

# **CERTAIN INVESTIGATION ON HAND-OBJECT POSE ESTIMATION USING HYPER-TUNED SSD-MOBILENETV3 AND YOLOV5 MODEL WITH KINEMATICS IMPLEMENTATION ON SIX-AXIS ARTICULATED ROBOT FOR OBJECT GRASPING APPLICATION**

## **ABSTRACT**

The recent advancement in deep learning methods in Hand-Object pose estimation is an necessary techniques for safe grasping of the objects when the Human-Robot Collaborative tasks is taking place in industries. The identification of position of oriented hand-object from a 2D image is critical problem due to external circumstances like occlusion, low lighting, and cluttered scene, poor detection and blur images. In this research work, the traditional and modern object detection approaches have been reviewed with possible collaborative robot applications. The drawbacks of current processes in human-robot interactive function are analyzed. The traditional image processing is not efficient in locating the object. The earlier version of R-CNN families has limitation in computational complexity, poor accuracy and difficulties in deployment in computing devices for real-time applications. The progress of SSD and YOLO of single shot multi-stage detectors gives a wide opportunity to overcome the above problems in object detection and localization.

The method uses a hyper-tuned MobileNetV3 with Single Shot Detection (SSD) model and YOLOv5 with CSPDarknet53 for the oriented hand-object detection along with its position in an image with better accuracy and less computational memory usage. The detected pose of hand and object is not directly solve the above issues of safe grasping the occluded object. For this, the novel approach is identified and estimated the safe grasping position of object, even though the object is occluded with hand during the collaboration use cases of hand over of object from human to robot. The hyper-tuning of the model architecture provides better improvement in accuracy and without compromising the latency in the detection of hand-object pose and its orientation. In order to overcome the drawbacks of heavy computation cost, high latency and less speed, the Network Architecture Search and NetAdapt Algorithm is used in MobileNetV3 that outperform the network search from the adaptive learning for multi-scale feature extraction and anchor box offset adjustment due to auto-variance of weight in the level of each layers. The squeeze-and-excitation block eliminates the computation and latency of the model. Hard-swish activation function and feature pyramid networks are used to avoid overfitting the data and stabilizing the training of image datasets. Based on the comparative analysis of SSD - MobileNetV3 with its predecessor and YOLOV5 were carried out, the results obtained are 92.8% and 89.7% of precision value, recall value of 93.1 and 90.2%, mAP value of 93.3% and 89.2 respectively. The proposed methods enables better grasping for robots with its own kinematics and trajectory functions by providing the pose estimation and orientation of hand-objects with position tolerance in the range of -1.9 to 2.15 mm along x axis, -1.55 to 2.21 mm along y, -0.833 to 1.51 mm along z axis and orientation range of -0.2330 to 0.2730 along z-axis.