DOI: 10.2195/lj_proc_azizpour_en_202211_01
URN: urn:nbn:de:0009-14-55795

Synthetic Data Generation for Robotic Order Picking

Synthetische Datengenerierung für die Kommissionierung mit Robotern

Moein Azizpour
Nafiseh Namazypour
Alice Kirchheim

Department of Technology of Logistics Systems
Helmut Schmidt University
University of the Federal Armed Forces Hamburg

Abstract: Advances in robotics, especially in computer vision, have led to the increasing use of robots in order picking. Deep Learning methods using CNN for computer vision purposes have shown good object detection and localization results. However, training neural networks requires a large amount of domain-specific labelled data. In this work, we generated synthetic data and converted it to the appropriate format to be fed to neural network. For this purpose, randomized camera angles, backgrounds, and object configuration are used for data augmentation. A generalized and balanced dataset is ensured by varying these parameters based on the properties of natural objects.

[Keywords: logistics, computer vision, synthetic data generation, order picking, pick and place]

Kurzbeschreibung: Fortschritte in der Robotik, insbesondere in der Computer Vision, haben zu einem zunehmenden Einsatz von Robotern in der Kommissionierung geführt. Deep-Learning-Methoden, die CNN für Computer-Vision-Zwecke verwenden, haben gute Ergebnisse bei der Objekterkennung und -lokalisierung gezeigt. Das Trainieren neuronaler Netze erfordert jedoch eine große Menge an objektspezifisch markierten Daten. In diesem Beitrag haben wir synthetische Daten generiert und in ein geeignetes Format konvertiert, um damit neuronale netzte zu trainieren. Zu diesem Zweck werden randomisierte Kamerawinkel, Hintergründe und Objektkonfigurationen zur Datenerweiterung verwendet. Durch die Variation dieser Parameter auf der Grundlage der Eigenschaften natürlicher Objekte wird ein allgemeiner und ausgewogener Datensatz gewährleistet.

[Schlüsselwörter: Logistik, Computer Vision, Erzeugung synthetischer Daten, Kommissionierung, Aufsammeln und plazieren]

1 Introduction

The constantly growing area of E-Commerce demands the need for fast and cost-effective customer order processing and order picking solutions. In logistics, considering a rapid change in workforce demography and high cost-per-pick, the use of Robotic-Order-Picking (ROP) is accelerating. However, fully automatic industrial applicable ROP in logistics still does not exist and they have limited flexibility in the shape and category of the to-be-picked objects. To overcome this inflexibility, newer ROP approaches build upon the advancements in Computer Vision and, more specifically, on utilizing Deep Learning and convolutional neural networks (CNN). A reliable vision system is required for a successful pick and place application in a pick station.

Deep learning has expanded to meet the needs of intelligent vision systems, but it requires a large amount of domain-specific labelled data. Despite all the achievements of this state-ot-the-art technique, providing large and high-quality data is an indispensable part of any deep-learning-based approach. The upside is that the amount of data is increasing exponentially every day. According to Forbes Magazine [1], 2.5 quintillion bytes of data are created each day at our current pace. However, in specific domains like computer vision, we often suffer from the scarcity of high-quality labelled data. This can be due to various reasons, including having limited access to the classified data or the quality of the available data does not meet the expectations.

Generating synthetic data is a promising alternative to overcome such deficiencies. State-of-the-art methods for synthetic data generation create labelled data that reflects features of the underlying real-world data like texture, scale, and shape. It is an active research field with the aim of replacing time-consuming, susceptible to human error and expensive manual data collection. The lack of quality data is also noticeable in logistics and order picking

DOI: 10.2195/lj_proc_azizpour_en_202211_01
URN: urn:nbn:de:0009-14-55795

applications [2], although some researchers like C. Mayershofer, collected a relatively sizable amount of logistics objects and annotated them for public use [2]. The objective of this paper is the generation of synthetic data for grocery order picking in logistics and investigation of the varying parameters involved.

