A COMPUTER VISION-BASED APPROACH FOR STORAGE LOCATIONS OCCUPANCY DETECTION USING DEEP LEARNING

[1]Łukasz JELEŃ, [2]Marek KARKULA, [3]Dariusz OLEARCZUK

[1]*Wroclaw University of Technology, Department of Computer Engineering, Wroclaw, Poland, EU,*
lukasz.jelen@pwr.edu.pl

[2]*AGH University of Science and Technology, Faculty of Management, Kraków, Poland, EU,*
mkarkula@agh.edu.pl

[3]*Optidata Sp. z o.o., Kraków, Poland, EU, dariusz.olearczuk@optidata.pl*

Abstract

Increasing the efficiency of processes in warehouse facilities is now required in every industry. One of the important decision-making problems is the proper utilization of storage space. The paper presents research results on the application of architecture for storage location occupancy detection based on computer vision methods and deep learning models. The paper contains a detailed description of the developed solution and an estimation of the solution performance.

Keywords: Occupancy management, storage, computer vision, deep learning, image processing

1. INTRODUCTION

More than two years ago, global supply markets and the way the companies operate has changed significantly. Previously durable and reliable supply chains were made to be broken, practically overnight. Many companies were forced to redefine how they sourced and executed their logistics processes. A significant increase in demand for e-commerce services has also been noticed. Therefore, digitisation and automation of logistics processes has become a necessity. This trend also applies to warehouse processes, moreover, the complexity and speed requirements of warehouse operations are constantly increasing. The use of computer vision methods can help increase automation and improve warehouse processes. For object detection and recognition problems, one of the research areas intensively used currently is computer vision (CV). Computer vision is a rapidly growing field of science that falls under machine learning (ML) and artificial intelligence (AI). It involves algorithms that make decisions based on observed data and visual features of objects. Computer vision plays an increasingly important role in improving the tracking and efficiency of logistics processes. Important application areas today are warehouse systems and operations within these systems.

1.1. Research motivation and objectives

The growing popularity of machine learning algorithms and their utilization in different fields of human life allowed us to take this opportunity to investigate the applicability of such methods in two aspects of storage location analysis. It often happens that we would like to estimate if the cargo fills the load location completely. Furthermore, we also would like to know if there are any unwanted objects on the site. This motivation led us to assess the computer vision methods. In the first and main approach, we evaluated location occupancy with an application of conventional image processing techniques. Our second task involved the detection of people within the location area. The subsequent sections describe the methodology of both approaches along with the depiction of the database and results.

1.2. Related works

Noteworthy are the growing number of papers related to the automation and digitalization of warehouse processes. Warehouse object detection algorithm based on fusion of DenseNet and Single Shot

Detection (SSD) was presented by Chen et al. [4], while Patel and Chowdhury [11] have demonstrated the use of deep neural networks to classify mixed palletizing operations. In his study [16], Xianhui proposed an intelligent logistics inspection system based on big data analysis. The author developed a deep learning-based method for parcel detection and tracking in warehouse facility. Data from the images are automatically processed by the deep neural network model and uploaded to the warehouse task scheduling subsystem. Vukicevic et al. in [15] showed a proposed solution for smart warehouse 4.0. They developed a system for reliable detection of cargo units (pallets) based on QR code recognition using IP cameras and computer vision algorithms.

2. PROBLEM FORMULATION AND METHODOLOGY

Literature review shows that the usage of artificial intelligence methods in warehouse facilities can significantly improve and optimize their operation [5]. In this paper, we took an opportunity to analyze the computer vision approach to estimate location occupancy. Additionally, we carried out experiments that allowed us to detect unwanted objects in the location field of view. Computer vision is a part of artificial intelligence that processes the visual information providing a user with the information about the image content. It became a very popular method in visual quality control of the different processes. Based on provided estimation, we can build a decision support system that can estimate load occupancy as a percentage of total load size.

