

LOW RESOLUTION FACE RECOGNITION ON CCTV IMAGES USING A COMBINATION OF SUPER RESOLUTION AND FACE RECOGNITION MODELS

ANTONIUS RILDO PRAMUDYA GONDOSISWOJO¹, GEDE PUTRA KUSUMA²

^{1,2}Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina

Nusantara University, Jakarta, Indonesia, 11480

E-mail: ¹antonius.gondosiswojo@binus.ac.id, ²inegara@binus.edu

ABSTRACT

Closed-Circuit Television (CCTV) serves as an essential device in contemporary society due to its capacity to capture images in public spaces, thereby contributing to the suppression of crime rates. However, a prevalent issue encountered is the small size of the images detected by CCTV, measuring only 32 x 32 pixels, resulting in inadequate facial recognition due to visual blurriness. To address this challenge, researchers opt to enhance the image resolution using the Super Resolution (SR) method before subjecting it to Face Recognition (FR) technology. This combined approach is referred to as Low Resolution Face Recognition (LRFR). In this research, the investigators aim to identify the optimal combination of SR and FR models. The SR models utilized include U-Net, EDSR, and Bicubic Interpolation, while the FR models employed are ResNet50 and MobilenetV2. As a result, six combinations of SR and FR are derived. The dataset employed for this study is LFW (Labelled Faces in The Wild). Based on the evaluation results, the study concludes that the most effective combination of SR and FR models is U-Net and ResNet50, achieving an accuracy rate of 85%, precision of 87%, recall of 85%, and a processing time of 11.454 seconds. Additionally, this combination successfully enhances the image resolution from 32 x 32 pixels to 128 x 128 pixels.

Keywords: *CCTV Images, Super Resolution, Low Resolution, Face Recognition, LFW Dataset*

1. INTRODUCTION

In today's modern world, technology plays an indispensable role in daily life, permeating various aspects such as work, education, healthcare, and security [1]. One notable technology widely used in the field of security is closed-circuit television (CCTV), which has become a common feature in public spaces and even residential areas [2]. CCTV utilizes one or more video cameras to capture video and audio data, operating through closed transmission either wirelessly or via cables. Its primary purpose is to enhance security and reduce crime rates by monitoring and deterring criminal activities. Major cities like Beijing and London have implemented extensive networks of surveillance cameras to combat crime, with Beijing boasting over 470,000 surveillance cameras [3].

Researchers are now working towards developing a "smart CCTV" system that goes beyond the traditional role of simply recording images. The aim is to incorporate Face Recognition (FR) technology within the CCTV system, enabling

it to recognize and identify detected faces. This integration of FR technology into CCTV systems eliminates the need for additional devices and facilitates faster data processing [1].

However, a challenge arises due to the small size of the images captured by CCTV, typically measuring only 32 x 32 pixels. To overcome this limitation, researchers employ Super Resolution (SR) techniques to enhance the image resolution before applying FR technology. This combination, known as Low Resolution Face Recognition (LRFR), is utilized to ensure accurate facial recognition even with low-resolution images [1].

SR techniques involve the process of restoring high-resolution images from low-resolution inputs and finding applications in various real-world domains such as medical imaging, satellite imaging, surveillance, security, and astronomy. Notable SR models utilized in this context include U-Net and EDSR. U-Net is a modified and expanded fully convolutional network that yields precise segmentation results with limited training data. On the other hand, EDSR is a fully

convolutional neural network specifically designed for super-resolution tasks [1].

Moving forward, researchers plan to evaluate different combinations of SR and FR methods to determine the most accurate and efficient approach. The evaluation will be focused on the assessment of the accuracy levels of these combinations, as well as the processing speed, with the aim of identifying the optimal combination for low-resolution face recognition. [1].

2. RELATED WORKS

2.1. Previous works related to Super Resolution

In this research, various Super Resolution (SR) methods were investigated.

The first study by Lupitha and Santoso [4] utilized the VGG19 model and SRGAN method for SR image generation and classification. They combined VGG19 with SRGAN to produce enhanced SR images and employed a CNN VGG19 model for classification, achieving an accuracy of 91.5% on their custom dataset.