In this paper, first, the previous works are reviewed. Then the changes in synthetic data generation in recent years and the future of this industry are described in Data Generation section, the whole process of synthetic data generation from 3D models and software tools to annotation and data preparation is disclosed.

2 Literature Review

Previous research has been surveyed based on the different methods of generating synthetic datasets. The two techniques include deep-learning-based and 3D-model-based. These two methods are selected as they are associated with image data. Drawing numbers from a distribution is another technique but it is only applicable in synthetic numeric data. Variational autoencoder and generative adversarial network (GAN) models are deep-learning-based techniques that improve data utility by feeding models with more data and in 3D-model-based, a 3D model of the objects is used with different viewpoints and background to generate more data. Recently, generation and use of synthetically generated data has gained popularity among researchers and has been used in several applications. Hinterstoisser et al. used a set of 64 objects to create a synthetic dataset with 3D models and varying backgrounds and item positions, ensuring a unified distribution of foreground objects in each frame [3]. They used 2D labels and the Faster-RCNN algorithm for the base of comparison. Weichao Qiu investigated the model robustness with synthetically generated data and addressed the domain gap between real and synthetic data [4]. He also presented two domain adaptation algorithms with and without an existing CAD model of the object. In the survey of a paper written by Nikolenko a more detailed investigation was conducted about the application of synthetic data and the method of generation with various methods like GANs as a deep-learning-based method and improving CGI-based (Computer Aided Imaginary) generation [5]. The latter can be categorized as 3D-based method and requires following the conventional workflow of creating 3D models of synthetic data in computer vision by setting up the environment with a camera, placing the objects in a controlled scene, and rendering synthetic images. Schoepflin et al. focused on creating a pipeline for generating synthetic data in intralogistics for visual object identification on load carriers [6]. For this purpose, they defined several real-life scenarios in handling load containers and used 3D scanner for capturing 3D data in a similar pipeline described in previous research by Nikolenko [5].

However, to the best of the authors' knowledge, no research has been conducted on generating synthetic data of fruits and vegetables in food logistics. Furthermore, the effect of domain randomization on neural network performance was not investigated.

3 History and Future of Synthetic Data Generation

Data is an integral part of machine learning and intelligent systems. In the context of privacy-preserving statistical analysis, in 1993, the idea of original fully synthetic data was created by Donald Rubin [7]. After that, by advances in technology, the role of synthetic data became more important and it finally led to today in which the portion of synthetic data is almost the same as the real data [8]. The prediction of synthetic data share, up to 2020, is shown in Figure 1. This figure shows the whole trend of data used for AI from real or synthetic point of view from 2020 to 2030. It is not restricted to images, which are in the focus of this paper, and does not make a differentiation with respect to properties of underlying data. Today, most synthetically employed data have their origin in easy to generate data as distribution of figures with known underlying property. But it is expected that not only the utilization of data for AI will face a sharp increase in the next years, but also synthetic data will completely overshadow real data in AI models [8].

It is estimated that by 2024, 60 percent of the data used to develop AI and analytics projects will be synthetically generated and the synthetic data market will continue to grow more than 10 % p.a. since most of the synthetic data market serves:

1. Test data management which is expected to grow 11.6 % compound annual growth rate (CAGR)
2. AI training data generation, which is expected to grow at 22.2 % compound annual growth rate (CAGR)

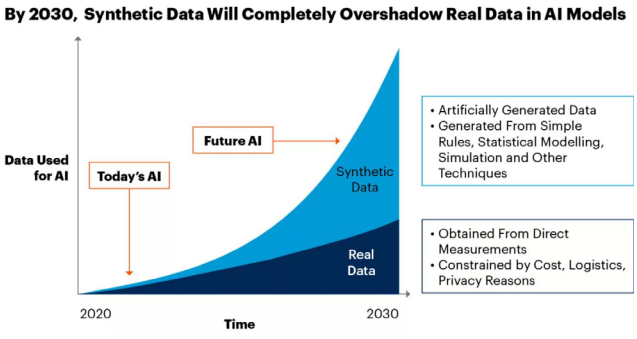

Figure 1. Prediction the future of synthetic data[8]

DOI: 10.2195/lj_proc_azizpour_en_202211_01
URN: urn:nbn:de:0009-14-55795

Synthetic data is often created with the help of algorithms and is used for many activities, including test data for new products and tools, model validation, and training AI models.