Today, we can divide the CV field into conventional and deep learning approaches. The conventional approach will use the image information and process it with known image processing methods and extract features that are used in a classification framework to make a decision. With the introduction of deep learning approaches, we process the visual information only with the neural network with a specified architecture. Here, we investigate the ability of the application of both methodologies. We use conventional image processing algorithms such as morphological operations for load occupancy estimation and YoLo v.3 deep neural network for people and object detection within the site.

2.1. Database

For the purpose of this study, we have collected surveillance videos from different cameras and under different conditions (see **Figure 1**). These videos were recorded with 1280×720 and 1280×1080 (px) resolution using Hikvision and Axis CCTV cameras. In the described research, we collected 30 movies of a variable length. From these videos, we sampled 20 frames per second, what provided enough samples for our analysis.

Figure 1 Experimental setup example

2.2. Load occupancy estimation

Recently, numerous studies report the importance of automated visual inspection in various aspects of industry [10]. The main purpose of visual inspection is to calculate object parameters such as their size, shape, and orientation. To properly estimate the size of a load occupancy within the cargo area, we proposed an image processing framework depicted in **Figure 2**. In the presented approach, we use morphological and contextless operations to determine load presence and calculate its size. Such a scheme requires a reference frame on which a cargo area needs to be specified. Load space is defined as a rectangular region of interest

(ROI). **Figure 3** shows an example of ROI selection. This procedure is made manually by a user on a reference frame. Further processing is possible within the region according to the above-mentioned framework. When ROI is selected, the image is preprocessed with a median filter and erosion morphological operation. Before the application of the median filter, the image is converted to grayscale and then before erosion, it needs to be segmented. Segmentation is the task that allows for the determination of objects and their separation from the background. Here we used one of the simplest segmentation methods known as thresholding or binarization. In this task, we select a threshold, and then for each pixel, we assign a new value 0 (black) for the background or 1(white) for an object. In this paper, the load is treated as an object and its determination is based on Equation 1.

$$y_{ij} = \begin{cases} 0 & \text{if} \quad f_{ij} \leq T \\ 1 & \text{if} \quad f_{ij} > T \end{cases} \tag{1}$$

where y_{ij} is a pixel value in the binarized image and f_{ij} is a gray-level value in the input image.
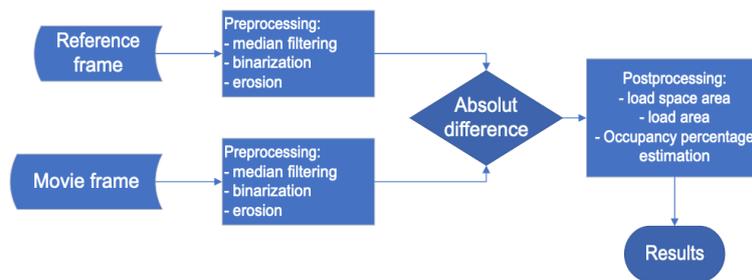
**Figure 2** Diagram of a proposed framework
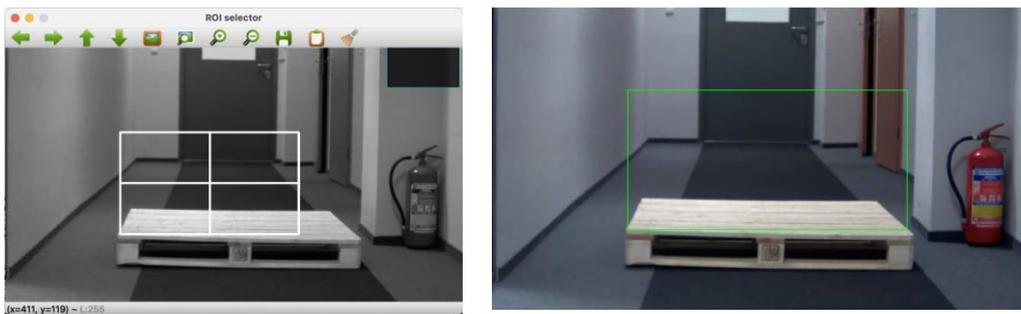
Figure 3 Caro area definition. Region of interest selector (left) and defined area (right)

