International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Face Recognition on Low Resolution CCTV Video using GhostFaceNets

Indrabayu[1], Andi Bukti Djufrie[2], Muhammad Amri Akbar[2], Budi Armansyah[2], Muhammad Fadhil Bin Bahrunnida[1], Nublan Azqalani[1]

[1] Informatics Department, Hasanuddin University, Makassar, Indonesia
[2]Department of Innovation and Technology Development, Makassar City Regional Research and Innovation Agency, Makassar, Indonesia

E-mail address: indrabayu@unhas.ac.id, Bukti.Djufrie@gmail.com, amriakbar72@gmail.com, budi.armansyah047@gmail.com, fdl.mks97@gmail.com, nublan.azqalani@gmail.com

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: This research aims to develop a face recognition system for low-resolution CCTV cameras in school areas, as part of the Jagai Anakta' program initiated by the government of Makassar. The system is based on the GhostFaceNets algorithm, which uses ghost modules, feature fusion, and attention mechanism to achieve high accuracy and efficiency in face recognition. The system is trained and evaluated on a custom dataset collected from SD Telkom Makassar, which comprises 18 individuals, including 10 students and 8 teachers. The dataset is preprocessed and augmented to enhance its diversity and robustness. The system's performance is measured using three evaluation stages: loss and accuracy based on the ArcFaceLoss matrix, loss and accuracy using .bin files, and accuracy and suitability using CCTV video. The results show that the system can handle up to five objects in the real-time frame with high confidence, but it faces difficulties when the number of objects increases. This study advances the GhostFaceNets algorithm for enhanced performance in low-resolution image processing and multi-object detection in real-time scenarios. Traditional benchmarks of GhostFaceNets primarily involve single-face detection in still images. This research extends this by adapting the architecture to effectively handle multiple faces in real-time video feeds. Key modifications include the removal of the DFC attention branch to prevent significant data representation loss and adjustments in the number of embedding layers. The integration of L2 regularization and the use of PReLU activation functions further refine the model's training effectiveness on the custom-compiled dataset. The system also shows a high degree of generalization and adaptability to different environmental conditions and scenarios. The system can be a useful tool for monitoring and providing early warning of criminal activities in school areas using a data-driven approach.

Keywords: Face Recognition, GhostFaceNets, Low Quality Image, Makassar, Custom Datasets

1. INTRODUCTION

Surveillance systems are technologies that enable the observation or monitoring of a series of information and behavior within a certain area. They typically involve the use of closed-circuit television (CCTV) cameras for visual data collection, as well as a series of algorithms for detection and tracking. These technologies have undergone significant development over the last few decades. However, the implementation of surveillance systems, especially in Makassar, is still not widespread. The installation of CCTV cameras at several locations has been a valuable asset for achieving the vision of a safe and comfortable Makassar Smart City.

The Jagai Anakta' program is one of the flagship initiatives of the government of Makassar to achieve the smart city status. This program aims to integrate the use of technology and the participation of the public in ensuring the safety and well-being of children in their daily activities, especially in school areas. However, crimes against children still persist in Makassar. According to data from the Operations Control Bureau, (Biro Pengendalian Operasi, Mabes Polri), collected by the Central Statistics Agency (Badan Pusat Statistik), South Sulawesi ranks first among the provinces with the highest number of crimes against personal freedom in 2021, with 375 cases, consisting of 9 cases of kidnapping and 366 cases of child exploitation [1]. Therefore, this research

E-mail:author's email

attempts to develop a system that can monitor and provide early warning of criminal activities in school areas using a data-driven approach that is suitable for the environmental conditions and the available infrastructure.

Face recognition is a biometric technology that aims to identify or verify the identity of a person based on their Face features. Face recognition has been developed for various applications such as security, surveillance, authentication, entertainment, and social media. The development of Face recognition can be divided into several stages, depending on the methods and techniques used. The first stage is the geometric-based methods, which use the distances and angles between Face landmarks to represent and compare faces. The second stage is the appearance-based methods, which use statistical or machine learning techniques to extract features from the whole face or its regions. The third stage is the feature-based methods, which use local descriptors such as SIFT, SURF, or LBP to capture the local texture and shape information of faces. The fourth stage is the deep learning-based methods, which use convolutional neural networks (CNNs) or other deep architectures to learn high-level and discriminative features from large-scale face datasets. The fifth stage is the lightweight methods, which use efficient and compact models to reduce the computational cost and memory demand of face recognition on resource-constrained devices.

The development of Face recognition has also faced various challenges and issues, such as pose, illumination, occlusion, expression, ageing, plastic surgery, and low resolution. These factors can affect the performance and accuracy of face recognition systems in real-world scenarios. To address these challenges, researchers have proposed different solutions and improvements, such as knowledge distillation, quantization, low-rank approximation, ghost modules, long-range dependencies, and bio-inspired models. These solutions aim to enhance the robustness and efficiency of face recognition systems while maintaining high accuracy. Moreover, researchers have also introduced various evaluation protocols and benchmarks to measure and compare the performance of different face recognition methods. Some of the major and challenging face datasets are LFW, MegaFace, MS-Celeb-1M, IJB – A / B / C / D / E / F / G / T / S, CASIA - WebFace, VGGFace/VGGFace2, CelebA / CelebA-Spoof / CelebA – Mask / CelebA – HQ / CelebA - 3D / CelebA - 3D -Mask / CelebA - 3D – Spoof / CelebA-3D-HQ/CelebA-3D-HQ-Mask/CelebA-3D-HQ-Spoof [2]–[4]. These datasets provide different levels of difficulty and diversity for face recognition tasks.