Another study by Yulianto [5] proposed a new model called CMU-Net (Cosine Magnitude U-Net) for SR. CMU-Net incorporated the CM (Cosine Magnitude) method into the U-Net model and employed a combination of BCE Loss, Cosine loss, and Magnitude loss. Yulianto achieved an accuracy of 99.17% on the LFW dataset using the CMU-Net model.

In a different approach, Zhang [6] developed the SOUP-GAN (Super-resolution Optimized Using Perceptual-tuned Generative Adversarial Network) model. This method involved data preprocessing, LR to SR transformation using Multi-scale SR network, and the application of 3D perceptual loss with GAN. The evaluation results demonstrated that the utilization of SOUP-GAN outperformed other models in medical imaging applications such as MRI and CT SCAN.

Felipe and Vanegas [7] focused on deep learning methods for Video Temporal Super-Resolution (VTSR). They compared various PyTorch models, including DeblurGANv2 and EDVR, using basic, reconstruction, and multilevel schemes. The VTSR method employed in their research achieved a 99% SSIM score on the huaweiRED dataset.

Furthermore, Chen [8] proposed the HAT (Hybrid Attention Transformer) model, which involved pretraining using the ImageNet dataset. HAT achieved the best results in terms of PSNR and SSIM for scale factors 2, 3, and 4. Pretraining on

ImageNet significantly contributed to the performance of the HAT model.

Saarimaki [9] conducted a comparative study on several SR models capable of reconstructing images with near-original quality. After testing various models with different datasets and scaling factors, the EDSR (Enhanced Deep Super Resolution) model was found to provide the best results on datasets such as set5, set14, urban100, DIV2K, and TAMPERE17.

Moving on to Makwana's research [10], the Super Resolution CNN (SRCNN) was proposed for transforming LR images into HR images. Two models, SRCNN and bicubic interpolation, were employed, and the accuracy of SRCNN was evaluated to be 97% using the DIV2K dataset.

Additionally, Abdulfattah [11] offered a solution using the DCSCN (Deep CNN with Skip Connection and Network In Network) model for image enhancement. They created their own dataset for training and evaluation, conducted pre-processing with feature extraction and reconstruction network, and performed testing. The evaluation results showed that the DCSCN model achieved a PSNR and SSIM value of 37.614dB / 0.9588 on a scale factor of 2 using the Set5 dataset.

Medeiros [12] focused on developing a framework to predict PSNR and SSIM results for the SRMD (Super-Resolution for Multiple Degradation) and SRGAN (Super-Resolution GAN) models. The framework incorporated SRMD and SRGAN as base models and employed loss functions such as content loss, adversarial loss, and residual loss. The researcher also introduced kernel, residual image frameworks, and multiple residual image frameworks. The evaluation results indicated that the existing models were already capable of detecting low-resolution images quite well.

Lastly, Sun and Chen [13] utilized the CAR (Content Adaptive Resampler) model for transforming HR images into LR images. They employed the ResamplerNet method for the transformation process. The evaluation results showed that CAR outperformed other methods on several datasets used in their study.

2.2. Previous works related to Face Recognition

After conducting a literature review, several studies were found to be relevant to the current research.

In the field of facial recognition (FR), significant advancements have been made in the state of the art. An interesting study conducted by Mazzia [14] proposed the use of the ArcFace method

in the latest FR system by combining the CASIA and VGGFace2 models and training them on the MS1MV2 dataset. The study involved face detection and feature extraction processes, and the results were highly impressive, achieving accuracies of 99.83% on the LFW dataset, 98.02% on the YTF dataset, 95.45% on the CALFW dataset, and 92.08% on the CPLFW dataset.

Another approach proposed by Abuzneid [15] combined BPNN (Back-propagation neural network), PCA (Principal Component Analysis), and LBPH (Local Binary Pattern Histogram) in FR. The method involved face registration with a minimalized cost function and deep face registration techniques. The results from BPNN were incorporated into the ResNet50 model (deep Residual Network). In testing using the LFW dataset, this method achieved an accuracy of 98.55% when using ResNet50.