2.2.1. Preprocessing

During image acquisition, due to technological imperfections, a system introduces various noises to the obtained image. These errors might include illumination conditions, and sensor or lens errors. In image processing, we can model these situations and with the application of filers, one can try to minimize the influence of the noise. In our study, we needed a choice of a filer that can preserve the edges of the objects. For this reason, the median filer was adopted. This is a non-linear filter that operates over a window with a predefined size. Here, we used a 15×15 pixel window within which a median intensity level was selected as a new pixel value. The application of this filter allowed us to remove the speckle noise that is often called a "salt and pepper" noise.

In the next step, after thresholding of the filtered image, we applied morphological operations to remove small unwanted objects. The idea of morphological operations is based on Minkowski's set theory and requires a definition of an additional component called a structuring element (SE). The structuring element is a shape defined as a matrix. It is used in morphological operations to interact with the object in the image according to

Equation 2. Here, we defined a disk-shaped SE of a 5×5 size, which means that we were able to remove all unwanted objects that are smaller than the structuring element.

$$I \ominus SE = \bigcap_{p \in supp(I)} (\widetilde{SE} + \vec{p}) \tag{2}$$

where I is the binary image, $\widetilde{SE}$ is a reflection of SE and *supp(I)* is a set of object in the image - image support.

When both images (reference and incoming frame) are preprocessed we performed a simple subtraction of the two images with normalization. The normalization is performed by clipping all the pixel values that fall below 0 during subtraction.

2.2.2. Postprocessing

After preprocessing and difference determination our framework is able to determine if the cargo area changed. Change in the scene reveals the information that the load has been placed in the location. In the postprocessing part of our methodology, we calculate the area of the load as well as its bounding box. The area of the load is determined as a number of white pixels in the ROI and is used as an estimation of the visible size of the cargo. The bounding box is calculated for all objects in the region and is utilized for the evaluation of load occupancy. The reason for such assessment is that the space covered by just a small part of the cargo cannot be used for additional purposes. The overall occupancy percentage is calculated as a ration between the bounding box area and the cargo area, both expressed in pixels.

2.3.   People detection

The second task of the described research is dedicated to the detection of people at the loading site. It is an important issue and from the management point of view a very crucial one. Conventional approaches described earlier could also be used for this task, but they require a step-by-step framework that will use collected data and apply algorithms to extract features. These features will then be used as an input to a classification system. To be able to correctly extract features, one has to know the problem domain well. As a remedy to that problem, in 1990 LeCun et al. described a neural network that was able to make correct decisions without a conventional feature extraction approach [7]. From the literature review, we can notice that the above-mentioned approaches are applied in a variety of applications from industrial [8] data analysis to medical data [2] and image classification [3, 14] and segmentation [9]. In this study, we investigated the problem of people and object detection in the videos from our database. The deep learning approaches are very well suited for this task [12]. In 2016, Redmon et al. proposed a new approach to object detection called "You Only Look Once" (YOLO) [13]. This approach proved to be useful in both object and people detection [1, 12]. The main advantage of this method is fairly straightforward teaching of new patterns but also a large repository of predefined and pretrained models. In our study, we made use of one of the YOLO networks - YOLO v3. These methods use full images for training and they do not require feature extraction and therefore making them more universal to use. Nevertheless, as reported by researchers, deep learning algorithms typically outperform the conventional machine learning methods, but to make correct decisions these approaches require a very large set of training examples and usually a complicated and long learning curve.

3.   EXPERIMENTAL RESULTS AND ANALYSIS

For the purpose of this study, we implemented a simple image processing framework in Python v3.9 and the OpenCV library. For the object and people detection, an *imageAI* library was incorporated into the project. During our studies, we analyzed the videos from the collected database and calculated load parameters. Additionally, we run experiments with the YOLOv3 neural network to check its performance in our laboratory conditions. The detection results were intentionally recorded in a different environment to somehow mimic the unpredicted warehouse conditions.

