Matthieu Daniel Assa

Real-time task identification in manual assembly steps using neural network based object detection

Human operators will be a central part of future production systems and the difficulty of the operator's tasks will rise. To minimize human errors, a smart camera system has been developed which can monitor human performed assembly processes and provide feedback to operators in real-time in industrial environments. The two main components of a smart camera system are an object detection algorithm and an interface which assesses the correctness of executed assembly steps based on the objects detected on camera frames. The first research questions addressed in this thesis was: "Which object detection algorithms are suitable for inference on an edge device and how do they differ in accuracy and speed?". Neural networks trained with Google AutoML did either not fulfil the requirement of real-time predictions or were not accurate enough to detect completed work steps. A single shot detector MobileNet V1 was re-trained in TensorFlow with quantization aware training and deployed on a central processing unit as well as a tensor processing unit after a TensorFlow Lite conversion and edge tensor processing unit compilation. These neural networks resulted in accurate and fast predictions. Models had trouble to detect objects with high intraclass variability and to distinguish between objects with low interclass variability. The quantization did not significantly reduce the accuracy of detected objects, but converting models to TensorFlow Lite and compiling them for the tensor processing unit did. However, this process only slightly affected the accuracy of predictions regarding correctly performed assembly steps. The second research questions was "which information is needed in combination with object detection to identify and assess the correctness of assembly steps?". The interface was based on a bin approach designed to robustly assess to completion of work steps. A sequence of classifications and overlaps was defined based on the worker instructions so that the interface knew which work step should be completed. Work steps consisting of a classification to recognize objects and overlaps to detect a longer interaction between objects were detected accurately. However, the system had trouble to detect overlaps consisting of a short interaction between objects. A challenge encountered to detect completed assembly steps were occlusions due to operator's hands or head. The accuracy of detected work steps is expected to increase with a different hardware, a multi-camera system and clearer worker instructions.