I. INTRODUCTION

Deep learning is a subfield of machine learning whose algorithms are inspired by the structure of the human brain. Computer vision is one of the technological advances that utilizes deep learning. The utilization of deep learning in computer vision can be used in image processing, object classification, and object detection. Based on the reference [1], the development of deep learning in several studies has experienced rapid development, marked by many research fields that have achieved good results and are accompanied by continuous improvement of Convolution Neural Network (CNN), computer vision has arrived at a new peak.

According to the paper [2], CNN as a successful model of deep learning, has the ability of hierarchical learning features, and the research shows that the feature extracted by CNN has a more vital ability to discrimination and generalization than handcrafted features. CNNs are like traditional neural network, but with deeper layers. It has weights, biases and outputs through a nonlinear activation. The neurons of the CNN are arranged in a volumetric fashion such as, height, width and depth [3] [4].

At this time, many object detection models have been published to the public. However, with so many object detection models, it will also cause problems such as determining the model suitable for the needs of a project which will take time because it requires comparisons between each model first. The purpose of the study in this paper is to apply three models that utilize CNN in order to detect objects such as cars and buses using a custom dataset, then analyze the data generated from three models that have successfully detected car and bus objects by comparing parameters such as Mean Average Precision (mAP), Average Recall (AR), and model size, so it is hoped that it can be used as a recommendation in selecting the model to be used in a future object detection project.