Furthermore, a study by Farid Naufal and Ferdiana Kusuma [16] tested several CNN models such as MobileNetV2, VGG16, DenseNet201, and Xception. They used the softmax optimization function and the Adam optimizer in their experiments. The results showed that the Xception architecture achieved the highest accuracy of 98.8%, while MobileNetV2 achieved an accuracy of 98.1% with a computation time of approximately 4081 seconds.

In another research by Dhanny [17] CNN architecture was employed in FR. They utilized the Haar cascade method in the OpenCV library for face detection and built a CNN model using the Tensorflow library. In the testing phase involving a dataset of 90,000 images, this research achieved an accuracy of 97% with the processed dataset.

In a study by Akbar [18], a combination of pre-processing methods including histogram equalization (HE), discrete wavelet transforms (DWT), discrete cosine transform (DCT), and stationary wavelet transform (SWT) was proposed along with a DNN (deep neural network) model in FR. Testing using the AR Face database showed that the combination of these four methods resulted in an accuracy of 92.73%.

Meanwhile, Almatarneh and John [19] explored the use of the AlexNet model in FR by combining eight algorithms, including PCA, 2D-PCA, LDA, ICA, DCT, SVM, CNN, and AlexNet. They conducted testing by introducing noise to the dataset and achieved an accuracy of 92% using DCT with the ORL dataset.

A survey conducted by Saragih [20] over the past five years on the best CNN models in FR found that the AlexNet and GoogLeNet models provided

the highest accuracy levels. Their research showed accuracy levels ranging between 99.72% and 100% when using the CelebFaces dataset.

In another study by Ridha [21], the AlexNet model was used in creating a Muslim fashion catalog by employing the HAAR cascades classifier method in the pre-processing stage. Increased accuracy was observed when using resolutions of 720p and 1080p, with respective improvements of 1.01% and 1.69%.

Furthermore, Firdaus and Munir [22] combined the VGGFace architecture with ANN for classification in FR. This study employed the YOLOv4 model for face detection. The research yielded an accuracy of 79.58% with the processed dataset.

Lastly, Nowak-Trzos and Horzyk Cracow [23] compared the PCA, MLP, SVM, and ANN methods in FR. They found that MLP achieved the highest accuracy of 90% for 20 individuals and 70% for 40 individuals using the Chicago Face database. The CNN accuracy achieved was 75.4% using the LFW database with 20 individuals.

3. METHODOLOGY

3.1. Research Stages

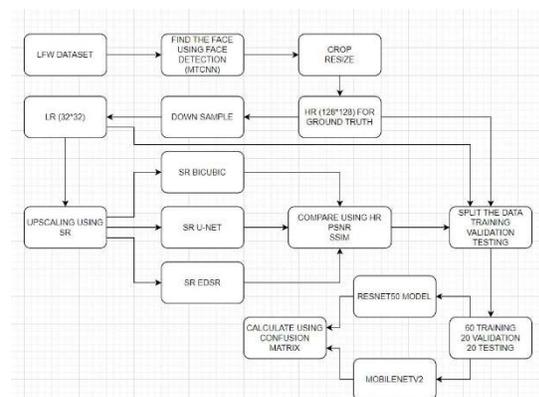


Figure 1. Research Stages

The research project involved several stages. Firstly, the Labeled Faces in the Wild (LFW) dataset was used, comprising various face images collected under diverse conditions, including pose variations, expressions, lighting, and backgrounds. Subsequently, face detection was carried out utilizing the Multi-Task Cascaded Convolutional Networks (MTCNN) method by Zhang [24]. MTCNN, recognized as a widely used technique for accurately detecting and identifying faces in images, made use of a cascaded convolutional neural

network to progressively identify and adjust bounding boxes around faces.

After successful face detection, the subsequent step involved the cropping of excessive background, focusing solely on the face region for analysis purposes. To ensure a consistent face image size, the faces were resized to 128 x 128 pixels [25] This adjustment aimed at obtaining high-resolution face images, which could then be uniformly processed in subsequent stages. Following this, the processed face images were reduced to 32 x 32 pixels [25], creating low-resolution face images tailored for the super-resolution stage.

