Using Artificial Intelligence to Facilitate Assembly Automation in High-Mix Low-Volume Production Scenario

Article *in* Procedia CIRP · January 2022

DOI: 10.1016/j.procir.2022.05.103

Atal Anil Kumar
University of Luxembourg

Peter Plapper
University of Luxembourg

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 107 (2022) 1029–1034

www.elsevier.com/locate/procedia

55th CIRP Conference on Manufacturing Systems

Using Artificial Intelligence to Facilitate Assembly Automation in High-Mix Low-Volume Production Scenario

Alexej Simeth*[a], Atal Anil Kumar[a], Peter Plapper[a]

[a] *University of Luxembourg, 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg, Luxembourg*

* Corresponding author. Tel.: +352-466-644-9540. *E-mail address:* alexej.simeth@uni.lu

Abstract

Artificial Intelligence (AI) based algorithms are being used increasingly to support industrial robots in the automation of assembly processes. The objective of this work is to detect bores of different geometry, appearance, and inner structure for automating a high-mix low-volume assembly line. Two most widely used AI algorithms namely, Single Shot Detection (SSD) and You Only Look Once (YOLO) have been used to perform the bore detection process. The results obtained by these algorithms have been compared with the conventional detection algorithm (Gradient filter) using the standard metrics used for evaluating the performance of the detection algorithms. The obtained results demonstrate the efficiency and robustness of AI-based algorithms for the detection as they exhibit better performance than the conventional detection method.

Keywords: Artificial Intelligence; automation; Single shot detection; You only look once; high-mix low-volume;

1. Introduction

Assembly process plays a significant role in manufacturing systems as it is one of the last processes within a production operation [1]. A competent manufacturing line is outlined by a highly efficient and optimized assembly process [2]. The present condition of assembly operations is characterized by an increasingly changing demand (mass customization) while the manufacturing line faces a trade-off between the increased output of automated assembly systems and the flexibility and adaptability of manual assembly.

This increase in changing demand caused by shorter product life cycles and higher customization has resulted in increased variance in product portfolio with low volumes of up to batch size one. Complete variance must be covered for every product which makes the automation of the assembly line difficult. It is not economical to stop the line for changing the tools, positions and process parameters for small volumes with high product variance [3]. The introduction of customized products also means that certain process parameters are not known in advance and need manual intervention for the assembly process. With the advancement in Artificial Intelligence (AI) algorithms, it is now possible to deduct decisions and actions from unknown multi-dimensional correlations in sensor data [4], which can then be made available for automation of customized individual products. One scenario where the demand for customization has introduced challenges in the production line is the pick and place process. Pick and place can be defined as the act of picking things up from one location and placing them in another [5], and requires a high level of precision and repeatability [6]. The development of robots with enhanced technologies has helped in building systems that are smarter and more flexible. Despite this, there are several tasks in a pick and place sequence where direct human intervention is still being required due to the excellent capability of a worker to observe and compute the object and interact with the surrounding environment [7].

Vision-based object detection for various steps in a pick and place operation has been proposed by numerous researchers since many years now [8]. Normally, the manufacturing lines have the sequence of pick and place (which includes the perception task) programmed in advance. However, when the assembly scenarios are presented with a batch size of one with random objects being placed for detection and assembly, the task becomes challenging and requires a flexible and robust solution. With the advancement in computational power and AI-algorithms, vision-based approaches have improved tremendously to solve complex problems that have had no effective

10.1016/j.procir.2022.05.103

solutions in the past. One of the competent approaches is the Neural Network (NN) based object detectors [9]. Two of the efficient and high performing NN-based object detectors, Single Shot Detector (SSD) and You Only Look Once (YOLO) are gaining a lot of popularity in recent years.

We present a comparison of these NN-based methods with conventional detection methods applied to localize bores in wooden-like boards with varying features. The methods shall be used to automate an assembly process for highly individual products consisting of a pick and place, and gluing process in a high-mix low-volume assembly scenario. Workpieces are joined by placing them into bores in a carrier workpiece and then bonded with glue. In the first step, the authors focused on the gluing process automation which is presented in [10]. As the next step, for the overall automation, a flexible pick and place process working on every workpiece is targeted. Bores have to be recognized on wooden-like boards with varying features to insert specific workpieces.