Face recognition is a dynamic and evolving field that has witnessed significant progress and innovation in recent years. The development of Face recognition has been driven by the advances in computer vision, machine learning, deep learning, and biometrics. Face recognition has also been influenced by the increasing demand and application of this technology in various domains and scenarios. Face recognition is expected to continue to grow and improve in the future, as new methods and techniques are developed and new challenges and opportunities are encountered.

GhostFaceNet [5], an algorithm that performs face recognition on low-quality images, is a deep neural network that uses ghost modules, feature fusion, and attention mechanism to achieve high accuracy and efficiency. Ghost modules are lightweight convolutional layers that generate more feature maps from fewer parameters, reducing the computational cost and memory consumption of the network. Feature fusion is a technique that combines the features from different layers of the network, enhancing the discriminative power and robustness of the network. Attention mechanism is a method that focuses on the most relevant regions of the face image, improving the performance of the network in handling occlusion, pose, and expression variations. GhostFaceNet has been shown to outperform other state-of-the-art face recognition models on several public benchmarks, such as LFW, CFP-FP, AgeDB-30, and MegaFace.

The motivation for using GhostFaceNet [5] in this research is its outstanding performance on several public datasets that are commonly used as benchmarks, as well as its low computational burden due to the Ghost Module. This can be very beneficial for real-time implementation and custom dataset creation, which involve various augmentation processes. The image data augmentation processes aim to simulate possible changes in the detection scope, such as rain and variations in brightness and light intensity.

2. RELATED WORKS

The deployment of deep learning-based face recognition (FR) models on resource-constrained embedded devices such as mobile phones faces challenges due to the high computational cost and memory demand of these model [6]–[8]. Most FR models rely on a large number of parameters to achieve high accuracy [9], [10]. Recent research in the field of FR has shown significant progress in addressing these challenges. Some methods use knowledge distillation (KD) to transfer the knowledge from large pre-trained state-of-the-art (SOTA) FR models to smaller models [11]. Other methods use quantization techniques to reduce the model size by using lower-precision representations [12], or low-rank approximations to reduce the computational complexity [13]. Designing lightweight deep neural networks has emerged as one of the most promising approaches to improve the trade-off between speed and accuracy in recent years [14], [15], [16], [17], [18]. Lightweight models are characterized by having low computational complexity, typically in the range of 1G floating point operations (FLOPs). Some examples of lightweight

models that achieve competitive results in image classification are SqueezeNet [14], MobileNets [15], ShuffleNets [16], [17],VarGNet [18], and MixNets [19]. However, only a few studies have proposed to use lightweight deep learning architectures as the backbone of FR models. For instance, MobileFaceNet [20], ShuffleFaceNet [21], VarGFaceNet [22], and MixFaceNets [23] use MobileNetV2 [15], ShuffleNetV2 [17], VarGNet [18], and MixNets [19] respectively as the backbone of their FR models. Recently, very lightweight backbones named GhostNetV1 and GhostNetV2 were introduced for image classification tasks [24], [25]. GhostNetV1 proposed a novel Ghost module that generates more features using fewer parameters, which was later extended in GhostNetV2 to incorporate long-range dependencies as well. Experiments demonstrated that GhostNetV1 and GhostNetV2 outperform their competitors at different levels of computational complexity[24]. GhostNetV1 notably surpassed MobileNets [15], ShuffleNets [16], [17], and many other models in image classification tasks, while GhostNetV2 improved upon the first version.

In developing a face recognition system, it's crucial to consider not only the algorithm's accuracy and computational time but also the system's stability in response to the number of know (registered) and unknown (unregistered) faces, based in a model trained with a custom dataset. In the study "Attendance system using Machine Learning-Based face detection for meeting room applications" [26], the results show a system that can detect faces, send data to the server, display data, and store unknown faces in one folder on the node. The system is relatively stable when detecting registered users. However, the system is less stable when detecting unregistered users. Instability results can occur when detecting the faces of three or more people in one camera. Moreover, there is a minimum distance to detect faces accurately, which is about 2 meters. This poses a significant problem to solve, how to make the system performance unaffected by the number of registered and unregistered objects and the distance of the object to the camera. This research will optimize the algorithm selection, parameter and hyperparameter tuning, and custom data preprocessing to address this problem.

Data retrieval methods, parameter and hyperparameter tuning, preprocessing and augmentation of custom data are essential steps in developing optimal machine learning models [26]–[30]. These steps depend on the objectives and the constraints of the system and the available infrastructure.

3. DATASET PREPARATION

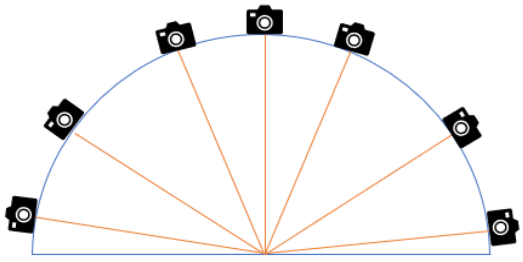
A pivotal phase in the GhostFacenets model development is dataset preparation, encompassing two primary tasks: data collection and data preprocessing. Data collection involves gathering and choosing face images for training and evaluation, while data preprocessing refines these images through several methods to ensure compatibility with the model's input requirements. In this sub-chapter, we will elaborate on these two sub-steps in detail and explain how they influence GhostfaceNets model performance.

A. Data Retrieving

The data used in this research is primary data collected directly at the research location. The first data was captured using a camera with seven viewpoints of the object; 5, 30, 65, 90, 115, 150, and 175 degrees tilt to the object. The data collection process can be illustrated by Figure 3.1 below. suggest that using multiple viewpoints can improve the robustness and accuracy of face recognition systems.

