# 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, including security monitoring, medicine and self-driving vehicles [1]. However, although computer vision has been integrated into industrial applications (e.g., safety and process monitoring) [2]., less research has addressed the issue of equipment detection.

As industrial plants are typically hundreds of meters long, it often becomes frustrating to identify equipment parts for maintenance or replacement. Ore processing plants treat several hundred tons of ore per hour, and the production capacity is constant. Therefore, it is often difficult to properly identify the equipment within a list of thousands of parts in a medium- to large-scale plant.

This work has been commissioned by Metso Outotec Oyj. Metso Outotec offers digital solutions that enable customers to automate their processes in the mining, aggregates and metals industries. In order to ensure that these processes operate as smoothly as possible, it is important to address optimizing them at all stages of production. Recently, Metso has successfully applied computer vision in applications for identifying foreign objects in crushing processes [3], for detecting defects in copper casted anodes [4], as well as for recognizing froth in flotation cells [5], to name a few. However, it is yet to discover if computer vision applications can be used to facilitate maintenance. For these reasons, Metso Outotec has requested an application that would further optimize the processes by leveraging state-of-the-art computer vision algorithms in equipment recognition.

Even though various methods have been implemented for detection of objects in a countless number of fields [3], these methods heavily rely on extensive data collection and training of models in order to accurately identify objects. Moreover, complications arise, as it is often not possible to collect huge amounts of training images from industrial environments due to privacy and confidentiality issues. Luckily, for this project, the images can rather easily be collected from a 3D simulator model of a gold refining plant. However, using the rendered images from a 3D simulator limits the accuracy of the models, as such models do not perform as well on real images due to the domain shift phenomenon [2]. The domain shift occurs when the environmental conditions at the time of capturing training and test images change, as discussed in [2].

Hence, this thesis proposes a cross-domain object detection approach as a solution to automatically localize and identify the equipment in a large environment in order to minimize the delay in production as compared to manual identification.

### 1.2 Thesis objective

The main goal of this thesis is to identify a suitable state-of-the-art object detection technique and to enhance its performance on the custom equipment dataset. The proposed method 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 method should provide a solution for optimizing the laborious process of data collection and labeling. Furthermore, the produced model should address the cases when new objects are incrementally added to the dataset. Such optimization is important not only because training the model from the scratch is a time-demanding process, but also because large plants contain thousands of objects, thus making scalability a critical requirement. Finally, a minimal proof-of-concept application should be prepared to demonstrate the performance.

### 1.3 Methodology

In order to accomplish these objectives, the thesis will first explore state-of-the-art object detection frameworks, libraries and algorithms. Similarly, domain adaptation algorithms will be analyzed in an object detection setup. The most suitable methodologies will then be used to train a cross-domain object detection model.

In order to circumvent regulations regarding accessibility and confidentiality, the dataset utilized for training the model in the experimental scenario is based on the 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 algorithms.

To achieve higher performance in the object detection, the domain shift phenomenon will be addressed using cross-domain object detection algorithm, namely Unbiased Teacher, which in turn uses the Faster-RCNN implementation in detectron2 framework as a detector base. This thesis will contribute to the current knowledge by introducing instance-level domain classifier, which is appended to the Unbiased Teacher network. Additionally, the study will evaluate the feasibility of other transfer learning strategies, such as fine-tuning and class incremental learning, to further enhance scalability of the model. The model will then be integrated into a prototype web application for demonstration purposes. The produced model 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 a minimal proof-of-concept solution based on analyzing and combining different components of existing state-of-the-art models. In addition, this solution will be wrapped in a prototype web application. However, preparing an actual real-life dataset and implementing the solution for a real plant remains outside the scope of this study due to time constraints. Although the proposed method attempts to optimize the data collection and labeling process, this will in practice require many months before the dataset and the model based on real data would be ready for training.

For the user interface, a prototype will be provided in order to showcase the performance of the model. However, the thesis will primarily focus on deep learning algorithms rather than methods to deploy a model in production. For this reason, the prototype will only offer basic functionality. Finally, due to time constraints, a video-compatible model will remain outside the scope of this work.

### 1.5 Structure of the thesis

The rest of this thesis is divided into four chapters. Chapter 2 reviews the literature on object detection models, domain adaptation, the latest cross-domain object detection techniques, and 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 those of other methods using average precision metrics. Chapter 5 summarizes this work by discussing the proposed architecture and 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.
3. Author 1. Reference placeholder 1. July 2022.
4. Author 2. Reference placeholder 2. July 2022.
5. Author 3. Reference placeholder 3. July 2022.