The rest of the paper is organized as follows: section 2 highlights the current state-of-the-art indicating why we selected the two NN-based algorithms, section 3 provides the overview of the experimental set-up and the four different boards used for the study, section 4 presents the results of the detection obtained using SSD and YOLO, while the comparison of the obtained results with the conventional gradient filter detection is presented in section 5. The last section 6 of the paper summarizes the results and lists the future work to be done to perform the final task.

2. State-of-the-Art

NN-based object detection algorithms have gained a lot of interest from the researchers all over the world in the last decade [11]. There are mainly two kinds of NN-based object detectors [12]. The first kind is a two-stage detector in which the detection architecture involves an object region proposal (region of interest, ROI) with conventional computer vision methods or deep networks, followed by object classification based on features extracted from the proposed region with bounding box regression. Two-stage methods have the highest detection accuracy but are typically slower [9]. The second kind of NN-based object detectors are the one-stage detectors which predict bounding boxes over the images without the region proposal step and as a result consume less time and are more suited for real-time applications. One-stage detectors are in general, superfast and structurally simpler. One of the efficient one-stage detectors was proposed by Wei Liu et al. [13]. The authors proposed a novel method of detecting objects in images using a single deep neural network and called this detection method as the Single Shot Detector (SSD). This detection method is simple, easy to train, and is very straightforward. Several improved versions of SSD algorithms can be found such as DSSD, FSSD, or ASSD [14].

Another widely used one-stage detector was proposed by Redmon et al. [15] known as YOLO, an acronym for You Only Look Once. It uses regression problem to solve the object detection. Improved versions of YOLO have been developed such as YOLO v2 [16], v3, v4, and YOLOX [17]. Based on the requirements of the study (to have a robust and available algorithm for all types of board and to have a real-time implementation), we decided to implement earlier versions of one-stage detectors, specifically YOLO v2 and the SSD algorithms for our work.

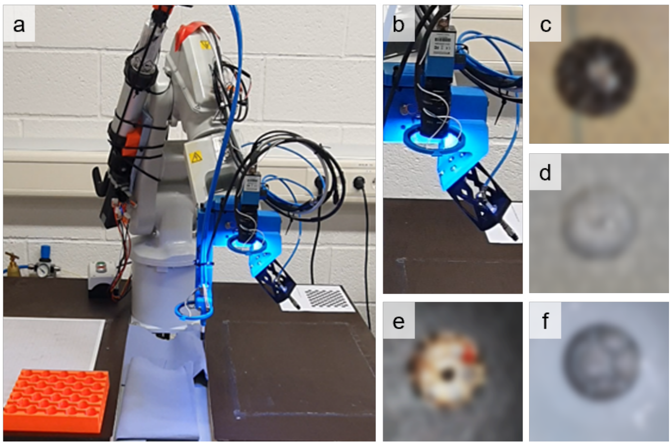

Fig. 1. (a) Overall setup. (b) End effector with camera, ringlight, and vacuum gripper. (c-f) Example of the boards 1:c-4:f used for study with bores

3. Methodology

3.1. Experimental setup

For this research, the selected pick and place process is applied in a batch size one production scenario, i.e., every product is different regarding its predecessor and successor. The considered process is conducted on the test set-up depicted in Fig. 1. In (a), the overall robotic system is shown. A camera system, vacuum gripper, and glue nozzle are mounted on the end effector of an industrial ABB robot installed on a workbench. The orange carrier is utilized to provide workpieces to be inserted. The boards are provided in the empty workspace on the right side. Behind the board area, a chessboard pattern is used for camera-robot calibration following the method presented in [18] so that pixel coordinates can be transformed to real world positions. A close-up of the end effector is given (b). In the images (c-f) examples of the four different boards used in this research are shown. The boards vary in material, dimension, color, inner structure, amount and dimensions of bores, and appearance. For instance, color and dimensional variation ranges from metallic white to matt black and +/-15 mm in bore diameter. It is further possible, that debris from previous milling process of the bores may stay inside them (cf. (c)).

