CHAPTER 1

INTRODUCTION

This chapter includes the following subtopics, namely: (1) Rationale; (2) Statement of the problem; (3) Objective and Hypothesis; (4) Scope and Delimitation; and (5) Significance of the study.

1.1 Rationale

Punishment or crime involves any act or behavior that is unlawful or against the provisions of law or other legal-related standards. The one who was involved in a crime is called a criminal or a suspect involved in a criminal act. After this activity, crime prevailed in each nation, and crime prevention was also part of the historic process. This technique can create a society that is free from criminals with helpful intentions. However, it must be pointed out that as per the records, only 23% of crimes are reported to law enforcement organizations. The subject of this study was acknowledged suspect as a fugitive beforehand. In reference to the Indonesian Dictionary (KBBI), a fugitive means a person who is known by the police or who has escaped a prison [24].

Several strategies for apprehending persons participating in criminal activities have been developed and used over time. The technique is intended to detect offenders in places prone to crime successfully. For example, CCTV cameras can be installed in public and private areas to monitor suspicious activity [2]. According to studies [19], CCTV is effective in reducing crime rates. However, constantly monitoring surveillance footage is an exhausting and routine process that needs visual concentration. This might increase the chance of misidentification [12]. Automated surveillance with intelligent technologies, such as facial recognition systems, can successfully handle these issues.

Facial recognition is a biometric technology that provides substantial benefits in terms of high accuracy and low noise. This approach uses programmed computer software to take face features as input. The output is a box that restricts the face of someone suspected to belong to a suspect. Identification and verification can be accomplished via Closed-Circuit Television (CCTV) images or video recordings. Although CCTV systems can detect criminal activities, the recorded images are frequently hazy or blurry, making it difficult to identify suspects. Furthermore, the criminals frequently utilize various things to cover up their facial features or wear masks to avoid detection. Furthermore, the suspected individual can change his or her appearance physically, as seen by wearing glasses, clothing, head coverings, various haircuts, and other changes.

There have been many studies on suspect identification based on face recognition, one of which is research by Hyunbin [12] et al. The system is described by FaceNet embedding

technology. The aim was to verify each passing person's face in CCTV footage, then extract an area of the face and integrate it in FaceNet, and then match it to a mugshot database depending on the number of similarities. FaceNet embedding uses CNN [22] (Convolutional Neural Network) to extract features from a person's face. The CNN model's principal goal is to enable accurate generalization of previously unknown data [17]. To increase the generalization of the network, it is required to add depth to the network i.e., the number of layers, and width of the layers i.e., the number of nodes in the layer [18]. However, one typical issue that develops when working with big networks is overfitting, which occurs when networks correctly evaluate training data but exhibit an enormous number of faults in testing data. The obvious answer to this problem is to apply smaller networks and combine more contextual information in the form of previous representations such as modeling deformations [14, 27, 32], yet this technique is ineffective for more complicated situations. The pooling method is another approach to handling generalization and overfitting of the network. Because of its computational effectiveness, CNN typically utilizes two forms of pooling: max and mean pooling. However, max pooling does not ensure the generalization of test data and failures when the magnitude of the major component is less than that of the minor component. Nevertheless, mean pooling might not work in cases when there are some zeros in the pooling region [1]. As a result, it is required to utilize a particular pooling mechanism, such as hybrid pooling. According to [28, 29], hybrid pooling has been proven to significantly increase CNN generalization abilities.

According to the current gap, identifying suspects with FaceNet embedding using standard pooling approaches remains a crucial challenge. This thesis proposed hybrid pooling in FaceNet embedding to identify suspects from CCTV footage. This proposed method is a solution to improve the generalization of CNN on FaceNet embedding to improve the performance of the system.

1.2 Statement of the Problem

FaceNet is one of the face recognition algorithms that can be implemented in suspect identification systems. FaceNet uses embedding to extract face features based on a convolutional neural network. However, CNN is mighty at managing pattern recognition tasks. However, the advantage of recognizing some unseen data is that it also has a weakness in predicting it. Several things can be done to enhance the generalization of CNN, such as using a special pooling layer called hybrid pooling. Hybrid pooling has been shown to improve some of the shortcomings of traditional pooling. Using hybrid pooling will improve the accuracy and generalization capabilities of the FaceNet architecture. In other ways, this process will allow the system to be more reliable and efficient in suspect recognition under different conditions.

1.3 Objective and Hypotheses

In order to solve the generalization CNN problem and provide higher performance than models using conventional pooling, this research aims to present a suspect identification system that uses hybrid pooling on FaceNet embedding.