Figure 1 Data Retrieval Scenario

Data collection was conducted directly at SD Telkom



Makassar. Students and teachers who consented to have their facial data taken sat at a certain point, and the camera captured their facial data from several perspectives to maximize the extraction of each individual's facial features. The camera used seven viewpoints of the object; 5, 30, 65, 90, 115, 150, and 175 degrees tilt to the object.

In addition to collecting facial data using a camera, each individual was also asked to walk through the school area where CCTV was installed. They walked past CCTV in several scenarios, either individually or in groups. Capturing data via CCTV video was intended to facilitate the model to learn facial features directly from the results obtained on the device that would be used. The CCTV video data was then processed to extract the face images of each individual using a face recognition algorithm. The facial images are then organized in one folder according to the identity of each individual with the following folder structure (Figure 2):

```
.                    # dataset folder
├── Nama_1           # person folder
│   ├── 100.jpg # face image
│   ├── 101.jpg # face image
│   └── 102.jpg # face image
├── Nama_2           # person folder
│   ├── 111.jpg
│   ├── 112.jpg
│   └── 113.jpg
├── Nama_....
│   ├── 707.jpg
│   ├── 708.jpg
│   └── 709.jpg
```

Figure 2 Dataset Folder Structure

The folder structure as shown in the image above is used to simplify the data labeling process for each image as is typically done when using deep learning algorithms. The folder name indicates the identity of the individual whose face images are stored in that folder. This way, the model can automatically assign the correct label to each image based on the folder name

B. Data Prerpcessing and Augmentation

At the pre-processing and augmentation stage, the existing data is processed and modified to make it suitable for the algorithm input. In this research, the pre-processing stage includes cropping and resizing the images to a fixed size of 112x112 pixels. The augmentation stage includes adding different levels of brightness and darkness to the images, as well as applying a rain filter to simulate adverse weather conditions. Figure 3.3 below illustrates the pre-processing and augmentation process performed on each image. These techniques are intended to increase the diversity and robustness of the data and improve the generalization ability of the model.

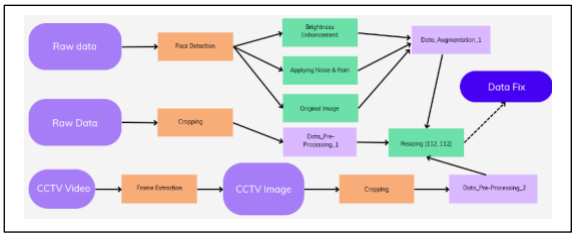


Figure 3 Pre-processing and Augmentation Process

1) Data cropping, is the process of refining image data, which initially contains numerous non-facial elements, to isolate and concentrate solely on the individual's face region. This process is intended to reduce the noise and complexity of the data so that features that are not related to the face are excluded from the training process, thus reducing the computation required by the algorithm to recognize the face.

2) Brightness enhancement, is the manipulation of the luminance level of image data, either by increasing or decreasing it. This process is performed to introduce variability in the data before feeding it into the algorithm training process, so that the facial recognition model can operate robustly regardless of the light intensity changes in the detection and recognition environment

3) Applying noise and rain filter, is the process of adding noise and rain effects to image data using a coding technique. This process aims to enhance the model's performance in outdoor settings, where the CCTV cameras are exposed to various weather conditions. This is relevant for some schools in Makassar City that have outdoor CCTV installations.

After applying cropping, brightness enhancement, and noise and rain filters to all data, the data is resized to 112x112 pixels (resizing). This is done to reduce the computational complexity of the neural network, as larger input images require the network to learn from four times as many pixels, which increases the training time for the architecture.

4. METHODOLGY

This research employs a custom dataset collected by SD Telkom Makassar. The dataset comprises 18 individuals, including 10 students and 8 teachers. After performing a series of dataset preparation procedures, the face recognition model training process is conducted based on the Ghost Face Net algorithm. To evaluate the model performance at each epoch or iteration, this research adopts three evaluation stages: evaluation of loss and accuracy based on the ArcFaceLoss matrix without intervention, evaluation of loss and accuracy using .bin files created manually using several separate lines of code, and evaluation of accuracy and suitability of the model in recognition using previously captured CCTV video (direct observation).

A. GhostFaceNets Model Training

This research constructs a model based on the GhostFaceNets algorithm, which has two architectural versions: version 1 and version 2. The main difference between these two versions is that GhostFaceNet version 2 incorporates the DFC attention branch during the data training process, as the DFC attention branch is efficient and capable of capturing long-distance dependencies between pixels located at different spatial locations. These two versions are then trained using same custom data with several parameter and hyperparameter tuning. Several aspects are considered during the process of constructing a model based on GhostFaceNets, such as:

1) For Selection of model type and architecture : The choice of model type and architecture depends on the characteristics of the data and the capabilities of the existing infrastructure. In this research, GhostFaceNets

version 1 demonstrates better and more optimal results in predicting data. This is because the use of the DFC attention branch in GhostFaceNets version 2 apparently causes computational savings, which results in less representation of the data to be trained, as dependencies between pixels are removed. However, the images used cover enough specific areas of the face that require more information.

2) Number of Embedding_layer Dimensions: The embedding layer is a representation of an image in matrix form. The larger the dimensions of the embedding layer specified, the more information from the image it can accommodate. However, this can also result in a lot of redundant or irrelevant information from the image and require more computing power. After conducting several experiments, it is found that the optimal number of embedding layer dimensions is 512 for each architecture.

