

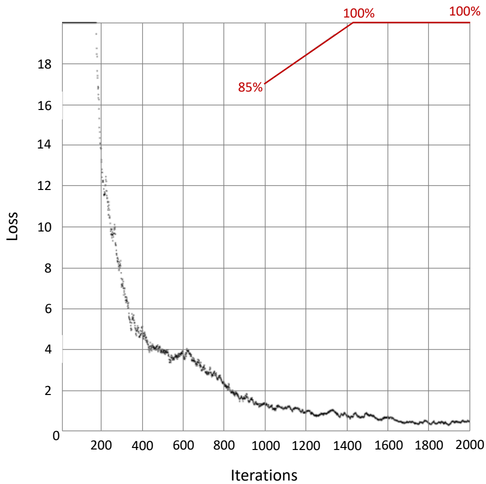

Figure 7: Training of CNN_1

Comparing Figures 7 and 8 shows that by re-training after adding an article in CNN's configuration, training seems to start from beginning. This is indicated by the fact that the course of training loss is similar for Figures 7 and 8. On the other hand, Figure 9 shows the initial training and Figure 10 the re-training resulting in a different course in Figure 10 meaning that the CNN's weights can be refined during re-training (Figure 10) in contrast to re-configuration (Figure 8).

Comparing Figure 7 and 9 regarding to mAP, training with an "empty" class at CNN_2 (Figure 9, no images of class one are used) affects the CNN's detection performance negatively in early training stage as mAP does not reach 100%.

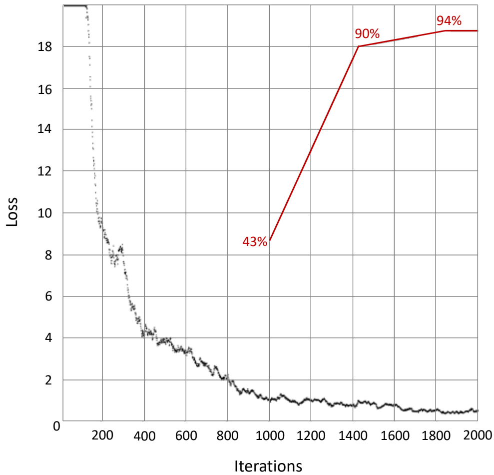

Figure 8: Training of CNN_1a

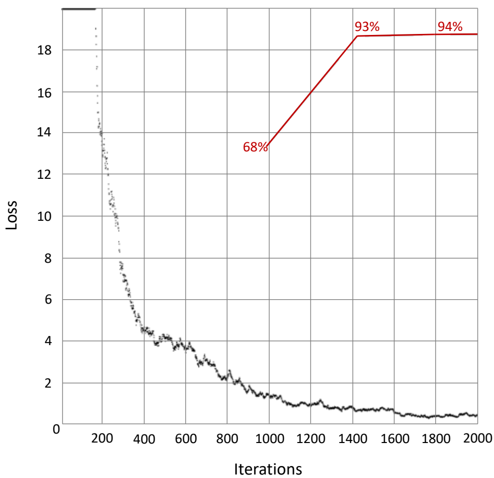

Figure 9: Training of CNN_2

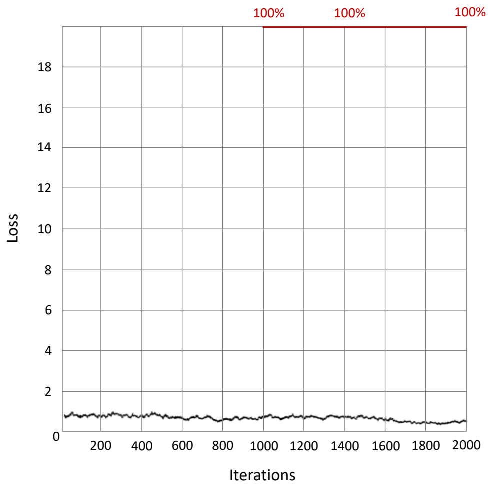

Figure 10: Training of CNN_2a

5.2 Use of negative samples

Numbers in Figures 11-16 are related to the validation data sets to which 20% of the images belong. The distribution for class differs, as distribution was defined by random numbers. Compensating this, presented numbers are relative, providing the rate of TP and FP for different classes in relation to the number of images. A rate higher than 100% results from multiple detections for one image that can occur in early stages of training but normally disappears with training duration.