The use of hybrid pooling in FaceNet embedding within the suspect identification system is expected to significantly improve identification accuracy compared to the baseline methods that only use max pooling or average pooling. This is because hybrid pooling can combine the advantages of both methods, namely the ability of max pooling to extract the most dominant features and the ability of average pooling to maintain more detailed spatial information [16, 28, 29]. Thus, hybrid pooling can produce more robust feature representations, thereby enhancing the performance of the identification system. In addition, FaceNet has proven effective in generating low-dimensional yet highly discriminative face embeddings [22], making it very suitable for identification tasks. The combination of hybrid pooling and FaceNet is expected to produce a more accurate and reliable identification system.

1.4 Scope and Limitation

This session consists of two parts: limitation and scope of research.

1.4.1 Limitations

This thesis has several limitations, such as the specific definition of the suspect, only one person for each CCTV frame, and CCTV installed in indoor areas.

Definition of Suspect

This study focuses on people who are obtaining fugitive status. The purpose of the research is to identify the individuals suspected to be fugitives who passed by CCTV-recorded areas. The method used involved an analysis of CCTV recordings to find similarities between the faces captured on them and those found on wanted persons' lists or mugshot datasets. For instance, if the CCTV footage shows a man whose face is similar to one of the mugshot datasets recorded by CCTV within a given locality, then such a person becomes suspected, and relevant authorities may take further actions, such as an investigation or temporary detention, to ascertain his or her actual identity. Hence, this study plays an important role in speeding up the process of identifying and arresting criminals, especially where time is of the essence and rapidity is required.

Only One Person for Each CCTV Frame

There are reasons why there is only one person in each video in the CCTV dataset that was used for this study. When the CCTV video contains many individuals, there is a significant probability of overlap or occlusion, which occurs when one person's face is partly or entirely obscured by another person. This can result in unsuccessful detection or inaccurate identification. The presence of a single individual in each video minimizes this risk, leading to more reliable and accurate detection. Furthermore, a video showing a single individual improves the annotation procedure by removing any ambiguity in giving the label to each identified face. Implementing this approach speeds up the process of creating the dataset and minimizes the possibility of mistakes in data labeling [12].

CCTV Only Installed in Indoor Areas

Creating a closed-circuit television (CCTV) dataset in indoor locations offers numerous benefits over outdoor locations, particularly in the field of suspect identification studies utilizing facial recognition technology. Firstly, indoor lighting is easier to regulate and modify [15]. The significance lies in the fact that optimal lighting conditions improve the quality of images obtained by closed-circuit television (CCTV) cameras, thereby increasing facial recognition accuracy. Secondly, the lighting outdoors can change significantly depending on the time of day and weather conditions, which could potentially impact the facial recognition system's optimal performance [15]. Indoor locations frequently include a consistent and arranged background, which facilitates improved facial recognition. The challenges posed by a crowded and diverse outdoor background may reduce the accuracy of the system, making it difficult to distinguish faces from other objects. Thirdly, data from indoor CCTV tends to have lower visual noise compared to outdoor. The noise can come from many sources, including additional background motion, which may affect the ability of facial recognition systems to accurately identify persons [15]. Given all of these factors, building a CCTV dataset in indoor locations can provide better and more reliable outcomes for studies on suspect identification by facial recognition.

1.4.2 Scopes of Research

The research scope of the thesis consists of the scope of knowledge, the scope of facility, the scope of user, and the scope of usability. First, the scope of knowledge encompasses various areas, including the Disguised Face Dataset, suspect identification, face detection, feature extraction, deep learning, and face recognition. Second, the scope of the facility includes the application of advanced sensor and camera technologies to gather face data with exceptional precision and the implementation of closed-circuit television (CCTV) systems equipped with advanced facial recognition technology at important sites. Third, the scope of users for this study includes police and CCTV surveillance. Last, the scope

of usability includes the pursuit of fugitives who have eluded apprehension, as well as examining the footage obtained from CCTV

1.5 Significance of the Study

The novelty of this research is the use of hybrid pooling within FaceNet embedding on suspect identification systems. The goal is to improve the generalization of CNN-based FaceNet embedding as a feature extraction so that it can improve the performance of the system.

The systems built in this research can be developed and used in various aspects, such as helping law enforcement agencies find fugitives, especially those on the search list of people caught by CCTV cameras automatically. In addition, the system can also be used as a face-based authentication security system in restricted access areas. For example, it can control access to the lab room, allow access only to registered individuals, and trigger an intruder alarm if an unregistered person tries to enter.