3) Addition of L2 regularization : L2 regularization or Ridge regression is a method for adding penalties as model complexity increases. The regularization parameter (lambda) penalizes all parameters except the intercept, so that the model generalizes to the data and avoids overfitting. Overfitting is a situation where the model will fit too closely to each object, for example if there are 18 objects during the training process, the overfitting condition causes the model to tend to overpredict one of the objects. This is due to the complexity of the layers in the architecture.

4) Addition of Activation Function : Activation functions in nerural networks determine whether neurons should be activated or deactivated. They modify each neuron's output in a layer, facilitating complex and non-linearity, these functions allow neural networks to learns and represent more intricate data patterns, going beyond what's achievable with mere linear transformations. In this research, the activation function used is PReLU. The Parametric Rectified Linear Unit (PReLU) activation function is a variation of the ReLU (Rectified Linear Unit) activation function that has an additional parameter that allows smoother gradients at negative values. PReLU was introduced to address the problem of "dying ReLU" that can occur when training artificial neural networks using ReLU functions.*architecture.*

In the context of the GhostFaceNets model in this study, the embedding layer plays a role in producing better feature representation, L2 regularization helps control model complexity, and replacing ReLU with PReLU creates more complex non-linearity in the model. These are all techniques for optimizing model's performance and generalizability to model data.

B. Evaluation file in .bin format

As stated at the beginning of this sub-chapter, one of the methods for evaluating the GhostFaceNet model is to use a file with the .bin format, which contains pairs of images that are either genuine or forged. The creation of this .bin format file involves a coding procedure with the assistance of TensorFlow, in Figure 3.

In this code snippet, the image_list variable stores the names of random image pairs that are used as a reference in evaluating the model. The name in the first sequence is paired with the name in the second sequence, and so on. The is_same_list variable then contains the Boolean value for each pair of images in the image_list, indicating whether they are TRUE or FALSE. An illustration of a .bin format file can then be seen in Figure 4

The model evaluation at each epoch produces two values of loss and accuracy, respectively. This facilitates the algorithm to find the optimal parameters and hyperparameters that maximize the accuracy values and minimize the loss values.

```
# Save to bin
import pickle
import tensorflow as tf
from skimage.io import imread

image_list = ...
is_same_list = ...

bb =
[tf.image.encode_jpeg(imread(ii)).numpy()
for ii in image_list]
with open("data_fix.bin", "wb") as ff:
    pickle.dump([bb, is_same_list], ff)
```

Figure 4 bin evaluation

In this code snippet, the image_list variable stores the names of random image pairs that are used as a reference in evaluating the model. The name in the first sequence is paired with the name in the second sequence, and so on. The is_same_list variable then contains the Boolean value for each pair of images in the image_list, indicating whether they are TRUE or FALSE. An illustration of a .bin format file can then be seen in Figure 5 below:

Figure 5 File with .bin format

The model evaluation at each epoch produces two values of loss and accuracy, respectively. This facilitates the algorithm to find the optimal parameters and hyperparameters that maximize the accuracy values and minimize the loss values.

C. Testing the effectiveness of the model via CCTV video

After the algorithm generates different models in each iteration, the model with a combination of high accuracy and low loss is then tested repeatedly on video data that has been captured using CCTV in the Telkom Makassar Elementary School area.

The CCTV video data used follows three main scenarios: in the first scenario, 18 people whose data has been collected will enter individually through the area covered by CCTV. The area covered by CCTV is allowed to contain some individuals who are not registered in the model. This is to evaluate the model's performance and efficiency.

The second scenario is similar to the first scenario, but 18 individuals who have been registered in the model enter the frame in pairs between one student and one teacher as a companion. The final scenario involves all individuals entering the frame simultaneously.

This entire scenario aims to evaluate the model's ability to recognize 18 individuals as well as other individuals not registered in the model. The results of the analysis are then examined using the confusion matrix, ROC curve, and accuracy score.

5. RESULT

As discussed in the previous sections, this research focuses on analyzing the performance of GhostFaceNets on a custom dataset of 18 individuals at SD Telkom Makassar based on CCTV video. A sample of the custom dataset that has undergone the pre-processing and augmentation procedures can be observed in Figure 6 below:

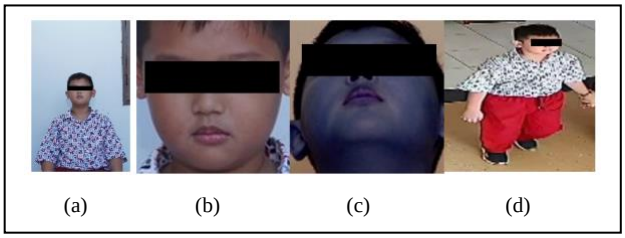


Figure 6 some examples of data processed in the algorithm: (a) original data, (b) cropped data, (c) augmented data, and (d) CCTV video data

The data was trained using GhostFacenet V1 and V2, with several adjustments to the range of parameters and hyperparameters. A total of 20 experiments were conducted, where each architecture was tuned using 10 different parameter ranges and hyperparameters with 200 epochs for each experiment. The observation results indicate that the best model was obtained at the 9th epoch in the 8th experiment using GhostFacenet V1. The learning rate in the 8th experiment for each epoch can be seen in Figure 7 below