In the super-resolution process, methods such as SR U-NET, SR EDSR, and SR BICUBIC [25] were employed to enhance the quality of low-resolution face images by restoring lost details. The evaluation of the super-resolution results encompassed a comparison of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) values [26]. PSNR quantified the restoration quality by contrasting the original and restored images, while SSIM assessed the structural similarity between them.

Following the image processing stage, the face dataset was divided into training, testing, and validation subsets, all of which played a pivotal role in training, testing, and evaluating the face recognition models through the use of unseen data during the training phase. Two distinct face recognition methods were subsequently employed to ascertain the accuracy, precision, and f1-score in face recognition. These methodologies underwent evaluation using the pre-divided training, testing, and validation subsets.

Lastly, in order to meticulously assess the performance of the face recognition system, a confusion matrix was used, facilitating the calculation of True Positive, True Negative, False Positive, and False Negative values [27]. From the outcomes of the confusion matrix, a range of evaluation metrics including accuracy, precision, recall, and f1-score could be computed. Hence, these research stages constituted a coherent sequence employed to execute face recognition utilizing the LFW dataset, incorporating a variety of techniques such as face detection, image preprocessing, super-resolution, and the evaluation of performance through the application of pertinent metrics.

3.2. Dataset

In research studies, researchers also utilized another dataset called LFW (Labelled Faces in The Wild), which is a public dataset consisting of 5,749

different labels or individuals, with a total of 13,233 face images at a resolution of 256 x 256 pixels. Figure 3 shows an example of each individual in the LFW dataset.

Each label name corresponds to a person's name separated by an underscore "_". The dataset contains 5,749 different labels or individuals. Out of the total dataset, 1,680 images have more than one image, while 4,069 images consist of only one facial image. The dataset can be downloaded from the website <http://vis-www.cs.umass.edu/lfw/>.

This dataset encompasses variations in pose, facial expressions, lighting conditions, age, gender, occlusion, and imbalanced class distribution. The images in this dataset are captured in conditions similar to real-world environments, making it an excellent reference for developing frameworks for face recognition algorithms. The usage of this dataset aims to minimize overfitting in the CNN classification model [28].



Figure 2. Sample dataset LFW (Labelled Faces in the Wild)

3.3. Data Pre-Processing Process

The researcher commenced by utilizing the LFW dataset, consisting of 13,233 images with 5,749 distinct classes. The subsequent step involved facial detection and cropping of facial regions within the LFW dataset. Furthermore, the researcher identified labels with more than 40 images.

3.3.1. Crop on the face area

In the cropping step for facial area, as shown in Figure 4, the aim is to reduce background variations in facial images with the expectation of improving classification accuracy. This cropping process is performed on the dataset before dividing it into training, validation, and testing data. One popular method for facial area cropping in face recognition

cases is using MTCNN (Multi-Task Cascaded Convolutional Networks) [24]. MTCNN is a Python program that serves as a downloadable library tool available at <https://pypi.org/project/torch-mtcnn/>.

By utilizing MTCNN, the facial area cropping process can be efficiently and accurately performed. This method assists in focusing on the relevant facial regions within the images while disregarding backgrounds that might influence classification. Thus, the usage of MTCNN as a tool for cropping facial areas becomes a crucial step in dataset preparation before proceeding to the training, validation, and testing stages of face recognition.

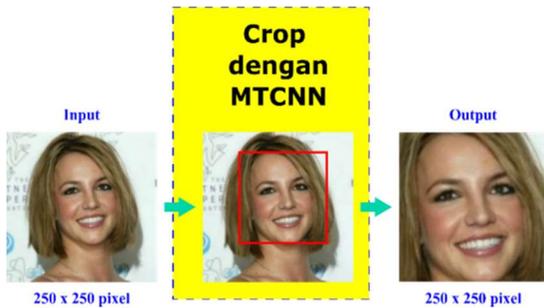


Figure 3. The MTCNN algorithm is used to precisely cut the face area in the dataset as a pre-processing step.