Figure 11 shows the course of TP and FP for class one and the average for classes two to sixteen over the re-training phase after every 100th iteration. For re-training only images of class one have been used resulting in a constantly decreasing TP-rate for classes two to sixteen.

Figure 11: Re-training without negative samples (CNN_2b)

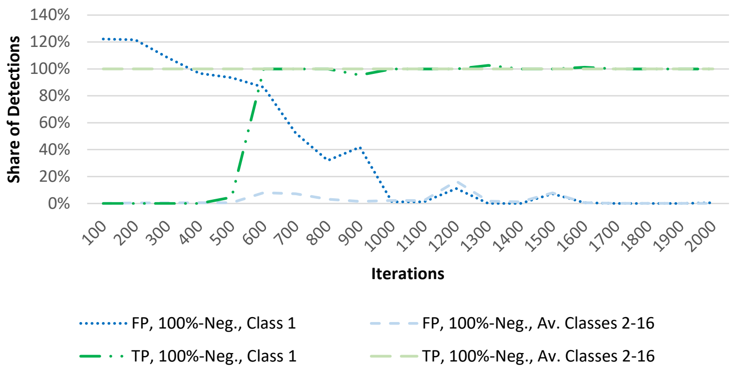

Figure 12: Retraining with 100% of Negative Samples (CNN_2a)

Figure 12 shows the result for the same experiment but using all images off all classes. This results in TP-rates for all classes near 100% and rates of near 0% as well.

Consequently, the data of existing classes is crucial for re-training to remain sufficient object detection performance for these classes.

5.3 Amount of negative samples

This section presents results from re-training a CNN that was trained with images from classes two to sixteen with images of all class. The share of images of classes two to sixteen used varies between 0% to 100% in different steps, all images of class one were used. Figures 13 and 14 show the number of TP for class one (cf. Figure 13) and classes two to sixteen (cf. Figure 14). The lower the number of images of classes two to sixteen, the faster a TP-share of around 100% is reached for class one. For all experiments, except 0%, the number of TP-share for classes two to sixteen remain at about 100% with some outliers above 100% resulting from multiple detections for one image.

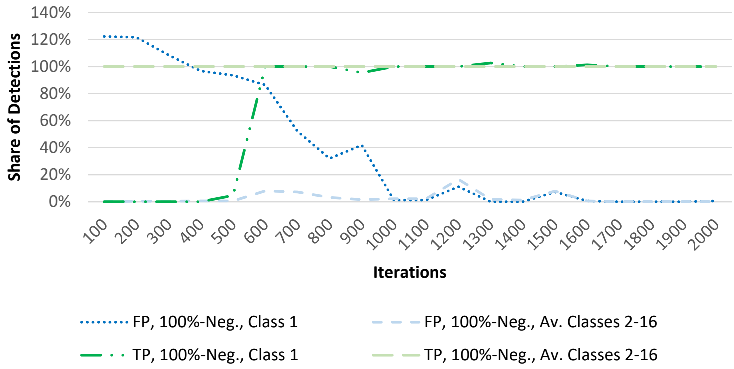

Figure 13: True positive detections for class one

A similar effect regarding FP can be observed comparing Figures 15 and 16. A faster decrease of FP-share of class one results from a higher number of images of classes two to sixteen (Figure 15). The share of FP for classes two to sixteen increase after re-training start near zero but coming back to the area of zero after some peaks.