Figure2 demonstrates how manually annotated datasets differ from synthetically generated ones in terms of workflows and iteration. By eliminating the need to review and audit datasets and avoiding time-consuming data collection and clean-up steps, synthetic data open the door to more rapid iteration [9].

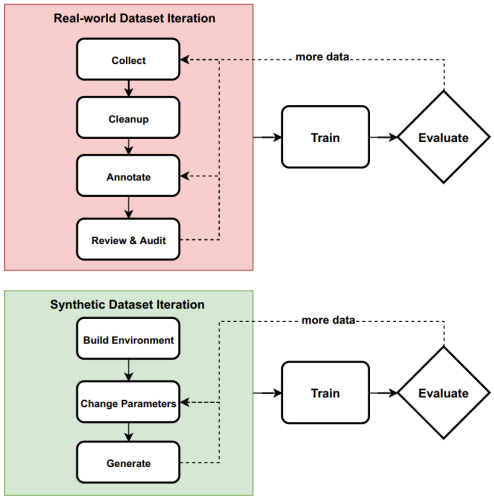

Figure 2. Comparison between real data and synthetic data generation processes.

The most important advantages of synthetic data generation are as follows:

- Accurate ground truth [4]: Manual labelling is prone to human errors and is an exhaustive task. Automatic labelling of the generated data set is conducted with high accuracy; thus, it needs less data cleaning afterwards.
- Controllability [10]: The controllability enables two things: 1) understand the model through controllable data, 2) train the model through interaction
- Overcoming real data usage restrictions [8]: Real data may have usage constraints due to privacy or other regulations. Synthetic data can replicate all essential statistical properties of real data without exposing real data, thereby eliminating the issue.
- Creating data to simulate not yet encountered conditions [8]: It means where real data does not exist, synthetic data is the only solution.
- Immunity to common statistical problems: These can include item nonresponse, skip patterns, and other logical constraints [8].

Though synthetic data has a number of advantages that can help, it also has some challenges. The most important drawbacks are [8]:

- The quality of the model depends on the data source: The quality of synthetic data is highly correlated with the quality of the input data and the data generation model. Synthetic data may reflect the biases in source models.
- User acceptance is more challenging: Synthetic data is an emerging concept and it may not be accepted as valid by users who have not witnessed its benefits before.
- Output control is necessary: Especially in complex datasets, the best way to ensure the output is accurate is by comparing synthetic data with authentic data or human-annotated data. this is because there could be inconsistencies in synthetic data when trying to replicate complexities within original datasets.

3.1 Image Segmentation

Image segmentation aims to recognize and understand what is in the image at the pixel level. Every pixel in an image belongs to at least one class, as opposed to object detection, where the bounding boxes of objects can overlap [11]. Different types of image segmentation are bounding boxes, 3D cuboids, semantic segmentation, polygon segmentation, landmark segmentation, line segmentation and instance segmentation. In this paper, instance segmentation of the images within the dataset is needed. The bounding boxes are also used to demonstrate the results.

3.2 Image Annotation

There is no single standard format when it comes to image annotation. In this paper, the Common Objects in Context (COCO) format is utilized, as the COCO format is one of the most popular formats in object detection. It is a dataset of complex everyday scenes containing common objects in their natural context. It has 91 objects types with a total of 2.5 million labeled instances in 328k images [12]. In most of the intelligent object detection works, this format is used. So, it would be easy to feed the generated data to every network and compare the results.

The annotations are stored using JSON. For object detection, COCO follows the format shown in Figure 3 [12].