3.3.2. Re-processing Sort Dataset

The LFW dataset used in this research consists of 5,749 individual labels. However, manual selection of individuals was necessary because many face images only had one available image per individual. In this research process, a Python script was used to extract classes or individuals that had more than 40 face images. This was done because deep learning CNN classification algorithms require a large number of datasets for each class, individual, or label. After the selection process, 19 classes or individuals that met the criteria were obtained, as detailed in Table 2. Furthermore, each individual label had to be manually checked by inspecting each image to ensure that the face area cropping using MTCNN was done correctly. The total number of selected individuals, as seen in Table 2, included 1,867 images.

The process of selecting individuals with more than 40 face images using Python programming allowed the researchers to obtain a sufficient dataset for analysis using deep-learning CNN algorithms. Additionally, each image was manually examined to ensure the success of the face area cropping process using MTCNN. These steps are crucial in ensuring the quality of the dataset used in this research. With careful selection and

verification, a total of 1,867 images that met the criteria for further analysis were obtained.

Table 1. Class, individual, or label selection results in the LFW dataset containing more than 40 facial images.

No	Label Name	Total Face Image
1	Ariel Sharon	503
2	Arnold Schwarzenegger	124
3	Colin Powell	111
4	Donald Rumsfeld	60
5	George W Bush	44
6	Gerhard Schroeder	142
7	Gloria Macapagal Arroyo	52
8	Hugo Chavez	106
9	Jacques Chirac	213
10	Jean Chretien	41
11	Jennifer Capriati	78
12	John Ashcroft	44
13	Junichiro Koizumi	58
14	Laura Bush	54
15	Lleyton Hewitt	42
16	Luiz Inacio Lula da Silva	48
17	Serena Williams	42
18	Tony Blair	54
19	Vladimir Putin	51
Total		1867

3.4. Low Resolution Face recognition method

LRFR (Low-Resolution Face Recognition) is a method used to detect faces in small images. It consists of two stages. The first stage involves enhancing the resolution of low-resolution images to a higher resolution. The second stage employs face recognition (FR) to detect whether a human object is present in the image or not. An example application of LRFR can be seen in Figure 2.



Figure 4. Examples of Using LRFR [29]

In this research, the aim is to find the best combination of LRFR methods for the current state.

This is influenced by the fact that LRFR is widely used in devices with low computational power, which affects both accuracy and speed. The proposed model for the super-resolution (SR) stage utilizes the U-Net [30] and EDSR models. These models have the highest values of PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index), which indicate that the generated images closely resemble the original ones [4]. For the face recognition (FR) stage, the researcher proposes the use of the ResNet50 [31] and MobileNetV2 [32] models. These models offer high accuracy even with low computational requirements, thus accelerating the face recognition process.

The research aims to find the best combination of these models in terms of accuracy and minimal computational cost. The suggested combinations can be found in Table 1. The evaluation of paired models can be performed using a confusion matrix, a widely used tool for analyzing the performance of classification models [26]. Additionally, the speed of the models will be taken into consideration to ensure that real-time face recognition capabilities are achieved on low-powered devices.

By combining the U-Net, EDSR, ResNet50, and MobileNetV2 models, the research aims to identify the optimal combination that achieves a balance between accuracy and computational efficiency. The overall performance of LRFR systems will be enhanced, particularly in applications where low computational power is a constraint.

Overall, this research aims to address the challenge of achieving accurate and computationally efficient face recognition in LRFR applications. By leveraging the U-Net [30], EDSR, ResNet50 [31], and MobileNetV2 [32] models, the study seeks to find the optimal combination that offers both high accuracy and low computational requirements, thereby enhancing the overall performance of LRFR systems.

Table 2. LRFR model combination

	Model LR 1 (U-Net)	Model LR 2 (EDSR)
Model FR (ResNet50)	U-Net + ResNet50	EDSR + ResNet50
Model FR (MobileNetV2)	U-Net + MobileNetV2	EDSR + MobileNetV2

3.5. Data Processing Process

The obtained results were subsequently resized to 128 x 128 pixels to achieve high resolution for testing high-resolution images alongside low-resolution images. After resizing, the researcher reduced the image resolution to 32 x 32 pixels. Following the downsampling of the entire dataset, three super-resolution (SR) techniques were applied to enhance the image resolution back to 128 x 128 pixels. The researcher employed two face recognition (FR) algorithms to develop a face recognition system. The FR process was performed on both high-resolution (128 x 128 pixels) images, whether they were enhanced or original, and low-resolution (32 x 32 pixels) images.