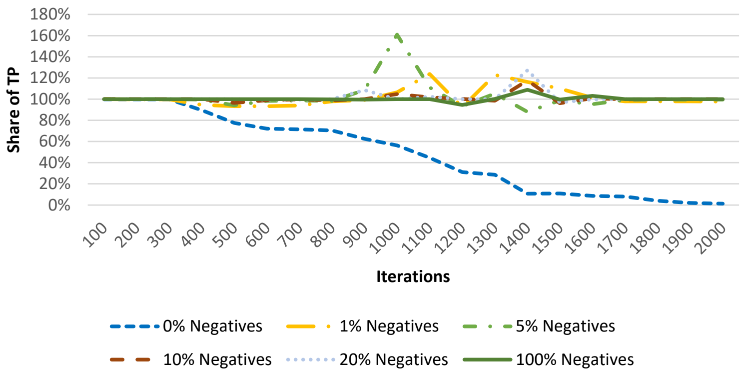

Figure 14: True positive detections for classes two to sixteen one

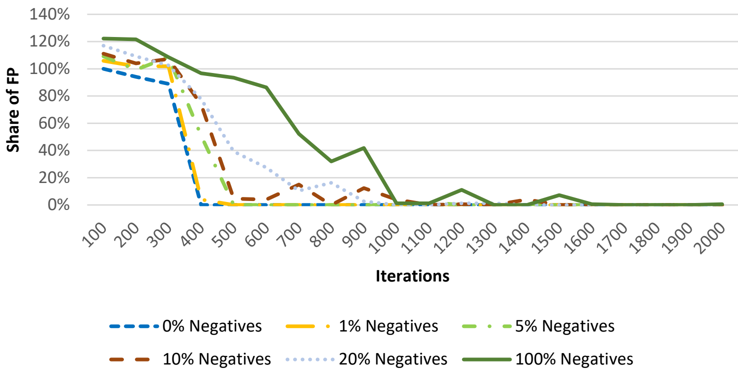

Figure 15: False positive detections for class one

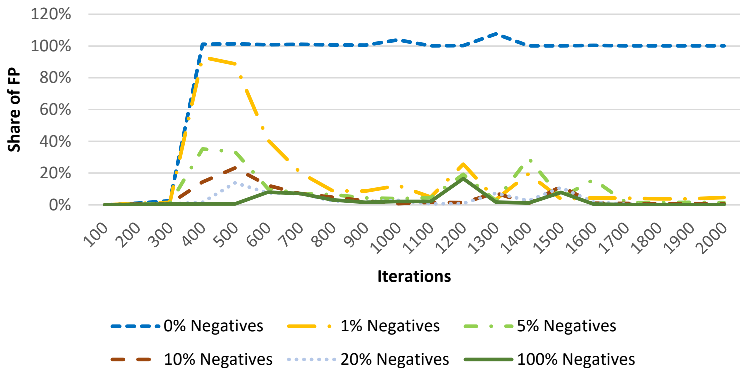

Figure 16: False positive detections for classes two to sixteen

6 Conclusion

This paper introduced state-of-art approaches of automating logistics warehouses and object detection for picking. Further, the requirements for object detection in dynamic logistic scenarios were discussed and from an industrial approach view. Experiments with CNNs examining the configuration and maintenance of CNNs for object detection in warehouse were conducted. Therefore, a custom data set of similar looking ceramic cups was defined and images recorded by a Picture Recording Machine. YOLO algorithm was used to train different CNNs to compare the object detection performance of different CNN configurations.

While the general use of CNNs for object detection is well established, the use of CNNs for object detection in the context of industrial settings can be expended. Existing approaches do not cover industrial settings, and most existing research only addresses the problem regarding a limited number of classes being treated by one single CNN. In the context of product lifecycles, changes to warehouse assortments occur frequently, and remains unconsidered in object detection research. For industrial applications, however, this resembles a serious challenge.

The experiments conducted in this paper provide an idea of how an object detection system for picking in logistics environment may be designed using multiple CNNs instead of one CNN processing the whole assortment. Therefore, different states of CNNs were compared and the impact of increased number of classes as well as the amount of images from known classes during re-training was analyzed. The results indicate that multiple CNNs are suitable for object detection in warehouses if a concept for continuous data gathering and CNN update, respectively maintenance, is applied. The experiments have been conducted in a laboratory environment, but the transformation from a laboratory CNN to warehouse employment was treated yet (Rieder and Verbeet, 2020).

In further research two different domains must be addressed: First, real-world applications in the field of logistics must further validate the presented results. The application of the presented approach to an industrial warehouse can also help to overcome the limitation of using laboratory images only. Furthermore, the number of articles must be increased to a real-world scenario.

Second, further investigations of how multiple CNNs interact with each other must be conducted. This provides the potential that different CNNs might be configured in a less complex way, leading to shorter training phases, increased picking performance and less resource usage in general.

Acknowledgments

This work is part of the project “ZAFH Intralogistik”, funded by the European Regional Development Fund and the Ministry of Science, Research and Arts of Baden Württemberg, Germany (F.No. 32-7545.24-17/3/1).

