# Equipment identification through image recognition

Author: Saidnassimov Darkhan

Master’s thesis Introduction chapter

2022

# 1 Introduction

### 1.1 Problem statement

In recent years, computer vision algorithms have received much attention due to their potential applications in a vast variety of fields. As the technologies advance, endless possibilities arise in numerous fields, including security monitoring, medicine and self-driving vehicles.[2] However, although over the last decade computer vision has emerged in industrial applications (e.g., safety and process monitoring) [2], less research has addressed the issue of equipment detection.

Industrial plants typically are hundreds of meters long and often it becomes frustrating to identify minor equipment parts. Identifying parts becomes relevant once they start requiring replacement or maintenance, as the plant would not be able to run at full capacity without them. Ore processing plants treat several hundreds of tons of ore per hour, and the production capacity is constant. [2]. Therefore, quite often it is troublesome to properly identify the equipment within a list of thousands of various parts in a medium- to large-scaled plant.

Even though various methods have been implemented for detection of objects in various domains [2], these methods heavily rely on extensive data collection and training of models in order to work. Complications arise, as it is often not possible to collect huge amounts of training images from the plants due to privacy and confidentiality issues. Luckily, in this work, the images can rather easily be collected from the 3D simulator. However, this limits the accuracy of the models, as the models will not perform as well on real images.

Hence, this thesis proposes a cross-domain object detection algorithm as a solution to automatically localize and identify the equipment in a large environment in order to minimize the delay before the production goes online again.

This work has been commissioned by Metso Outotec Oyj. Metso Outotec aims to offer applications that would allow their customers to automate their mineral and metal refining processes. In order to help these processes run as smoothly and automatically as possible, it is important to minimize the downtime between maintenance stops. Recently, due to promising results, computer vision has been integrated into various fields of the business. For these reasons, Metso Outotec has requested an application, which would leverage state-of-the-art computer vision methods in equipment recognition.

### 1.2 Thesis objective

The main goal of this thesis is to identify suitable state-of-the-art object detection methods and to adapt these methods in order to develop an algorithm to be implemented in a minimal proof-of-concept application for an industrial plant. The proposed algorithm should be able to identify an object in a real image given a labeled dataset of rendered images from a 3D equipment model and a smaller unlabeled dataset of real images.

Additionally, the developed algorithm should provide a means for optimizing the laborious process of collecting and labeling data. Furthermore, the thesis must consider cases when more objects are appended to the dataset. Such optimization is important not only because it is a time demanding process, but also because large plants contain thousands of objects, thus making scalability a critical requirement.

### 1.3 Methods

In order to accomplish these objectives, the thesis will first explore state-of-the-art object detection frameworks, libraries and methods. Similarly, the domain adaptation methods will be analyzed in an object detection setup. The most suitable methodologies will then be used to train a cross-domain object detection model. However, a modern deep learning model requires a vast amount of data. Most traditional object detection algorithms use consistent environmental conditions to obtain training and testing images. Unfortunately, the resulting accuracy often suffers due to the domain shift phenomenon [2]. The domain shift occurs when the conditions at the time of capturing training and testing images alter, as discussed in [2]. A typical example is a combination of CityScapes and FoggyCityScapes datasets, where images are taken in normal and foggy weather, respectively. Additionally, in an industrial environment, such as plants of the clients, it often becomes challenging to obtain images due to regulations regarding accessibility and confidentiality. For these reasons, the dataset utilized to implement the methods in our sandbox scenario is based on TLess open-source dataset [2]. Since the dataset was originally intended for pose estimation in 3D models, it will be converted into formats appropriate for the proposed object detection methods.

As mentioned earlier, one promising approach to fight domain shift phenomenon and achieve higher performance in object detection model would be through domain adaptation. In the given scenario, a cross-domain object detection model based on Faster-RCNN and Adaptive Teacher implementation will be proposed to tackle the objectives. The model will consider additional components to the adaptation network and evaluate their efficiency. Additionally, the study will consider other transfer learning methodologies, such as fine-tuning and class incremental learning to grant a potential for scalability of the model in production. Finally, the model will be integrated into a prototype web application for demonstration purposes.

These methods will be trained on rendered data from 3D models and evaluated on real images using mean average precision metrics. Finally, the proposed method will be evaluated using one equipment item from a real plant.

### 1.4 Scope

The thesis will be limited to proposing the solution upon analyzing and combining different components of existing state-of-the-art models. Preparing an actual real-life dataset and implementing the solution for a real plant remains out of the scope due to time constraints of the project.

Although the proposed methods attempt to optimize the data collection and labeling process, in practice this will require many months before the dataset, the model based on the real data would be ready for training.

When it comes to the user interface, a prototype will be provided in order to showcase the performance of the model. However, the thesis will be primarily focused on the deep learning algorithms rather than how to deploy a model in production. For this reason, the prototype will only serve basic functionality. Finally, due to time constraints, a video-compatible model will remain out of the scope.

### 1.5 Structure of the thesis

The rest of this thesis is divided into four chapters. Chapter 2 reviews the literature that is relevant to understand the key concepts of the thesis. The chapter covers one- and two-stage state-of-the-art object detection models, as well as introduces the reader to domain adaptation, a subfield of transfer learning. Consequently, the chapter reviews latest cross-domain object detection techniques. Finally, the chapter discusses up-to-date class incremental learning implementations. Chapter 3 defines the dataset used and outlines the proposed architecture of the model. Chapter 4 evaluates the solution and compares the results to other methods using average precision metrics. Chapter 5 summarizes this work by discussing the proposed architecture as well as by suggesting directions for future work.

### References

1. Wanyi Li, Fuyu Li, Yongkang Luo, Peng Wang, and Jia sun. Deep domain adaptive object detection: a survey. February 2020.
2. Poojan Oza, Vishwanath A. Sindagi, Vibashan VS, and Vishal M. Patel. Unsupervised domain adaptation of object detectors: A survey. May 2021.