In summary, the LFW dataset, after being sorted based on image size, was resized to 128 x 128 pixels (HR). Subsequently, the resolution of the dataset was reduced to 32 x 32 pixels, and three SR techniques were used to improve the image resolution back to 128 x 128 pixels. Then, two FR algorithms were employed to construct a face recognition system, with evaluations conducted on both HR and LR (low-resolution) images.

3.5.1. Decreasing Resolution for HR and LR Builds

This section will explain how HR images are created. The process involves the entire LFW dataset being gathered, and the dataset being read. Face detection using MTCNN is then applied to identify the faces in the images. The dataset is subsequently classified based on its resolution. Once the images are read, they are standardized to a resolution of 128 x 128 pixels. The resolution of the images is then reduced by the researcher to train the U-Net, EDSR, and Bicubic interpolation models. The resolution is downsampled from 128 x 128 pixels to 32 x 32 pixels.

3.5.2. Dataset Setup Process for FR

In this section, the researchers divided the dataset that had been upsampled using SR. The researchers performed a data shuffling process on the selected dataset in Table 2, and then grouped it into three parts: training data, comprising 80% of the dataset; validation data, comprising 10%; and testing data, also comprising 10%. This data-splitting method follows the approach used by Rai [33]. The detailed number of images in each data group can be found in Table 3. According to Table 3, there are

1,437 images for training, 129 images for validation, and 120 images for testing.

Table 3. The results of splitting the LFW dataset into groups of train, validation, and test datasets.

No	Label Name / Class / Individual	Total Face Image		
		Train	Val	Test
1	Ariel Sharon	426	47	30
2	Arnold Schwarzenegger	60	7	57
3	Colin Powell	96	11	4
4	Donald Rumsfeld	41	5	14
5	George W Bush	35	4	5
6	Gerhard Schroeder	119	13	10
7	Gloria Macapagal Arroyo	36	4	12
8	Hugo Chavez	91	10	5
9	Jacques Chirac	185	21	7
10	Jean Chretien	30	3	8
11	Jennifer Capriati	62	7	9
12	John Ashcroft	37	4	3
13	Junichiro Koizumi	50	6	3
14	Laura Bush	45	5	4
15	Lleyton Hewitt	36	4	2
16	Lui Inacio Lula da Silva	41	5	2
17	Serena Williams	32	4	6
18	Tony Blair	44	5	5
19	Vladimir Putin	45	5	1
Total		1512	168	187

The second step after the dataset separation process is setting several parameters during training. The parameters adjusted are the optimizer, learning rate, batch size, and epoch. The researchers used the Adam optimizer as the optimizer parameter. They set the learning rate parameter to 0.001. The batch size parameter was set to 32. Lastly, the researchers set the number of epochs to 100.

3.6. Performance Measurement

In this section, the researcher evaluates each model, both SR (Super-Resolution) and FR (Face Recognition). In the measurement process, the researcher employs several methods. For SR measurement, the researcher utilizes the PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) measurement methods. For FR measurement, the researcher employs the confusion matrix method.

3.6.1. SR Performance Measurement

The use of Super-Resolution (SR) is aimed at enhancing facial images, which are then interpolated

or up sampled to a resolution of 128 x 128 pixels. However, despite the image being in 128 x 128 pixels resolution, the image quality is still blurry. Before proceeding to the Face Recognition (FR) process in this thesis, measurements are taken from the SR results using two parameters: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). PSNR is a method for evaluating the similarity between the interpolated SR images and the high-quality HR reference images, or in other words, it measures the quality score of the image pair consisting of the low-resolution (LR) and high-resolution (HR) images. SSIM is a method used to measure the level of similarity between the SR images and the HR images.