3.2. Data set generation

To obtain realistic images for the detection model development a scenery imitating the assembly line is created. An industrial 5-megapixel RGB camera with lighting equipment is mounted on an industrial robot and the board is placed inside the robot's workspace. Images are taken from above perpendicular to the board surface (Fig. 1 (a)). From the original image a centred 1526x1526x3-pixel image is cropped for further processing. The taken images contain a varying number of bores. The four considered boards differ in material, surface texture, reflectiveness, and color. Additionally, the number of bores per image, the bore diameter, the bore appearance, and the inner structure of the bore and the overall lighting situation are changing (Fig. 1 (c-f)). The high number of variants are used to test the versatility of the applied algorithms.

For the data sets for SSD and YOLO (s. following subchapters) different numbers of images are taken for the considered boards. For board a) 212, board b) 128, board c) 182, board d) 135 images. The data sets for SSD and YOLO are the same. Each image is labelled with the positions of the bore centres and according bounding boxes. The data sets are split into test and training data in the ratio of 80:20. The training data is augmented via reflection, rotation, and changing brightness and contrast. In total six images are created out of one increasing the training set size. For the conventional detection approach 125 images of the same data sets are used for setting up and testing. The settings of the filter are optimized on 25 images manually. With the identified settings the filter is tested on 100 other images. For each board new filter settings have been generated.

3.3. Selected algorithms

The main target of the object detection algorithm is the robust localization of bores on images of boards. The algorithm should work with all board types and be applicable in a process in real-time. For later implementation, detection time is a key parameter since it will have an impact on tact time. Both conventional and AI-based methods are used to localize bores and derive a new position for robot manipulation. In pretests, a gradient edge detector as one of the robust conventional detection algorithm was identified as suitable method for bore detection since it was found to be working with all boards with a fast processing time. Other tested algorithms as Circular Hough Transform, or Gabor filter indicated a lower performance and hence have not been presented in this study.

3.3.1. Gradient filter

Derivative filters are used to detect discontinuities in images. The filters are designed to give no response in smooth regions of an image and to return significant values at points of abrupt changes in intensity level or texture, indicating an end or a start of a region. Such identification of jumps in intensity values is also known as edge detection. Classical gradient edge detectors apply matching of local image segments with specific edge patterns. Edges are detected by searching for maxima or minima in the first derivative (gradient) of the image [19].

3.3.2. SSD

The SSD detector directly performs convolutional prediction on the underlying feature extraction network and extracts feature map characteristics with different scales. For defined locations in the image the SSD predicts a set of default bounding boxes and confidence scores for each class. During training and evaluation, the identified bounding boxes are matched with the ground truth using intersection over union (IoU) [9, 13]. In contrast to the original SSD setup, the backbone, i.e. the feature extraction network, is changed from VGG-16 to the 50-layer deep pretrained convolutional neural network (CNN) ResNet50, which is trained on >1 million images from ImageNet database with 1000 object categories [20]. In conducted pretests for this study, ResNet101 did not yield significant benefits over ResNet50 while reducing classification speed. Thus, the smaller CNN was selected.

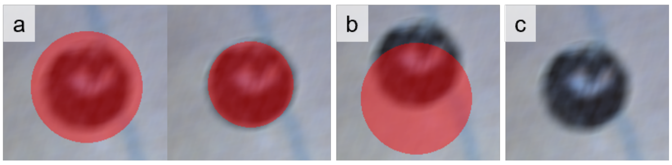

Fig. 2. Example of a) true positives, b) false positive, c) false negative (no detection)

3.3.3. YOLO

The basic idea of YOLO is the transformation of the object detection problem into a regression problem. A single neural network is applied to a full image. The image is divided into an SxS grid and for each cell of the grid the network directly returns the target bounding boxes and its classification categories [16]. The whole image is processed at a time. As conducted for the SSD approach, YOLO v2 is implemented with a changed backbone to CNN ResNet50.

3.4. Evaluation metrics for comparison

The result of the bore detection and localization on the image is evaluated based on the metrics precision and recall. In this case, true positive (TP) is defined as the number of actual positive cases, i.e. a bore is detected as a bore at the correct position. False positive (FP) describes the number of positions which are incorrectly detected as a bore. False negative (FN) is the number of bores, which are not detected as such, but as background. True negatives have not been considered since the focus was in detection of the bores. Examples of TP, FP, and FN are given in Fig. 2.

