HSLU Hochschule Luzern

Technik & Architektur
Master Thesis

MSE Mechatronics and Automation

Stereo Vision-Based Object Detection for Mechanical Parts

using Synthetic Data Generation and Convolutional Neural Networks

Infrastructure

Fig. 1: UR3e and Intel Realsense D415

Fig. 2: Mechanical parts

3D Reconstruction Process

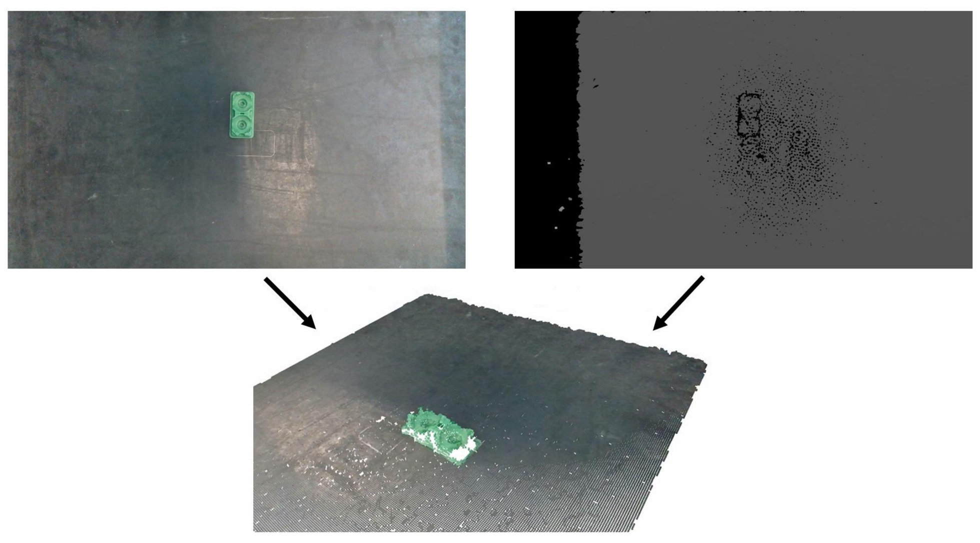

Fig. 3: **Point cloud generation** from depth and color image

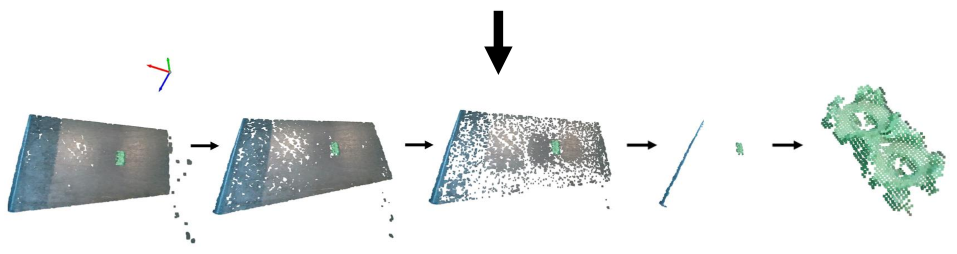

Fig. 4: Point cloud **processing** and **fusion**

Fig. 5: Filtered point cloud

Synthetic Data Generation

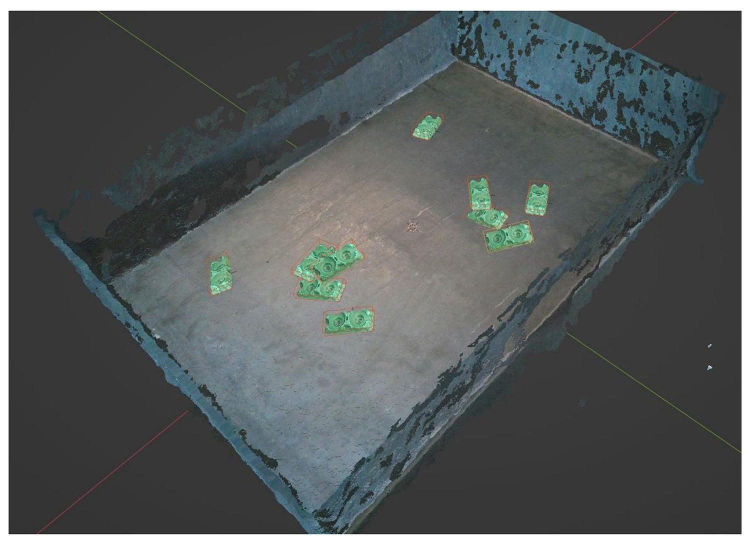

Fig. 6: **Synthetic Scene** with scanned objects and bounding boxes in Blender

Fig. 7: Synthetic **depth & color** image

2D Object Detection

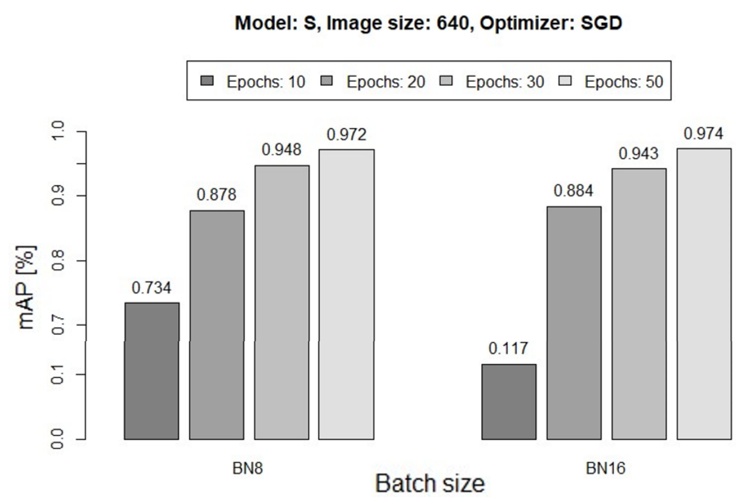

Fig. 8: Evaluation of **hyperparameters**

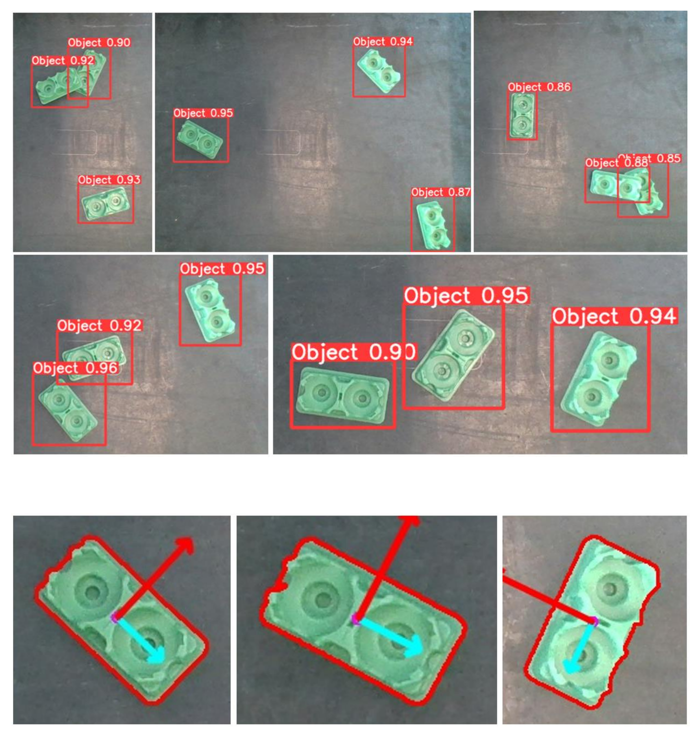

Fig. 9: **Prediction** on **real** images

Introduction

As part of this master research project, a collaborative effort between the Lucerne University of Applied Sciences and Mequadrat AG is progress to develop an advanced system for object detection, localization and subsequent robot-assisted removal of mechanical parts. This solution has the purpose to automate the process of picking objects from bins or containers, using advanced robotics and **3D computer vision** techniques (Fig.1). The primary objective of this thesis is to lay the foundation to design and implement a robust and efficient system that can accurately identify mechanical parts (Fig.2) within a logistic box, determine their precise location and orientation and facilitate their picking from the box through the use of robotic assistance. This process, called **bin picking**, incorporates robotics for motion planning and picking, 3D vision technology (stereo vision) for depth sensing and deep learning for localization of the objects.

Methods