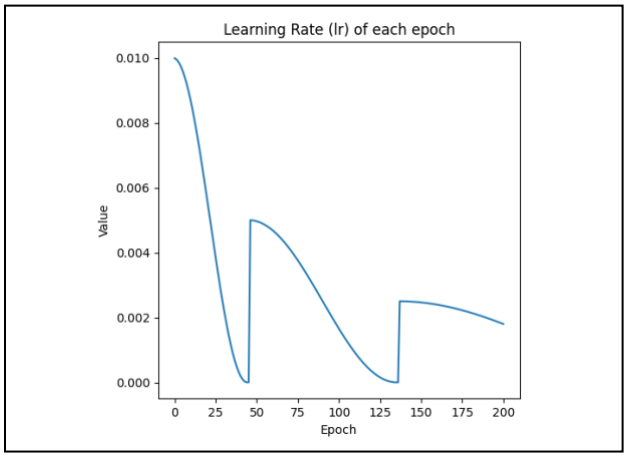


Figure 7 Graph of changes in learning rate in the 8th experiment

As shown in Figure 7, the lowest learning rate occurs around epoch 50 and epoch 130. However, the accuracy and loss graphs for these two epoch ranges do not demonstrate satisfactory results (either without or with .bin files). The changes in accuracy and loss values in the same experimental session are illustrated in Figures 8 and 9 below:

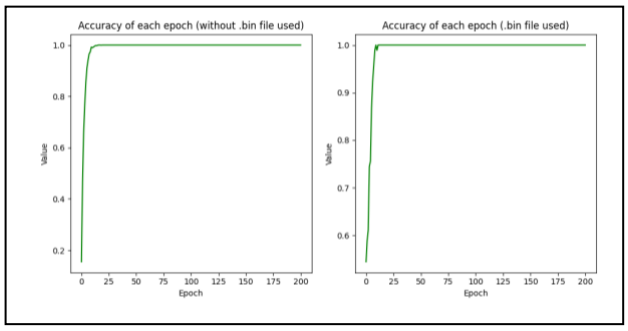


Figure 8 Graph of changes in accuracy value in the 8th experiment

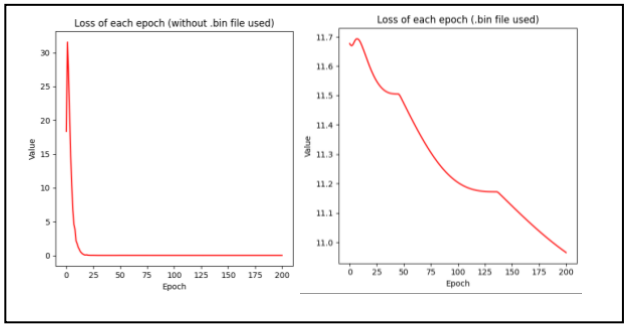


Figure 9 Graph of changes in loss value in the 8th experiment

During the experiments, the model was also evaluated against the FR/FV benchmark by employing the trained model as a feature extractor and computing the Cosine distance between feature vectors in all verification experiments. The FR/FV benchmark graph for each epoch in the 8th experiment is depicted in Figure 10 below.

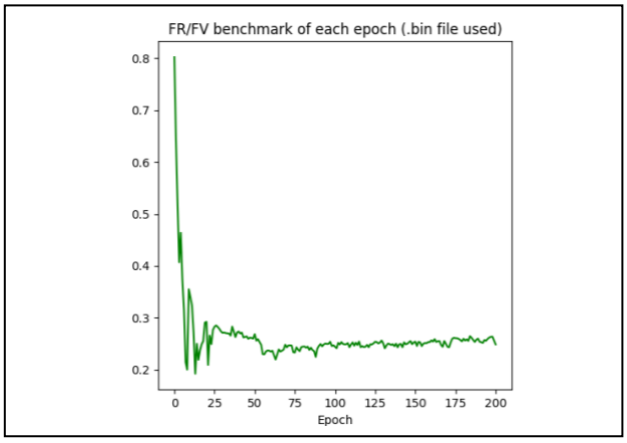


Figure 10 Graph of changes in FR/FV (Cosine distance) in the 8th experiment

In order to select the best model for the task, the evaluated method is by using several metrics for each experiment, such as accuracy, loss, and cosine distance. A qualitative analysis was also performed by visually inspecting the model outputs on the video data. The model that achieved high accuracy and low loss was chosen, and tested repeatedly on different video samples until the optimal results were obtained in the 8th experiment at epoch 9. This model, trained on a custom dataset, achieved an accuracy of 0.9730, a loss of 15.5367, and an AUC of 0.993. The ROC curve for this model is shown in Figure 11 below:

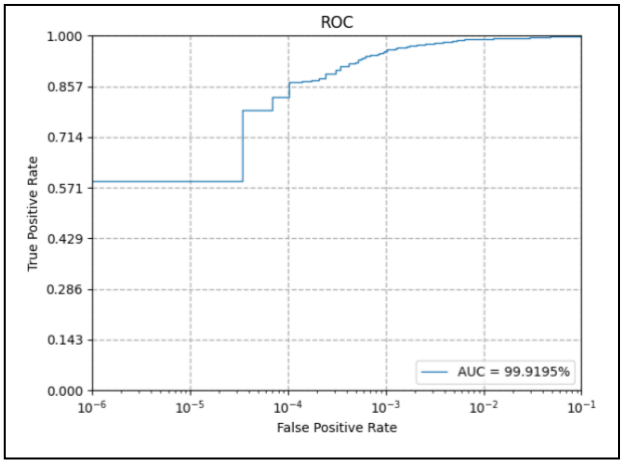


Figure 11 Graph of ROC curve at 9th epoch in the 8th experiment

The scenarios were designed to measure how the system handled the following situations: (1) only one known face in a frame, (2) only two known faces in a frame, and (3) all known faces in a frame. The results of the analysis are presented in Tables 1, 2, and 3, which show the value of cosine distance of each object for each scenario.