Figure 4 (left) shows the result of the bounding box detection and **Figure 4** (right) shows the estimated load area and its occupancy. From these results, we can notice that the bounding box (in red) is correctly calculated. The imperfection on the left image is due to the shadow that the load is producing on the pallet. Further image processing didn't improve the situation and introduced more background noise in the resultant image. Nevertheless, such a small inadequacy didn't have much impact on the occupancy estimation. In addition to the bounding box determination, we show the estimation of the load occupancy percentage (right image) and the load area. It is important to notice that there is a part of the cargo outside of the ROI and in such a case we only calculated the area that was inside the cargo area (in green).

Figure 4 Results of load occupancy estimation

The next experiment involved the detection of people within the cargo area and in a different scenario. We wanted to take that opportunity to verify the correctness of the chosen model. In **Figure 5**, we show the results of people detection in the cargo area. It is easy to notice that the YOLOv3 architecture is able to properly detect a person. From **Figure 5** (left) we can see that the person was detected even when only their legs were visible. What is important to note is the high classification accuracy of 96.52 % for this particular case. When a whole body was visible (see **Figure 5** - right) the accuracy raised to 99.39 %. For the remaining videos, this tendency was maintained.

Figure 5 Results of people detection in the cargo area

Figure 6 Object detection results

In **Figure 6**, we showed results of object detection but in a different scenario than in the previous experiment. Here we can see that other objects are also detected. In this case, we can see a laptop computer that was

also correctly classified. In addition to the laptop, the described architecture was able to detect handbags, cups, chairs, and a few other objects. In this scenario, a person was detected with an accuracy of 99.72 %.

A situation that drew our attention is the significant drop in classification performance when a person exited the viewing field. The highest classification for object detection was recorded at 84.5 % for the laptop (**Figure 6** - left) when the person was standing next to the computer. Contrary, when the person exited the frame, laptop classification dropped to 33.76 % (**Figure 6** - right).

4.   CONCLUSIONS

In this paper, we have taken an opportunity to evaluate computer vision approaches to the task of automated visual inspection. From the results presented in the previous section, we can conclude that image processing techniques provide sufficient information about the storage location occupancy. This allows for further problem investigation in numerous areas. One of the future extensions could be an introduction of an algorithm for automatic cargo area determination as well as an algorithm for the detection of a load exceeding the defined cargo area. The second conclusion that can be drawn in this study is that the application of deep neural networks provided a promising classification result as high as 99.72 %. In future work, a deep neural network scheme could be extended with additional samples. In such a case, we could build a classification system for load identification.

We believe that an introduction of a similar framework can simply warehouse operations and will additionally allow for object traceability and tracking.

REFERENCES

[1]   AHMAD, M., AHMED, I., ADNAN, A. Overhead View Person Detection Using YOLO. *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. 2019, pp. 0627-0633, Available from: https://doi.org/10.1109/UEMCON47517.2019.8992980

[2]   BANACHOWICZ, A., LIS-NAWARA, A., JELEŃ, M., JELEŃ, Ł. Convolutional Neural Networks for Dot Counting in Fluorescence in Situ Hybridization Imaging. In: Zamojski W., Mazurkiewicz J., Sugier J., Walkowiak T., Kacprzyk J. (eds.) *Theory and Applications of Dependable Computer Systems*. DepCoS-RELCOMEX 2020. Advances in Intelligent Systems and Computing. 2020, vol. 1173

[3]   CHAWLA, D., PUSHPALATHA, M., POORNIMA, S., SAXENA, P. Integrated Breast Cancer Analyzer and Predictor Using Machine Learning and Deep Learning. *Annals of the Romanian Society for Cell Biology*. 2021, vol. 25, no. 6, pp. 4460-4478.