The calculation of the metrics for the conventional approaches is based on a visual inspection of every classification result. The learning-based algorithms are evaluated automatically. The predicted bounding box is compared with the original ground truth bounding box and counted as TP if IoU is equal or higher than 0.5 and as FP otherwise. In the following this threshold is indicated by IoU@0.5. Precision, recall, and F1-score [21] are calculated for every image and the detection time is recorded and averaged over all detections.

4. Results

4.1. Gradient filter

The detection of the bores using the gradient filter is conducted according to the following process. At first, the images are converted to grayscale and then reduced in size during preprocessing. The gradient magnitude of the intensity values is calculated with a standard Prewitt gradient operator. The result is scaled to the interval [0,1], small details are amplified using logarithmic function and the image is binarized. This process cannot be held constant for all boards but is adapted to achieve a better detection result. After several image processing operations (area opening, filling holes, image dilation) the bores are identified by calculation of the roundness of objects remaining in the image. The final mask is generated by approximation of circles for the detected regions and the calculation of their centroids. The results for all boards for the gradient approach are given in Table 1. The detection time is approx. 17 FPS for every board. Compared to other filters tested, this is the fastest conventional filter. The detection results for board 1, 3, and 4 are similar, whereas, the result for board 2 is significantly worse. Especially, the recall is reduced and thus, more bores are not detected on board 2 than on the other boards.

Table 1. Results of gradient approach averaged on prediction level

Board	Precision	Recall	F1-score	Accuracy	FPS
1	0.9639	0.9423	0.9530	0.9102	16.64
2	0.8865	0.8210	0.8525	0.7429	16.19
3	0.9278	0.9890	0.9574	0.9184	17.27
4	0.9342	0.9748	0.9541	0.9122	17.81
Avg.	0.9281	0.9318	0.9293	0.8709	16.98

4.2. Neural network based algorithms

Both NN-based algorithms are trained on the same data set. I.e., the images per board are split into training and test-set and then used for both detectors. The proposed detectors, SSD and YOLO, are both implemented with the CNN ResNet50. The training of the detectors are done with 25, 50, 75, and 100 epochs. To use the same algorithm for every board without any changes, the setting with the highest average performance is presented in the following sections, which are 75 epochs for both SSD and YOLO. The detectors are trained in different settings. Firstly, they are trained on each board separately and used to classify solely this board type. These results are indicated by the board index (1-4). Secondly, the detectors are trained on all boards and used to classify all test sets. The entry 'All, 1 class' refers to the category-agnostic classifier, i.e. a single class 'bore' for all bores, whereas the detector 'All, 4 classes' has for each board a different bore class.

4.2.1. SSD

The results of SSD detector are given in Table 2. In the table, the mean average precision, recall, and the F1-score are given to evaluate the classification result for IoU@0.5. The quality of localisation is represented by the IoU-score. The classification speed is listed in the last column. The evaluation metrics indicate a good prediction quality for most of the SSD classifiers trained on one board only. Precision and recall are on a high level and the F1-score is on average above 95%. Only for board 2 the recall drops below 90%. The precision for board 1, 2, and 3 is on a similarly high level. Evaluating the classifiers trained on all boards, only the category-agnostic detector achieves comparable results. F1-score is slightly reduced compared to the average of four single board detectors, but precision and recall are in same range. The results of the SSD detector with four bore classes deteriorated significantly. Although the precision is still high, most of the bores are not identified indicated by the low recall value. Especially on boards 2 and 4 the detector misses nearly all bores. Thus, this detector is not considered further in this research. The highest average IoU-score of 0.90 and the most accurate localization of SSD is found at the category-agnostic classifier. The average IoU-score for SSD trained on single boards is 0.89. The prediction time is similar for all SSD detectors and varies between 32-35 FPS.

