CHAPTER I
INTRODUCTION

1.1 Background

Artificial Intelligence or AI has experienced a very significant development. There are several components in AI, one of example is computer vision. Computer vision is a major subfield of artificial intelligence and has grown in recent years due to rapid improvements in deep learning [1]. Technology that applies computer vision concepts has been widely used in several things such as fire detection [2], face tracking [3], and autonomous driving [4]. As time goes by, computer vision technology has penetrated into the retail product industry and the retail product industry is entering a new era where a product can be identified and classified automatically using computer vision technology. One example of the application of computer vision technology in the retail product industry is the Amazon Dash Cart. However, automatic recognition of retail products still has problems such as the similarity between one product and another so that this technology is difficult to distinguish. The similarity between one product and another is called intra-class variation. In general, identifying intraclass objects is a challenging task due to several things such as similar subordinate categories often have only minor differences in certain areas of their appearance, intraclass objects may display multiple appearance variations with different scales and different environmental factors such as lighting, background and occlusion may have a great impact on object identification [5].

The application of computer vision technology in the retail product industry has been done in many previous studies. Research by Bikash Santra et al. uses two stages, namely in the first stage using the Reconstruction Classification Network (RCNet) which is basically a Deep Supervised Convolutional Auto Encoder (SCAE) similar to the Supervised Auto Encoder (SAE) [6]. RCNet will serve as a classifier and assist in product recognition under various lighting in the store. The second stage in this study is to improve the classification performance of the first stage on the discriminative part of the product being searched (in an unsupervised manner) and arranged as an ordered sequence to describe the product uniquely then

the sequence is modeled using Convolutional LSTM (Conv-LSTM) and the classification results of the two stages will be combined and determine the product label of the test image [6]. Research conducted by Wenyong Wang et al. proposed a fine-grained image classification method that was improved using the self-attention method [7]. The method used in this study uses the Self-Attention Destruction and Construction Learning (SADCL) method with the VGG-16 and Resnet-50 base models and the proposed method is used to calculate precise fine-grained classification predictions and large areas of information in the reasoning process [7]. Research from Rajib Ghosh discusses a method for discussing various products on supermarket shelves by detecting text blocks on product labels using the Faster Region-Based Convolutional Neural Network (Faster R-CNN) with more than one Region Proposal Networks (RPNs) and then recognizing the text using a Recurrent Neural Network (RNN) classifier [8]. In addition, this research also proposes more than one RPN with various sizes in traditional Faster R-CNN to detect text blocks on product labels and recognize the text using an RNN classifier [8]. Research by Haitian Sun et al. proposed a method called TemplateFree where product detection is based on deep learning without using a single retail store shelf image [9]. The TemplateFree method proposed in this research concentrates on the characteristics of shelves in retail stores where each shelf can be segmented horizontally into layers and vertically into products and also horizontal and vertical separation candidates are sampled and optimized then vertical separation is refined [9]. In addition, the optimization and refinement of the vertical separatrix is assisted by the trained GoogLeNet.

From previous studies, no one discussed specifically about the problem of the high level of product similarity between one another and caused the model to be difficult to identify and classify these products. For that reason, in this research design a system for automatic recognition of retail products to overcome the problem of high similarity between retail products using the YOLOv8 algorithm and the Simple Framework for Contrastive Learning of Visual Representations (SimCLR) algorithm combined with the Bilinear Convolutional Neural Network (B-CNN). The combination applied in this study is to change the default base encoder of SimCLR, ResNet, with the B-CNN algorithm where the algorithm used

is VGG-16. This research is divided into two stages, namely the detection process and the classification process. The prepared dataset is divided into two parts, namely the coarse-grained class dataset and the fine-grained class dataset. The coarse-grained class dataset is used in the detection process and trained using the YOLOv8 algorithm then the training results are used to predict images and the predicted images are cropped based on the bounding box obtained from the prediction results. The cropped images based on the bounding box from the detection process are forwarded to the classification process to be trained using SimCLR algorithm combined with B-CNN. B-CNN combined with SimCLR is used in the base encoder part of SimCLR where the default base encoder of the SimCLR algorithm, ResNet, is replaced using the B-CNN algorithm. The B-CNN model used is the pretrained model, VGG-16. Finally, after obtaining the training results from the classification process, the training weights are transferred back to the model architecture resulting from the combination of SimCLR with B-CNN to be used in the training and evaluation process on the fine-grained class dataset.