[4]   CHEN, L., WANG, F., WANG, L. Research on warehouse object detection algorithm based on fused DenseNet and SSD. In: *Chinese Conference on Image and Graphics Technologies*. Berlin: Springer, 2019, pp. 602-611. Available from: https://doi.org/10.1007/978-981-13-9917-6_57.

[5]   DEVEREUX, M.G., MURRAY, P., WEST, W. Automated Object Detection for Visual Inspection of Nuclear Reactor Cores. *Nuclear Technology*. 2022, vol. 208, no. 1, pp. 115-128. Available from: https://doi.org/10.1080/00295450.2020.1863067.

[6]   KHAN, S.A., HASSAN, S.M., KUSI-SARPONG, S., MUBARIK, M.S., FATIMA, S. Designing an integrated decision support system to link supply chain processes performance with time to market. *International Journal of Management Science and Engineering Management*. 2022., vol. 17, no. 1, pp. 66-78. Available from: https://doi.org/10.1080/17509653.2021.2000900.

[7]   LECUN, Y., BOSER, B., DENKER, J., HENDERSON, D., HOWARD, R., HUBBARD, W., JACKEL, L. Handwritten Digit Recognition with a Back-Propagation Network. In *Advances in Neural Information Processing Systems*. Morgan-Kaufmann. 1990, pp. 396-404.

[8]   LUCKOW, A. et al. Artificial Intelligence and Deep Learning Applications for Automotive Manufacturing. In: *2018 IEEE International Conference on Big Data (Big Data)*. 2018, pp. 3144-3152. Available from: https://doi.org/10.1109/BigData.2018.8622357.

[9]   MISELIS, B., KULUS, M., JUREK, T., RUSIECKI, A., JELEŃ, Ł. Deep Neural Network for Whole Slide Vein Segmentation. In: Saeed, K., Homenda, W. (eds.) *Computer Information Systems and Industrial Management. CISIM 2018. Lecture Notes in Computer Science*. 2018, vol 11127.

[10]   MORDIA, R, VERMA, A. K. Visual techniques for defects detection in steel products: A comparative study. *Engineering Failure Analysis*. 2022, vol. 134, p. 106047. Available from: https://doi.org/10.1016/j.engfailanal.2022.106047.

[11]   PATEL, A.D., CHOWDHURY, A.R. (2022). Vision-Based Object Classification Using Deep Learning for Mixed Palletizing Operation in an Automated Warehouse Environment. In: Dave, H.K., Dixit, U.S., Nedelcu, D. (eds.) *Recent Advances in Manufacturing Processes and Systems. Lecture Notes in Mechanical Engineering*. Springer, Singapore. Available from: https://doi.org/10.1007/978-981-16-7787-8_79

[12]   RANST, W. V., DE SMEDT, F., BERTE, J., GOEDEMÉ, T. Fast Simultaneous People Detection and Re-identification in a Single Shot Network. In: *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 2018, pp. 1-6. Available from: https://doi.org/10.1109/AVSS.2018.8639489.

[13]   REDMON, J., DIVVALA, S., GIRSHICK, R.: You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 779-788.

[14]   RYBIAŁEK, A., JELEŃ, Ł. Application of DenseNets for Classification of Breast Cancer Mammograms. In: Saeed, K., Dvorský, J. (eds.) *Computer Information Systems and Industrial Management. CISIM 2020. Lecture Notes in Computer Science*. 2020, vol 12133.

[15]   VUKICEVIC, A., MLADINEO, M., BANDUKA, N., MACUZIC, I. A smart Warehouse 4.0 approach for the pallet management using machine vision and Internet of Things (IoT): A real industrial case study. *Advances in Production Engineering & Management*. 2021, vol. 16, No. 3, pp. 297-306. Available from: https://doi.org/10.14743/apem2021.3.401

[16]   XIANHUI, Ch. E-Commerce Logistics Inspection System Based on Artificial Intelligence Technology in the Context of Big Data. *Security and Communication Networks*. 2022, vol. 2022, 3418466. Available from: https://doi.org/10.1155/2022/3418466.