TABLE I. SYSTEM ANALYSIS USING CCTV VIDEO IN THE FIRST SCENARIO (ONE OBJECT PER SCENE)

		Predicted Label / Class (on Frame)																	
		1	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***	***11***	***12***	***13***	***14***	***15***	***16***	***17***	***18***
Original Label / Class (on Frame)	***1***	0,94	0,57			0,58		0,51		0,62			0,51			0,61			
	2	0,72	0,90		0,76														
	3			0,98	0,64	0,69	0,73					0,64			0,75				
	4		0,61	0,53	0,95	0,69	0,79		0,57	0,57	0,61						0,54	0,69	0,67
	5	0,80	0,61	0,55	0,73	0,94		0,73	0,82	0,64		0,61			0,51				
	6			0,87			0,97								0,77				
	7					0,53		0,94			0,57		0,52	0,56		0,68	0,80		
	8				0,69	0,51	0,65	0,50	0,97	0,54	0,61	0,80					0,61		0,56
	9	0,67				0,66		0,55	0,65	0,84				0,58		0,69			
	10				0,73			0,52		0,64	0,91					0,70	0,74		
	11								0,53	0,71		0,95		0,76			0,81	0,58	0,67
	12				0,53			0,66	0,64	0,61	0,64		0,84	0,56				0,60	0,62
	13					0,51	0,62			0,63		0,58		0,86		0,53	0,51		0,58
	14			0,69			0,77								0,90				
	15							0,84			0,65	0,56		0,62	0,70	0,85			

		Predicted Label / Class (on Frame)																	
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
	16				0,68			0,62		0,66	0,64		0,67		0,67		0,97		
	17				0,67	0,57		0,60	0,57	0,53	0,68			0,58		0,65	0,61	0,91	0,63
	18		0,67		0,50			0,57	0,51	0,67	0,68		054	0,59		0,66	0,63	0,59	0,87

TABLE II. SYSTEM ANALYSIS USING CCTV VIDEO IN THE SECOND SCENARIO (TWO OBJECT PER SCENE)

			Predicted Label / Class (on Frame)																	
			1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
	1	*9*	0,88								0,90									
	2	*10*		0,86								0,76								
	3	*11*			0,95								0,90							
	4	*12*				0,96								0,81						
	5	*14*					0,85									0,92				
	6	*13*						0,95							0,81					
	7	*15*							0,70								0,85			
	8	*16*								0,79								0,92		
	17	*18*																	0,68	0,80

The cosine distance score in the table is the highest score for each scenario in the frame. There are several cosine score values for each scenario, but the highest score is chosen so that when implementing the system in the field the minimum score of each existing maximum score can be used as a reference as a minimum standard of cosine distance. The cosine distance score between facial features states how similar a class of facial objects in the frame is to the results predicted by the model.

TABLE III. SYSTEM ANALYSIS USING CCTV VIDEO IN THE THIRD SCENARIO (ALL OBJECT ARE IN FRAME)

		Predicted Label / Class (on Frame)																	
		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*
Original Label / Class (on Frame)	*1*	0,78																	
	2		0,72																
	3			0,91															
	4				0,92														
	5					0,82													
	6						0,93												
	7							0,87											
	8								0,67										
	9									0,78									
	10										0,75								
	11											0,86							
	12												0,70						
	13													0,83					
	14														0,88				
	15															0,77			
	16																0,75		

Based on the results of observing the cosine distance score, it can be concluded that the optimal cosine score is 0.75. Apart from determining the cosine score, adding a tracking algorithm can also resolve differences in recognition results for each frame. However, in this research, this has not been implemented.

The several images below are examples of model recognition results for each scenario:

Figure 12 A sample of the model recognition result for scenario 1 (student's face)

Figure 13 A sample of the model recognition result for scenario 1 (teacher's face)

In the two example images of the recognition results (Figures 12 and 13), it can be seen that the model is optimal enough to perform face recognition based on the classes that have been trained in the first test scenario (one trained object captured for each frame). The problem of errors in assigning bounding boxes in the first scenario can be overcome by increasing the minimum cosine distance score to 0.75.

Figure 14 A sample of the model recognition result for scenario 2

Figure 15 A sample of the model recognition result for scenario 2

Figures 14 and 15 illustrate the model recognition results for the second scenario, where each frame contains only two objects that have been trained using GhostFaceNets. The table II and the figures show that the model's performance is relatively stable in recognizing the faces, with only minor reductions in the cosine distance scores. The cosine distance score indicates the similarity between the facial features of the objects in the frame and the predicted classes by the model. The higher the score, the more confident the model is in recognizing the faces. The results suggest that the model can handle multiple faces in a frame with some compromising the accuracy and consistency of the recognition.

Figures 16 and 17 present the model recognition results for the third scenario, where each frame contains more than three objects that have been trained using GhostFaceNets. The results reveal that the model, which is considered the most optimal in this research, has a poor performance when dealing with multiple faces in a frame. The cosine distance scores, which indicate the similarity between the facial features of the objects in the frame and the predicted classes by the model, are significantly lower in this scenario than in the previous ones. This implies that the model has a low confidence and accuracy in recognizing the faces. This poses a challenge for applying the system to videos that contain more facial objects, especially those that have not been trained in the model creation process.