Table 2. Results of SSD detector for IoU@0.5 with ResNet50 and 75 epochs of training

Board	Precision	Recall	F1-score	IoU-score	FPS
1	0.9759	0.9701	0.9730	0.8912	33.02
2	0.9859	0.8861	0.9333	0.8793	32.24
3	0.9430	1.0000	0.9707	0.8858	32.79
4	0.9728	0.9051	0.9377	0.8943	32.42
Avg.	0.9694	0.9403	0.9537	0.8876	32.62
All, 1 class	0.9172	0.9724	0.9440	0.8977	34.81
All, 4 classes	0.9696	0.3450	0.5089	0.8979	34.01

4.2.2. YOLO

In this research, the YOLO version 2 is applied with a modified backbone to the CNN ResNet50, which is named YOLO detector in the following. The results of the detector are presented in Table 3. For all detectors applied on a single board only, recall is close to one, on average 99.7%. There is more variation in the precision, but all values are above 90%. Thus, nearly all bores on all boards are detected, but some predictions are false. The severity of a false detection is depending on the location of falsely detected bores (centre vs. rim of image). A random sample visual inspection of predictions revealed that the falsely detected bores are without exception a second bounding box for a bore at or close to the image border, which is, for the presented use case not problematic. The detectors trained on all boards achieve different results. Whereas, the detector with four classes has similar values for precision, recall, and F1 as the average of the four detectors trained on single boards, the category-agnostic YOLO detector is on a significantly lower level and thus not further considered.

The IoU-score of the predictions is on average 0.86. The highest value is achieved by the YOLO detector trained on board 3. The prediction time of all YOLO detectors in the conducted trials is similar and in the range of 33-36 FPS.

Table 3. Results of YOLO detector for IoU@0.5 with ResNet50 and 75 epochs of training

Board	Precision	Recall	F1-score	IoU	FPS
1	0.9003	1.000	0.9475	0.8606	35.54
2	0.9349	1.000	0.9644	0.8347	35.31
3	0.9783	0.9890	0.9836	0.9004	33.14
4	0.9294	1.000	0.9634	0.8739	35.87
Avg.	0.9357	0.9973	0.9652	0.8674	34.97
All, 1 class	0.8220	0.8882	0.8538	0.7402	35.78
All, 4 classes	0.9275	0.9988	0.9618	0.8931	36.01

5. Comparison

The presented approaches Gradient, SSD, and YOLO achieve reasonable results depending on the boards. An overview of the F1-score of selected detectors is depicted in Fig. 3. ‘Single’ refers to detectors trained on one board only whereas ‘All’ indicates that the detector is trained on all board images with either one object class (1C), or four classes (4C). All detectors shown achieve on average for all boards a score above 0.92. Outliers are the gradient on board 2, and ’SSD All 1C’ on board 3. On board 2, the bore detection via the gradient filter yields an F1-score of 0.85, which is significantly lower. Especially the recall drops to 0.82. Thus, several bores are not detected. In 9% of the tested images, only the half of the bores or less are identified. A reason for the compromised performance on board 2 may be the high similarity in appearance of surface and bottom of the bore, which is challenging when the detection is based only on changes in intensity values.

The detectors ’SSD Single’ and ’YOLO Single’ perform well for every board. YOLO has a higher recall while, SSD has a higher precision, i.e., YOLO detects more bores but has a higher share of false positives. These, however, are multiple bounding boxes detected for the same bore at the outer part of images, which is not critical. Comparing the difference between different YOLO detectors, the results are for every board very close. For SSD, the deviation between boards is higher. On board 2, ’SSD All 1C’ outperforms all other presented approaches, but on board 3 the F1-score is 0.86, which is the lowest score of all applied detectors. Overall, the YOLO results are most consistent and achieve the highest average F1-score for both detectors.

Beside the classification, the localisation of the detected bores has an immanent role for the overall result. The detection quality for the gradient approach is evaluated manually. For the AI-approaches, the IoU-score is applied. In Fig. 4. a) the IoU

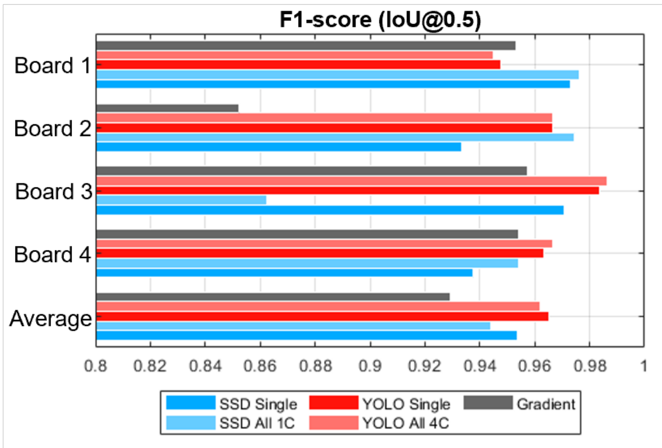

Fig. 3. F1-Scores of bore detectors for IoU@0.5. ’SSD Single’ and ’YOLO Single’ are the detectors trained on one board only. ’SSD All 1C’ is the category-agnostic SSD detector, ’YOLO All 4C’ the detector with four object classes.