Many thanks go to Richard Verbeet and Martin Kies for inspiring discussions, collaboration and support.

References

Bochkovskiy, Alexey. Yolo v4, v3 and v2 for Windows and Linux. [online] Available at: <https://github.com/AlexeyAB/darknet> [Accessed 20 May 2022].

Bochkovskiy, Alexey. Yolo_mark: Windows and Linux GUI for marking bounded boxes of objects in images for training Yolo v3 and v2. [online] Available at: <https://github.com/AlexeyAB/Yolo_mark> [Accessed 6 May 2020].

Bochkovskiy, Alexey, Wang, C.-Y. and Liao, H.-Y. M., 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. <http://arxiv.org/pdf/2004.10934v1>.

Bonkenburg, T., 2016. Robotics in Logistics: A DPDHL perspective on implications and use cases for the logistics industry. Bonn.

Bormann, R., Brito, B. F. de, Lindermayr, J., Omainska, M. and Patel, M. Towards Automated Order Picking Robots for Warehouses and Retail. In: , pp. 185–198.

Bozer, Y. A. and Aldarondo, F. J., 2018. A simulation-based comparison of two goods-to-person order picking systems in an online retail setting. International Journal of Production Research, [e-journal] 56(11), pp. 3838–3858. http://dx.doi.org/10.1080/00207543.2018.1424364.

Correll, N., Bekris, K. E., Berenson, D., Brock, O., Causo, A., Hauser, K., Okada, K., Rodriguez, A., Romano, J. M. and Wurman, P. R., 2018. Analysis and Observations From the First Amazon Picking Challenge. IEEE Transactions on Automation Science and Engineering, [e-journal] 15(1), pp. 172–188. http://dx.doi.org/10.1109/TASE.2016.2600527.

DujmešićIvona, N., Bajor, I. and Rožić, T., 2018. Warehouse Processes Improvement by Pick by Voice Technology. Tehnicki vjesnik - Technical Gazette, [e-journal] 25(4). http://dx.doi.org/10.17559/TV-20160829152732.

EHI Retail Institute. Robotics4Retail: Automatisierung und Robotisierung in Handelsprozessen. [online] Available at: < https://www.ehi.org/produkt/poster-robotics4retail/> [Accessed 22 August 2022].

He, K., Gkioxari, G., Dollár, P. and Girshick, R., 2017. Mask R-CNN. <http://arxiv.org/pdf/1703.06870v3>.

Hui, J. mAP (mean Average Precision) for Object Detection. [online] Available at: <https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173> [Accessed 19 May 2022].

Kaggle Inc. Lyft 3D Object Detection for Autonomous Vehicles. [online] Available at: <https://www.kaggle.com/c/3d-object-detection-for-autonomous-vehicles/overview/evaluation> [Accessed 19 May 2022].

Koster, R. B. M. de, 2018. Automated and Robotic Warehouses: Developments and Research Opportunities. Logistics and Transport, [e-journal] 38(2), p. 33–33. http://dx.doi.org/10.26411/83-1734-2015-2-38-4-18.

Kuznetsova, A., Rom, H., Alldrin, N., Uijlings, J., Krasin, I., Pont-Tuset, J., Kamali, S., Popov, S., Malloci, M. and Kolesnikov, A., 2020. The Open Images Dataset V4. International Journal of Computer Vision, [e-journal] 128(7), pp. 1956–1981. http://dx.doi.org/10.1007/s11263-020-01316-z.

Li, T., Huang, B., Li, C. and Huang, M., 2019. Application of convolution neural network object detection algorithm in logistics warehouse. The Journal of Engineering, [e-journal] 2019(23), pp. 9053–9058. http://dx.doi.org/10.1049/joe.2018.9180.

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L. and Dollár, P., 2014. Microsoft COCO: Common Objects in Context. <http://arxiv.org/pdf/1405.0312v3>.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y. and Berg, A. C., 2016. SSD: Single Shot MultiBox Detector, [e-journal] 9905, pp. 21–37. http://dx.doi.org/10.1007/978-3-319-46448-0_2.

Mayershofer, C., Holm, D.-M., Molter, B. and Fottner, J. LOCO: Logistics Objects in Context. In: , pp. 612–617.