Figure 16 A sample of the model recognition result for 3rd scenario

Figure 17 A sample of the model recognition result for scenario 3

6. CONCLUSSION

This research focuses on building a GhostFacenets model using a custom dataset applied to low-resolution CCTV cameras. The results of this study show that the trained model can handle up to 5 objects in the real-time frame, but problems arise when there are too many objects in the frame. Low resolution is a challenge in optimizing face recognition models.

The research findings indicate significant improvements in the GhostFaceNets architecture for processing low-resolution images with multiple face detection in real-time conditions. This is achieved by eliminating the attention branch of the DFC, which initially causes loss of data representation. In addition, optimizing the number of embedding layers also, incorporating L2 regularization, and adopting PReLU as the activation function plays an important role in improving efficiency. This modification proved effective in training the model on custom datasets, allowing it to handle multiple objects in real-time with increased accuracy and efficiency.

In the future, it is considered that the problems in this research can be solved by integrating GhostFaceNets with a tracking algorithm. Where changes in the cosine distance score for each frame will not affect the recognition results. Apart from that, adding more datasets can also be an additional solution. Where increasing the number of faces trained can be the basis for the model to recognize each face better, this is due to the richness of features that can be calculated by the model to avoid recognition process errors, especially for objects that have not been trained.

ACKNOWLEDGMENT

The author(s) would like to express their sincere gratitude to the Makassar City Regional Research and Inovation Agency for their generous financial support through the research grant program. This research benefitted greatly from the facilities and support provided by SD Telkom Makassar, whose contribution to this

project was invaluable. Special thanks are also extended to the Artificial Intelligence and Multimedia Processing (AIMP) Research Group for their guidance and invaluable insights, which significantly contributed to the success of this study. The collaboration and support of these organizations have been instrumental in achieving the goals of this research..

REFERENCES

[1] redaksi, "Di Balik Tingginya Kriminalitas Remaja di Sulsel," *Mediasulsel.com*, Jan. 31, 2023. https://www.mediasulsel.com/di-balik-tingginya-kriminalitas-remaja-di-sulsel/ (accessed Sep. 12, 2023).

[2] M. O. Oloyede, G. P. Hancke, and H. C. Myburgh, "A review on face recognition systems: recent approaches and challenges," *Multimed Tools Appl*, vol. 79, no. 37–38, pp. 27891–27922, Oct. 2020, doi: 10.1007/s11042-020-09261-2.

[3] D. Salama AbdELminaam, A. M. Almansori, M. Taha, and E. Badr, "A deep facial recognition system using computational intelligent algorithms," *PLoS ONE*, vol. 15, no. 12, p. e0242269, Dec. 2020, doi: 10.1371/journal.pone.0242269.

[4] L. Li, X. Mu, S. Li, and H. Peng, "A Review of Face Recognition Technology," *IEEE Access*, vol. 8, pp. 139110–139120, 2020, doi: 10.1109/ACCESS.2020.3011028.

[5] M. Alansari, O. A. Hay, S. Javed, A. Shoufan, Y. Zweiri, and N. Werghi, "GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations," *IEEE Access*, vol. 11, pp. 35429–35446, 2023, doi: 10.1109/ACCESS.2023.3266068.

[6] J. Deng, J. Guo, D. Zhang, Y. Deng, X. Lu, and S. Shi, "Lightweight Face Recognition Challenge," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 2638–2646. doi: 10.1109/ICCVW.2019.00322.

[7] Y. Deng, "Deep learning on mobile devices: a review," in *Mobile Multimedia/Image Processing, Security, and Applications 2019*, S. S. Agaian, S. P. DelMarco, and V. K. Asari, Eds., Baltimore, United States: SPIE, May 2019, p. 11. doi: 10.1117/12.2518469.

[8] Y. Martínez-Díaz *et al.*, "Benchmarking lightweight face architectures on specific face recognition scenarios," *Artif Intell Rev*, vol. 54, no. 8, pp. 6201–6244, Dec. 2021, doi: 10.1007/s10462-021-09974-2.

[9] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition".

[10] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper, "ElasticFace: Elastic Margin Loss for Deep Face Recognition," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 1577–1586. doi: 10.1109/CVPRW56347.2022.00164.

[11] F. Boutros, P. Siebke, M. Klemt, N. Damer, F. Kirchbuchner, and A. Kuijper, "PocketNet: Extreme Lightweight Face Recognition Network using Neural Architecture Search and Multi-Step Knowledge Distillation." arXiv, Dec. 13, 2021. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/2108.10710

[12] F. Boutros, N. Damer, and A. Kuijper, "QuantFace: Towards Lightweight Face Recognition by Synthetic Data Low-bit Quantization." arXiv, Jun. 21, 2022. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/2206.10526

[13] M. Iliadis, H. Wang, R. Molina, and A. K. Katsaggelos, "Robust and Low-Rank Representation for Fast Face Identification with Occlusions," *IEEE Trans. on Image Process.*, vol. 26, no. 5, pp. 2203–2218, May 2017, doi: 10.1109/TIP.2017.2675206.

[14] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size." arXiv, Nov. 04, 2016. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/1602.07360

[15] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.

[16] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices." arXiv, Dec. 07, 2017. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/1707.01083

[17] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11218. Cham: Springer International Publishing, 2018, pp. 122–138. doi: 10.1007/978-3-030-01264-9_8.

[18] Q. Zhang *et al.*, "VarGNet: Variable Group Convolutional Neural Network for Efficient Embedded Computing." arXiv, Apr. 29, 2020. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/1907.05653

[19] M. Tan and Q. V. Le, "MixConv: Mixed Depthwise Convolutional Kernels." arXiv, Dec. 01, 2019. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/1907.09595

[20] S. Chen, Y. Liu, X. Gao, and Z. Han, "MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices," in *Biometric Recognition*, J. Zhou, Y. Wang, Z. Sun, Z. Jia, J. Feng, S. Shan, K. Ubul, and Z. Guo, Eds., in Lecture Notes in Computer Science, vol. 10996. Cham: Springer International Publishing, 2018, pp. 428–438. doi: 10.1007/978-3-319-97909-0_46.