scores of presented detectors are given. The values of detectors trained on all boards are for both SSD and YOLO more accurate than ones trained on a single board. SSD detectors indicate a slightly improved localisation accuracy over YOLO, but for the ‘All’ detectors the difference is minimal as the IoU-scores are on average between 0.89-0.90. A reference example on a detection with similar IoU is given in Fig. 4. b). Compared to a positive detection result of the gradient detector depicted in Fig. 4. c) the localisation is slightly less accurate. Considering the bounding boxes itself, the difference of the bounding box centres is less than 6 pixel for the given example. Regarding the processing time, the AI-based approaches are approximately two times faster than the conventional approach with 36 FPS for YOLO, 34 FPS for SSD and 17 FPS for the gradient detector.

In a last step, the detection results of Gradient, SSD, and YOLO were utilized to place workpieces into identified bores with the robot (cf. Fig. 1). Initial trials showed that it is possible to insert workpieces with all three methods. The placement itself is evaluated based on outcome (successful or not) and occurrence of contact of the inserted workpiece with the surface of the carrier workpiece during placement. The initial trials leave the impression that the placement with the conventional Gradient method is more accurate. The assessment of the impact of the detection method on placement is still ongoing and subject to further studies.

6. Conclusions

This paper compares different neural network (NN) based and conventional methods to robustly detect production features (bores) on different boards to be applied in pick and place process in a batch size one assembly automation. The detection results are then used to compute the bore positions where workpieces shall be inserted. As NN-based algorithms, Single Shot Detector (SSD) and You Only Look Once (YOLO) are selected and compared against a conventional Gradient filter. All

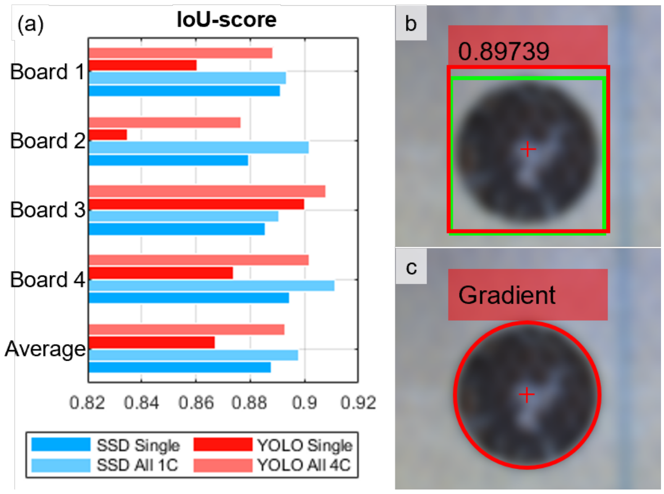

Fig. 4. a) IoU-scores of selected AI detectors. b) localisation result of AI detector with 0.9 IoU, the detected bounding box in red and the original bounding box in green. c) and conventional gradient approach.

methods are trained or optimized on the same generated data set based on images taken of the real product. For comparison of the approaches regarding classification quality, localisation accuracy, and speed, the metrics F1-score based on precision and recall, intersection over union (IoU) and visual inspection, and frame rate are utilized.

For all variants the NN-based detectors work well. When trained on all boards, they achieve good results on all boards. For gradient, this is not possible and certain pre- or postprocessing steps must be adapted according to the examined board. The classification results for all applied methods are on a high level ranging from 0.93 for gradient to 0.94 for SSD and 0.96 for YOLO but slightly improved with NN-methods. Regarding localization, the NN-approaches achieve in average an IoU of 0.9. Comparing the results with gradient detection the distance between the centres is on average $\sim$ 6 pixels. The detection speed of NN-detectors is significantly higher working with double frame rate on same machine. Applying the presented detection methods on a real robot, the conventional filter achieves a higher accuracy. Overall, it is shown, that fast one-stage detectors can be trained and applied on products with a high difference in appearance and used for the classification and detection tasks, which will improve the overall automation potential, especially for high-mix low-volume production scenarios. In future, the authors target to improve the NN-based detection to increase the location accuracy, which can be achieved by implementing updated versions or to predict instances instead of bounding boxes.