This master thesis incorporates a comprehensive control and parameterization of the **Intel Realsense D415 camera** within a **LabVIEW** environment. This stereovision camera can generate 3D point clouds of scenes (Fig.3). To process and filter point clouds, the LabVIEW software was expanded using a Python-based machine learning environment, facilitating the implemented **3D reconstruction** process. The reconstructed point clouds (Fig.5) are then transformed into realistic synthetic data using Blender, which served as the training dataset for a CNN-based deep learning algorithm. The newly released YOLOv8 algorithm was integrated, enabling **real-time object detection** and **localization**. This project covers the entire workflow, encompassing camera control, 3D reconstruction and object detection, providing a seamless approach for bin picking solutions.

Results

By optimizing the Intel Realsense camera, it was possible to generate accurate 3D point clouds. These were processed to create a representative model (Fig. 5) in the 3D reconstruction process.

The resulting point cloud was incorporated into a realistic environment in Blender, allowing it to generate **synthetic 3D point clouds** (Fig. 6), **rendered 2D color** and **depth images** (Fig.7) and their corresponding **2D and 3D labels** by using virtual cameras. Using this method, it becomes possible to extract all relevant information within a matter of seconds. These synthetic datasets were then utilized to train the **YOLOv8 algorithm**, a CNN-based object detection algorithm. Through extensive investigations, the optimal parameters for the deep learning model were identified (Fig. 8). Finally, the model's performance was evaluated with unseen real-world images, demonstrating exceptional results with high **accuracy** (**> 85%**) in the localization and evaluation of the orientation of the objects (Fig. 9).

Marvin Lichtsteiner

Advisor: Prof. Dr. Björn Jensen
Expert: Thomas Estier

Project partner: Stefan Nyffenegger
CEO mequadrat AG

mequadrat

MSE | MASTER OF SCIENCE IN ENGINEERING

FH Zentralschweiz