The “annotations” section includes a list of every individual object annotation from every image in the dataset [13]. Each annotation has an id (unique to all other annotations in the dataset). The image id corresponds to a specific image in the dataset. The category id corresponds to a single category specified in the categories section. Area is measured in pixels (e.g. a 10 px by 20 px box would have

DOI: 10.2195/lj_proc_azizpour_en_202211_01
URN: urn:nbn:de:0009-14-55795

an area of 200). Is Crowd specifies whether the segmentation is for a single object or for a group/cluster of objectsThe COCO bounding box format is [top left x position, top left y position, width, height].

The "categories" object contains a list of categories (e.g. apple, orange) and each of those belongs to a supercategory (e.g. fruits, vegetables) and an id which is an integer number assigned to the category.

```
annotation{
"id" : int,
"image_id": int,
"category_id": int,
"segmentation": RLE or [polygon],
"area": float,
"bbox": [x,y,width,height],
"iscrowd": 0 or 1,
}
categories[{
"id": int,
"name": str,
"supercategory": str,
}]
```

Figure 3. COCO annotation format

3.3 Dataset Generation

Unity is a powerful game engine to help generate synthetic datasets. It supports domain randomization for introducing variety into the generated datasets, 2D/3D object detection, semantic segmentation, instance segmentation, and keypoints (nodes and edges attached to 3D objects) [9].

This work uses Unity Perception package to generate and augment the synthetic dataset. This package has an advantage over its competitors like NVIDIA Isaac Sim and BlenderProc in terms of degree of randomization and having a highly customizable toolset [9].

For validating the quality of synthetically generated data, training is being conducted with Mask R-CNN network [14]. Mask R-CNN is chosen because of its capability to detect objects and simultaneously generate a segmentation mask for each object.

Mask R-CNN uses the backbone of ResNet 101 and extended R-CNN and faster R-CNN in predicting object masks [15] and achieves higher accuracy in a shorter time [16].

To create a dataset, the twelve most popular grocery items are chosen based on Statista [17] report "Per capita consumption of fruit in Germany by type by 2020/21''. These are red apple, yellow apple, green apple, banana, strawberry, tomato, potato, carrot, cucumber, onion, bell pepper, and three branches of these items as a banana branch, cucumber branch, and carrot branch.

Then, 3D models of each item were used to generate scenes and capture images. These models of fruits and vegetables are converted via Blender for the readable format in Unity. To implement Domain Randomization [18] a set of Randomizers is used, each undertaking a specific randomization task.

- Background: 50 different backgrounds are selected from random pictures with various textures and content.
- Number of objects: The type and number of objects within each frame are randomized.
- Object Placement and positioning
- Scaling: The distance of the objects to the camera and thus their sizes are varied within each capture.

In Total, 1000 images are captured with their annotations in JSON format. Figure 4 shows a sample including banana, orange, cucumber, tomato and banana branch.

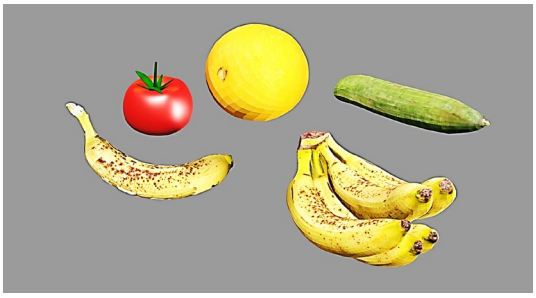

Figure 4. A sample of 3D-model configuration of items

For each capture, each item's bounding box and mask are generated and saved into a JSON file. Figure 5 demonstrates a sample of output in which the bounding boxes and masks are drawn.

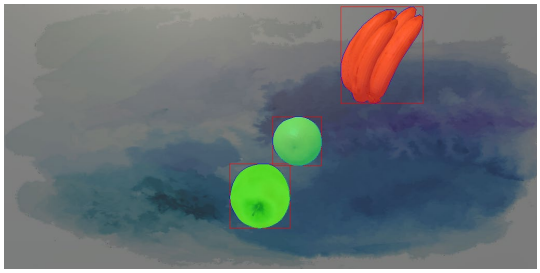

Figure 5. A sample of output which has both mask and bounding box

Statistics from the synthetic data generation are shown in Figure 6.