[21] Y. Martindez-Diaz, L. S. Luevano, H. Mendez-Vazquez, M. Nicolas-Diaz, L. Chang, and M. Gonzalez-Mendoza, "ShuffleFaceNet: A Lightweight Face Architecture for Efficient and Highly-Accurate Face Recognition," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South): IEEE, Oct. 2019, pp. 2721–2728. doi: 10.1109/ICCVW.2019.00333.

[22] M. Yan, M. Zhao, Z. Xu, Q. Zhang, G. Wang, and Z. Su, "VarGFaceNet: An Efficient Variable Group Convolutional Neural Network for Lightweight Face Recognition." arXiv, Nov. 24, 2019. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/1910.04985

[23] F. Boutros, N. Damer, M. Fang, F. Kirchbuchner, and A. Kuijper, "MixFaceNets: Extremely Efficient Face Recognition Networks." arXiv, Jul. 27, 2021. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/2107.13046

[24] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features from Cheap Operations." arXiv, Mar. 13, 2020. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/1911.11907

[25] Y. Tang, K. Han, J. Guo, C. Xu, C. Xu, and Y. Wang, "GhostNetV2: Enhance Cheap Operation with Long-Range Attention." arXiv, Nov. 23, 2022. Accessed: Sep. 12, 2023. [Online]. Available: http://arxiv.org/abs/2211.12905

[26] R. Muttaqin, N. -, S. Fuada, and E. Mulyana, "Attendance System using Machine Learning-based Face Detection for Meeting Room Application," *IJACSA*, vol. 11, no. 8, 2020, doi: 10.14569/IJACSA.2020.0110837.

[27] N. S. Irjanto and N. Surantha, "Home Security System with Face Recognition based on Convolutional Neural Network," *IJACSA*, vol. 11, no. 11, 2020, doi: 10.14569/IJACSA.2020.0111152.

[28] I. Ahmad, F. AlQurashi, E. Abozinadah, and R. Mehmood, "A Novel Deep Learning-based Online Proctoring System using Face Recognition, Eye Blinking, and Object Detection Techniques," *IJACSA*, vol. 12, no. 10, 2021, doi: 10.14569/IJACSA.2021.0121094.

[29] J. Alamri, R. Harrabi, and S. Ben, "Face Recognition based on Convolution Neural Network and Scale Invariant Feature

Transform," *IJACSA*, vol. 12, no. 2, 2021, doi: 10.14569/IJACSA.2021.0120281.
[30] M. Ishtiaq, S. H., R. Amin, M. A., and H. Aldabbas, "Deep Learning based Intelligent Surveillance System," *IJACSA*, vol. 11, no. 4, 2020, doi: 10.14569/IJACSA.2020.0110479.

Indrabayu was born on July 16, 1975 in Makassar, Indonesia. He was awarded Summa Cum Laude from the Doctoral degree in Artificial Engineering in Civil Application from Hasanuddin University, Makassar, Indonesia, in 2013. Also received M.E degree in multimedia and communication from Institut Teknologi on 10 November, Surabaya, Indonesia in 2005. Currently, he is a Professor at the Department of Informatics, Universitas Hasanuddin. His research interest includes artificial intelligence and multimedia processing.

Andi Bukti Djufrie received a master's degree in urban planning science from Hasanuddin University and a bachelor's degree in agricultural science. Currently, he holds the position of head of the agency and regional research and innovation for the City of Macassar. His bureaucratic experience is undeniable; he has held numerous high-ranking positions and is currently pursuing his doctoral studies at Hasanuddin University. For correspondence, please use the email account bukti.djufrie@gmail.com

Muhammad Amri Akbar received a doctorate in environmental science from Brawijaya University and a master's degree in regional development planning from Hasanuddin University, currently serving at the Makassar City Government's Regional Research and Innovation Agency as head of innovation and technology development, previously working at the Makassar City Regional Development Planning Agency as head of the Socio-Cultural and General Government Division. His correspondence account is amriakbar72@gmail.com. Interested in environmental, socio-cultural and technological innovation issues.

Budi Armansyah holds a Master's and Bachelor's degrees in Public Administration from Hasanuddin University, currently serves at the Regional Research and Innovation Agency of the Makassar City Government as a researcher in the field of innovation and technological development, and is also a lecturer at STIA Al Gazali Barru. His correspondence account is budi.armansyah047@gmail.com

Muhammad Fadhil Bin Bahrunnida Muhammad Fadhil Bin Bahrunnida (Fadhil), Graduating from Hasanuddin University with a Bachelor of Engineering in Informatics Engineering, has been instrumental in developing intelligent system-based software targeting regional challenges in Makassar since 2021. The current focus lies in advancing a career and research that integrates artificial intelligence with digital security. Correspondence is encouraged through the provided professional account at fdl.mks97@gmail.com.

Nublan Azqalani was born in Ujung Pandang, South Sulawesi, Indonesia, on May 8, 1999. He attended Hasanuddin University, where he earned a bachelor's degree in informatics. He joins member of Artificial Intelligence and Multimedia Processing (AIMP) Research Group during his studies. His research interests include Intelligent Transportation Systems and Computer Vision. He can be contacted at email: nublan.azqalani@gmail.com