Mok, C., Baek, I., Cho, Y. S., Kim, Y. and Kim, S. B., 2021. Pallet Recognition with Multi-Task Learning for Automated Guided Vehicles. Applied Sciences, [e-journal] 11(24), p. 11808–11808. http://dx.doi.org/10.3390/app112411808.

Oksuz, K., Cam, B. C., Kalkan, S. and Akbas, E., 2021. Imbalance Problems in Object Detection: A Review. IEEE transactions on pattern analysis and machine intelligence, [e-journal] 43(10), pp. 3388–3415. http://dx.doi.org/10.1109/TPAMI.2020.2981890.

Padilla, R., Netto, S. L. and da Silva, E. A. B. A Survey on Performance Metrics for Object-Detection Algorithms. In: , pp. 237–242.

Pal, S. K., Pramanik, A., Maiti, J. and Mitra, P., 2021. Deep learning in multi-object detection and tracking: state of the art. Applied Intelligence, [e-journal] 51(9), pp. 6400–6429. http://dx.doi.org/10.1007/s10489-021-02293-7.

Pathak, A. R., Pandey, M. and Rautaray, S., 2018. Application of Deep Learning for Object Detection. Procedia Computer Science, [e-journal] 132, pp. 1706–1717. http://dx.doi.org/10.1016/j.procs.2018.05.144.

Poss, C., 2020. Applications of Object Detection in Industrial Contexts Based on Logistics Robots. Freie Universität Berlin.

Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., 2015. You Only Look Once: Unified, Real-Time Object Detection. <http://arxiv.org/pdf/1506.02640v5>.

Rejeb, A., Keogh, J. G., Leong, G. K. and Treiblmaier, H., 2021. Potentials and challenges of augmented reality smart glasses in logistics and supply chain management: a systematic literature review. International Journal of Production Research, [e-journal] 59(12), pp. 3747–3776. http://dx.doi.org/10.1080/00207543.2021.1876942.

Ren, J., Guo, Y., Zhang, D., Liu, Q. and Zhang, Y., 2018. Distributed and Efficient Object Detection in Edge Computing: Challenges and Solutions. IEEE Network, [e-journal] 32(6), pp. 137–143. http://dx.doi.org/10.1109/MNET.2018.1700415.

Rennie, C., Shome, R., Bekris, K. E. and Souza, A. F. D., 2015. A Dataset for Improved RGBD-based Object Detection and Pose Estimation for Warehouse Pick-and-Place. <http://arxiv.org/pdf/1509.01277v2>.

Rieder, M. and Verbeet, R., 2020. Realization and validation of a collaborative automated picking system.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A. and Bernstein, M., 2015. ImageNet Large Scale Visual Recognition

Challenge. International Journal of Computer Vision, [e-journal] 115(3), pp. 211–252. http://dx.doi.org/10.1007/s11263-015-0816-y.

Saleh, K., Szenasi, S. and Vamossy, Z. Occlusion Handling in Generic Object Detection: A Review. In: , pp. 477–484.

Schwindhammer, T., 2022. Response to Inquiry by Amazon Pubic Relations. [Email] Message to M. Rieder. Sent Thursday 5 May 2022, 00:00.

Sultana, F., Sufian, A. and Dutta, P. A Review of Object Detection Models Based on Convolutional Neural Network. In: , pp. 1–16.

Thiel, M., Hinckeldeyn, J. and Kreutzfeldt, J., 2018. Deep-Learning-Verfahren zur 3D-Objekterkennung in der Logistik.

VDI, 1994. 3590. Kommissioniersysteme. Berlin: Beuth Verlag. <https://www.vdi.de/richtlinien/details/vdi-3590-blatt-1-kommissioniersysteme-grundlagen-1> [Accessed 19 May 2020].

Wahrmann, D., Hildebrandt, A.-C., Schuetz, C., Wittmann, R. and Rixen, D., 2019. An Autonomous and Flexible Robotic Framework for Logistics Applications. Journal of Intelligent & Robotic Systems, [e-journal] 93(3-4), pp. 419–431. http://dx.doi.org/10.1007/s10846-017-0746-8.

Yang, K., Qinami, K., Fei-Fei, L., Deng, J. and Russakovsky, O., 2020. Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy, [e-journal] 104, pp. 547–558. http://dx.doi.org/10.1145/3351095.3375709