DOI: 10.2195/lj_proc_azizpour_en_202211_01
URN: urn:nbn:de:0009-14-55795

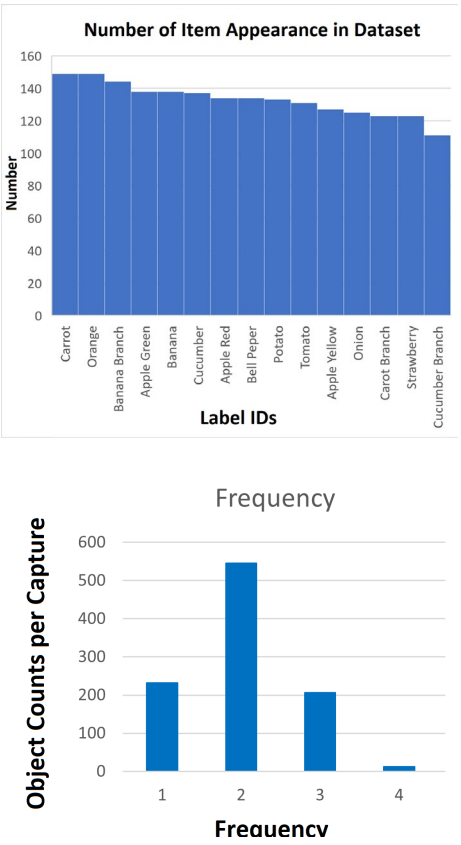

Figure 6. Statistics from the synthetic data generation: (a) Object counts per capture

(b) Number of item appearance in dataset

After generating all images and their annotations, they should be fed to a Mask R-CNN network. This required all the annotations to be reconfigured from Unity JSON output to COCO-JSON. The annotations are converted to COCO format, a specific JSON structure dictating how labels and metadata are saved for image data.

A script is written to read the JSON file, extract data of all instances of each image and then write those data in COCO-JSON format to use as an input for Mask R-CNN. The pipeline from 3D-model to object detection is shown in Figure 7.

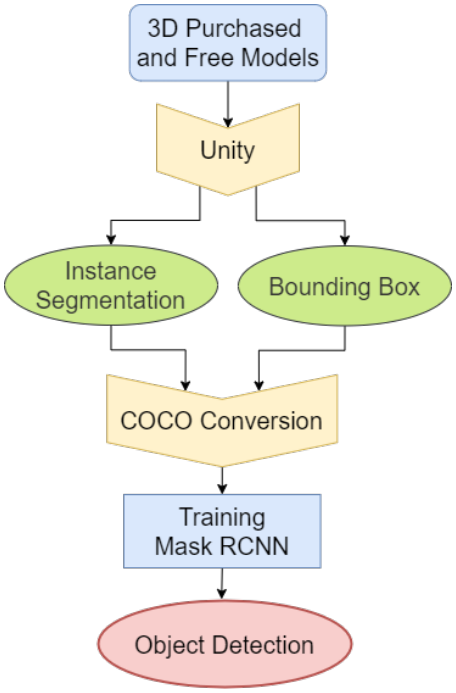

Figure 7. The whole process from 3D-model to object detection

Figure 8 shows the Autonomous Mobile Robot (AMR) equipped with a robot arm. This combination enables the navigation from picking the items from the high bay warehouse and bringing it to the pick station with the help of one integrated hardware solution. Figure 9 shows the box containing fruits as it is delivered to pick station.