References

[1] David Sanders and Alexander Gegov. AI tools for use in assembly automation and some examples of recent applications. *Assembly Automation*, 2013.

[2] Jose Luis Outón, Iván Villaverde, Héctor Herrero, Urko Esnaola, and Basilio Sierra. Innovative mobile manipulator solution for modern flexible manufacturing processes. *Sensors*, 19(24):5414, 2019.

[3] Peter Burggräf, Johannes Wagner, Matthias Dannapfel, Sarah Fluchs, Katharina Müller, and Benjamin Koke. Automation decisions in flow-line assembly systems based on a cost-benefit analysis. *Procedia CIRP*, 81:529–534, 2019.

[4] Rajendra Akerkar. *Artificial Intelligence for Business*. SpringerBriefs in Business. Springer International Publishing, Cham, 2019.

[5] Yong Zhang, Brandon K Chen, Xinyu Liu, and Yu Sun. Autonomous robotic pick-and-place of microobjects. *IEEE transactions on robotics*, 26(1):200–207, 2009.

[6] James M Ritchie, Richard G Dewar, and John E L Simmons. The generation and practical use of plans for manual assembly using immersive virtual reality. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 213(5):461–474, 1999.

[7] Ming-Yu Liu, Oncel Tuzel, Ashok Veeraraghavan, Yuichi Taguchi, Tim K Marks, and Rama Chellappa. Fast object localization and pose estimation in heavy clutter for robotic bin picking. *The International Journal of Robotics Research*, 31(8):951–973, 2012.

[8] Chien-Ming Lin, Chi-Yi Tsai, Yu-Cheng Lai, Shin-An Li, and Ching-Chang Wong. Visual object recognition and pose estimation based on a deep semantic segmentation network. *IEEE sensors journal*, 18(22):9370–9381, 2018.

[9] Licheng Jiao, Fan Zhang, Fang Liu, Shuyuan Yang, Lingling Li, Zhixi Feng, and Rong Qu. A Survey of Deep Learning-Based Object Detection. *IEEE Access*, 7:128837–128868, 2019.

[10] Alexej Simeth, Jessica Plaßmann, and Peter Plapper. Detection of Fluid Level in Bores for Batch Size One Assembly Automation Using Convolutional Neural Network. In A. Dolgui, A. Bernard, D. Lemoine, G. von Cieminski, and D. Romero, editors, *Advances in Production Management Systems. Artificial Intelligence for Sustainable and Resilient Production Systems. IFIP International Federation for Information Processing. APMS 2021, IFIP AICT 632.*, chapter IFIP AICT, pages 86–93. Springer International Publishing, Cham, 2021.

[11] Tuan-Tang Le and Chyi-Yeu Lin. Bin-picking for planar objects based on a deep learning network: a case study of USB packs. *Sensors*, 19(16):3602, 2019.

[12] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232, 2019.

[13] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. SSD: Single Shot MultiBox Detector. pages 21–37. 2016.

[14] Jingru Yi, Pengxiang Wu, and Dimitris N Metaxas. ASSD: Attentive single shot multibox detector. *Computer Vision and Image Understanding*, 189:102827, 2019.

[15] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:779–788, 2016.

[16] Joseph Redmon and Ali Farhadi. YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[17] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. YOLOX: Exceeding YOLO Series in 2021. jul 2021.

[18] R Müller, M Vette-Steinkamp, and A Kanso. Position and orientation calibration of a 2D laser line sensor using closed-form least-squares solution. *IFAC-PapersOnLine*, 52(13):689–694, 2019.

[19] Shashidhar Ram Joshi and Roshan Koju. Study and comparison of edge detection algorithms. In *2012 Third Asian Himalayas International Conference on Internet*, pages 1–5. IEEE, nov 2012.

[20] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[21] Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, jun 2006.