3.6.1.1. Metric PSNR (Peak Signal to Noise Ratio)

PSNR (Peak Signal to Noise Ratio) is a calculation used to measure the similarity between an image that has been reconstructed using super-resolution GAN and the original high-resolution image. It is a scale for measuring image quality that focuses on the maximum signal strength ratio possible compared to the image distortion caused by noise. A higher PSNR score indicates better super-resolution image quality [34]. The equation for PSNR can be seen in Equation 1.

$$PSNR(y, \hat{y}) = 20 \log_{10} \left(\frac{L}{\sqrt{MSE(y, \hat{y})}} \right) [dB] \quad (1)$$

when,

$$MSE(y, \hat{y}) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (y_{ij} - \hat{y}_{ij})^2$$

According to the Grm method [35], L represents the maximum pixel value. For example, the maximum value for an 8-bit image is 255. Furthermore, y represents the original image, and \hat{y} represents the image reconstructed with super-resolution. The variables M and N represent the resolutions of the two compared images, and their resolutions should be the same [26]. A higher PSNR value indicates a closer resemblance of the super-resolution reconstructed image to the original or ground truth image. The measurement method of PSNR is based on human visual perception.

3.6.1.2. Metric SSIM (Structural Similarity Index Measurement)

SSIM (Structural Similarity Index) is a measurement used to assess the quality of an image by comparing it to a reference image [26]. SSIM measures the similarity between two images based on three parameters: luminance, contrast, and structure. To calculate the difference in luminance between the two images, denoted as image x and image y or $l(x, y)$, which are the values of μ_x and μ_y then SSIM computes the average intensity values of each image being compared.

$$SSIM(y, \hat{y}) = \frac{1}{M} \sum_{i=1}^M s(y, \hat{y}) \tag{2}$$

$$s(y, \hat{y}) = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)}$$

$$\mu_x = \left(\frac{1}{N} \sum_{i=1}^N x_i\right) - \sum_{i=1}^N x_i = 0 \tag{3}$$

The comparison between the luminance values of the two images x and y, or $l(x, y)$ can be expressed as follows:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{4}$$

When,

$$C_1 = (K_1L)^2$$

Where C_1 is a constant value to prevent division by zero. L is a free parameter ranging from 1 to 255 for 8-bit grayscale images, and $K_1 \ll 1$ is a very small constant value.

To calculate the difference in signal contrast $c(x,y)$ between the two images x and y, the standard deviations σ_x and σ_y are required.

$$\sigma_x = \left(\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2\right)^{\frac{1}{2}} \tag{5}$$

Thus, the difference in signal contrast $c(x,y)$ can be calculated using the following equation:

$$c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{6}$$

To compute the difference in structure between the two images $s(x,y)$, normalization or division by the standard deviation is necessary:

$$\frac{(x - \mu_x)}{\sigma_x} \text{ and } \frac{(y - \mu_y)}{\sigma_y} \tag{7}$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3}$$

When,

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \tag{8}$$

3.6.2. FR evaluation

To calculate the accuracy of face image classification using a CNN, the prediction results are tabulated into a confusion matrix (Table 4). The confusion matrix can be used as a reference when calculating accuracy in a dataset with various categories or labels [36].

Table 4. Classification Performance Measurement using Confusion Matrix Tabulation

		Actual			Total	Precision
		Class	A	B		
Predicted	A	TAA	FAB	FAC	A'	
	B	FBA	TBB	FBC	B'	
	C	FCA	FCB	TCC	C'	
Total		A''	B''	C''	Total	
Recall						

The columns A, B, and C in the Table represent the actual classes or labels, while the rows A, B, and C represent the predicted results. A", B", and C" denote the correct predictions with the prefix T, and the number of incorrectly predicted data with the prefix F. By mapping the classification results to the confusion table, the matrix is then used to calculate accuracy, precision, recall, and F1-score (Lv et al., 2018). Columns A, B, C in Table 4 represent the actual classes or labels, while the rows A, B, C

represent the predicted results. A", B", and C" denote the correct predictions with the prefix T, and the number of incorrectly predicted data with the prefix F. By mapping the classification results to the confusion table, the matrix is then used to calculate accuracy, precision, recall, and F1-score (Lv et al., 2018).

$$Accuracy = \frac{TAA + TBB + TCC}{Total} \quad (9)$$

In the evaluation method of classification results using confusion matrix tabulation, besides measuring the overall accuracy of the classification, it can also be used to calculate accuracy rates for each class or column, such as precision and recall (Lv et al., 2018).