DOI: 10.2195/lj_proc_azizpour_en_202211_01
URN: urn:nbn:de:0009-14-55795

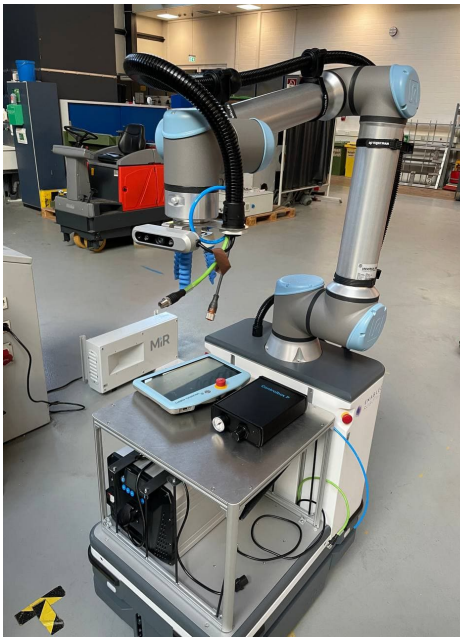

Figure 8. The robotic Arm and Softgripper system

Figure 9. A box containing fruits for pick and place

4 Conclusion and Future Works

In this work, we presented an initial approach to generating synthetic data. We aim to gain deeper insight into the statistical distribution of objects over the total dataset and its effect on the neural network's performance for a 3D-Object detection purpose. We provided 3D models and then set some domain randomization parameters to create a balanced dataset. The effect of including real data in the synthetic data will also be investigated in future research. Furthermore, the generated dataset and the trained neural network will be implemented in the robotic arm to pick and place grocery objects. It can also be a good practical benchmark for the whole process.

1 Literatur

[1] B. Marr, „How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read“, *Forbes*, 21. Mai 2018, 2018. [Online]. Verfügbar unter: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/?sh=80df1bb60ba9. Zugriff am: 16. August 2022.

[2] C. Mayershofer, D.-M. Holm, B. Molter und J. Fottner, „LOCO: Logistics Objects in Context“ in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA, 2020, S. 612–617, doi: 10.1109/ICMLA51294.2020.00102.

[3] S. Hinterstoisser, O. Pauly, H. Heibel, M. Marek und M. Bokeloh, „An Annotation Saved is an Annotation Earned: Using Fully Synthetic Training for Object Instance Detection“.

[4] Weichao Qiu, „Generating Human Images and Ground Truth Using Computer Graphics“. Master of Science, Uniersity of California, Los Angeles, 2016.

[5] S. I. Nikolenko, „Synthetic Data for Deep Learning“, 9/25/2019. [Online]. Verfügbar unter: http://arxiv.org/pdf/1909.11512v1.

[6] D. Schoepflin, D. Holst, M. Gomse und T. Schüppstuhl, „Synthetic Training Data Generation for Visual Object Identification on Load Carriers“, *Procedia CIRP*, Jg. 104, S. 1257–1262, 2021, doi: 10.1016/j.procir.2021.11.211.

[7] Wikipedia, *Synthetic data.* [Online]. Verfügbar unter: https://en.wikipedia.org/w/index.php?title=Synthetic_data&oldid=1071178758 (Zugriff am: 15. August 2022).

[8] C. Dilmegani, „What is Synthetic Data? What are its Use Cases & Benefits?“, *AIMultiple*, 19. Juli 2018, 2018. [Online]. Verfügbar unter: https://research.aimultiple.com/synthetic-data/. Zugriff am: 14. August 2022.

[9] S. Borkman *et al.,* „Unity Perception: Generate Synthetic Data for Computer Vision“, 7. Sep. 2021. [Online]. Verfügbar unter: https://arxiv.org/pdf/2107.04259.

[10] Weichao Qiu, „Solving Computer Vision Challenges with Synthetic Data“. for the degree of Doctor of Philosophy, Johns Hopkins University, Baltimore, Maryland, 2020.

[11] Shaip, *Quality Image Annotation Services | Image Labeling for Computer Vision.* [Online]. Verfügbar unter: https://www.shaip.com/offerings/image-annotation/ (Zugriff am: 14. August 2022).

