1. Introduction

Bananas as a global food industry mainstay can be used in a variety of different methods. Also the maturity of bananas is a significant factor in how it is used. One method of determining the maturity of fruits is by examining the physical features of the fruit, for example its size, shape, or color[18]. In Indonesia, maturity determination in the industry is still mostly done manually which is time-consuming, cost a lot of manpower, and has a high rate of inconsistency due to human involvement[2, 5, 9, 15]. These issues can be resolve by utilizing computer vision technology due to its swift, reliable, cost-efficient, and consistent nature[10]. There have been various cases which technology is used in examining the ripeness of fruits to automate the process such as using sensors, thermal imaging, color extraction, machine learning, and deep learning.

Convolutional Neural Network (CNN) is one of deep learn ing methods that is the most researched about in terms of using computer vision to classify fruit ripeness which is able to achieve a high degree of accuracy and precision in its classification[15]. CNN has multiple architecture structure (i.e. VGG, Inception, AlexNet, ResNet, etc) with each having their own strengths and weaknesses. This study aims to apply and evaluate the performance of Convolutional Neural Network or CNN, especially ResNet50 architecture for optimal banana ripeness classification. Residual network, commonly referred as ResNet, is a popular CNN deep network architecture that deals with vanishing gradient problem. ResNet solves this issue by creating shortcut connection that bypasses one or more of the network layers thereby minimizing the issue with vanishing information/gradient from the earlier layers into deeper layers[7, 11].

Further advancements have been made to make the any baseline CNN models, including ResNet50, to have better per formance by adding attention module [20]. Attention module helps any CNN models by highlighting important features so the model puts more 'attention' to it. One of the most lightweight and easiest to implement module is called Con volutional Block Attention Module (CBAM). This can give a slight boost to existing CNNs with hardly any overhead added. A prominent example of the usage of attention-enhanced model can be seen from the research by Zhao et al[21] in which they proposed a model that incorporated CBAM in its feature extraction network for their peach ripeness task. Their proposed model was able to accurately detect peach even under heavy occlusion conditions beating the state-of-the-art Mask RCNN model.

Hence the purpose of this research is to evaluate the performance of Convolutional Neural Network (CNN) model, in particular Residual Network (ResNet) fitted with attention module CBAM mechanism. The contributions of this study is addressing the limitations of previous studies, such as simpli f ied dataset by training the model on a complex dataset that have numerous variables, ranging from different background types, various angles, lighting, even some stickers on them to closely mimic the real world conditions. Moreover, a proper preparation steps and preprocessing approach is done in order to overcome other difficulties found in previous studies