1.2 Problem Identification

Figure 1.1. Illustration of intra-class variation in retail products

Many things must be considered in retail product recognition using computer vision technology such as the quality of the dataset used. Although the quality of the dataset used is very good, the current retail product recognition system still has obstacles such as the similarity between one product and another making it difficult to identify and classify the product. The similarity between one product

and another is commonly referred to as intra-class variation. Figure 1.1 shows an illustration of intra-class variation in retail products. Therefore, in this thesis, a retail product recognition system is designed using SimCLR combined with the B-CNN algorithm. The SimCLR model combined with B-CNN used in this thesis is expected to make the model able to distinguish products that have a high level of similarity well.

1.3 Objectives

The objectives of this thesis are as follows:

1. This thesis uses the YOLOv8 algorithm and SimCLR algorithm combined with B-CNN.
2. The model used in this thesis aims to solve the problem of retail product recognition, namely the similarity between one product and another.
3. This thesis evaluates the model using several evaluation metrics such as mean average precision (mAP) and accuracy.
4. In this thesis, a comparison is made with several models trained using the same dataset so that the difference between the proposed method and other models can be seen and also so that the performance of the proposed method can be measured.

1.4 Scope of Work

The assumptions and limitations of the problem in this thesis are:

1. In this thesis, a retail product recognition system is designed and trained using the Grocery Store Dataset.
2. In this thesis, the YOLOv8 algorithm is used in the detection process and the combination algorithm of SimCLR and B-CNN in the classification process where the B-CNN model used is the VGG-16 pretrained model.
3. In this thesis, the designed system is used to be able to distinguish products that have a high similarity between each other.
4. Evaluate the performance of the model using the mean Average Precision (mAP) metric at the detection process and the accuracy metric at the classification process.

1.5 Expected Results

The expected result of this thesis is that the model can identify and classify images that have similarities between one and another precisely and correctly according to their class. This thesis consists of two stages, namely the detection process and the classification process. The detection process uses the YOLOv8 algorithm and the classification process uses the modified SimCLR algorithm which is a combination of the SimCLR algorithm with the B-CNN algorithm where the algorithm used in B-CNN is the VGG-16 pretrained model. The model evaluation process uses the mAP metric for the detection process and the accuracy metric for the classification process. From the analyses carried out based on the evaluation results, it is expected to explain how the model's ability to identify and classify images accurately and can distinguish between one product and another product that has a high level of similarity. This research is also expected to contribute valuable knowledge in the field of artificial intelligence, especially in the field of retail product recognition.

1.6 Research Methodology

The methodology used for the process of completing this research consists of several stages such as:

1. Do literature study by searching and understanding journals, articles, web and other sources related to the topic of this thesis.
2. Create and design system workflows using approaches such as flowcharts and block diagrams.
3. Implement the system on the software.
4. Experiment and evaluate system performance.
5. Analyze the experimental results obtained and compare them with the results of other similar experiments and make conclusions.

1.7 Structure of Thesis

The structure of this thesis writing can be seen as follows:

- **CHAPTER II: BASIC CONCEPT**

  This chapter contains basic concepts about the algorithms used in this thesis.

- **CHAPTER III: SYSTEM MODEL AND METHOD**
  This chapter contains the system design and experimental details used in this thesis.
- **CHAPTER IV: EXPERIMENTAL RESULTS AND ANALYSIS**
  This chapter discusses the results and analysis of the experiments.
- **CHAPTER V: CONCLUSION**
  This chapter contains conclusions from the results of the research.