[12] T.-Y. Lin *et al.,* „Microsoft COCO: Common Objects in Context“, 1. Mai 2014. [Online]. Verfügbar unter: http://arxiv.org/pdf/1405.0312v3.

[13] „Create COCO Annotations From Scratch“, *Immersive Limit*, 1. Nov. 2019, 2019. [Online]. Verfügbar unter: https://www.immersivelimit.com/tutorials/create-coco-annotations-from-scratch. Zugriff am: 16. August 2022.

[14] Kaiming He, Georgia Gkioxari, Piotr Dollar und Ross Girshick, „Mask R-CNN“, 2018.

DOI: 10.2195/lj_proc_azizpour_en_202211_01
URN: urn:nbn:de:0009-14-55795

[15] R. Katırcı, E. K. Yılmaz, O. Kaynar und M. Zontul, „Automated evaluation of Cr-III coated parts using Mask RCNN and ML methods“, *Surface and Coatings Technology*, Jg. 422, S. 127571, 2021, doi: 10.1016/j.surfcoat.2021.127571.
[16] T. Chen, Y. Jiang, W. Jian, L. Qiu, H. Liu und Z. Xiao, „Maintenance Personnel Detection and Analysis Using Mask-RCNN Optimization on Power Grid Monitoring Video“, *Neural Process Lett*, Jg. 51, Nr. 2, S. 1599–1610, 2020, doi: 10.1007/s11063-019-10159-w.
[17] Statista, *Pro-Kopf-Konsum von Obst in Deutschland nach Art 2020/21 | Statista.* [Online]. Verfügbar unter: https://de.statista.com/statistik/daten/studie/247425/umfrage/die-beliebtesten-obstsorten-der-deutschen/ (Zugriff am: 15. August 2022).
[18] S. Z. Valtchev und J. Wu, „Domain randomization for neural network classification“ (eng), *Journal of big data*, Jg. 8, Nr. 1, S. 94, 2021, doi: 10.1186/s40537-021-00455-5.

Moein Azizpour schloss 2020 sein Masterstudium "Nutzfahrzeugtechnik" an der Universität Kaiserslautern ab und setzte anschließend seine akademische Laufbahn als wissenschaftlicher Mitarbeiter an der Helmut-Schmidt-Universität fort. Dort beschäftigt er sich mit der Automatisierung von Pick-and-Place-Robotern in der Lebensmittellogistik.

Address: Helmut-Schmidt-Universität, Universität der Bundeswehr Hamburg, Technologie von Logistiksystemen, Holstenhofweg 85, 22043 Hamburg, Germany, Phone: +49 40 6541-2126, E-Mail: moein.azizpour@hsu-hh.de

Nafiseh Namazypour studiert im Masterstudiengang "Automation and Control" an der Technischen Universität Kaiserslautern und setzte seine akademische Laufbahn als wissenschaftlicher Mitarbeiter an der Helmut-Schmidt-Universität fort. Dort beschäftigt sie sich mit Projekten im Bereich Computer Vision.

Address: Helmut-Schmidt-Universität, Universität der Bundeswehr Hamburg, Technologie von Logistiksystemen, Holstenhofweg 85, 22043 Hamburg, Germany, Phone: +49 40 6541-2126, E-Mail: namazypn@hsu-hh.de

Alice Kirchheim, Prof. Dr.-Ing., Professorin für Technologie von Logistiksystemen an der Helmut-Schmit-Universität in Hamburg. Alice Kirchheim studierte Informatikingenieurwesen an der TU Hamburg und wurde an der Universität Bremen zu einem Thema der Automatisierung logistischer Prozesse promoviert. Nach einer mehrjährigen Tätigkeit in der in der Industrie ist sie seit 2021 Professorin an der Helmut-Schmidt-Universität in Hamburg.

Address: Helmut-Schmidt-Universität, Universität der Bundeswehr Hamburg, Technologie von Logistiksystemen, Holstenhofweg 85, 22043 Hamburg, Germany, Phone: +49 40 6541-2126, E-Mail: alice.kirchheim@hsu-hh.de