- Calculating precision for each label or class

Precision, as defined in equation 2.12, is used to calculate the data extracted based on misclassified prediction information in each column of the confusion matrix (Lv et al., 2018).

$$PA = \frac{TAA}{A'} \quad (10)$$

When;

$$A' = TAA + FAB + FAC$$

- PA : Precision
- TAA : True Positive
- FAB : False Positive
- FAC : False Positive

- Calculating recall for each label or class

Recall or sensitivity, as defined in equation 2.20, is used to calculate the misclassified results in each column of the confusion matrix.

$$F1 A = 2 \times \frac{PA \times RA}{PA + RA} \quad (10)$$

3.6.3. Evaluation of Results of the Combination of SR and FR

In this section, the best combination of SR and FR in terms of speed will be determined by the researcher. One hundred LR images will be taken and processed through the SR method to measure the time taken by the model to enhance the resolution. Once the resolution is improved, the results will be

input into the FR method to measure the time taken by the model to identify faces. After the process is completed, all the recorded times will be summed up by the researcher. Lastly, the experiment will be repeated 10 times, and the results will be averaged by the researcher. The recording of each time will be indicated in Table 5.

Table 5. Image Classification Evaluation of Super Resolution Results Mapped into Confusion Matrix.

Experiment to	Processing Speed per Image (ms)		
	SR	FR	Combination
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
Average			

4. RESULT AND DISCUSSION

4.1. Super Resolution Evaluation Results

In this section, the researchers present the results of each Super-Resolution (SR) method using the LFW dataset.

4.1.1. U-Net

In this section, the researchers conducted image quality measurements. The results of the image measurements are presented in Table 6, which reveals the distance difference between the 128 x 128-pixel LR image, obtained through interpolation from the 32 x 32-pixel image using U-Net, and the original 128 x 128-pixel HR image. In this test, the U-Net GAN achieved a PSNR score of 31.04 dB and an SSIM score of 0.94 dB. The LR to SR image conversion process took 98,000 ms.

Table 6. Measurement of the performance of the LFW test dataset image interpolated with U-Net to a resolution of 128 x 128 pixels against the original 128 x 128-pixel HR image.

PSNR	SSIM	Conversion Time per Image (ms)
31,04	0,94	98,000

4.1.2. EDSR

In this section, the researchers conducted image quality measurements. The results of the image measurements are presented in Table 7, which shows the distance difference between the 128 x 128-pixel LR image, obtained through interpolation from the 32 x 32-pixel image using EDSR, and the original 128 x 128-pixel HR image. In this test, EDSR achieved a PSNR score of 31.07 dB and an SSIM score of 0.93 dB. The LR to SR image conversion process took 596,000 Ms seconds (9 minutes and 56 seconds).

Table 7. Measurement of the performance of the LFW test dataset image interpolated with EDSR to a resolution of 128 x 128 pixels against the original 128 x 128-pixel HR image.

PSNR	SSIM	Conversion Time per Image (ms)
31,07	0,93	596,000

4.1.3. Bicubic Interpolation

In this section, the researchers conducted image quality measurements. The results of the image measurements are presented in Table 8, which shows the distance difference between the 128 x 128-pixel LR image, obtained through interpolation from the 32 x 32-pixel image using bicubic interpolation, and the original 128 x 128-pixel HR image. In this test, bicubic interpolation achieved a PSNR score of 30.91 dB and an SSIM score of 0.94 dB. The LR to SR image conversion process took 1391 ms.

Table 8. Measurement of the performance of the LFW test dataset image interpolated with bicubic interpolation to a resolution of 128 x 128 pixels against the original 128 x 128-pixel HR image.

PSNR	SSIM	Conversion Time per Image (ms)
31,91	0,94	1,391

4.2. Discussion of Super Resolution Results

In this section, the researcher discusses the research findings. As shown in Table 35, the summary of PSNR and SSIM for each super-resolution (SR) method is presented. It can be

observed that the EDSR achieves slightly higher PSNR results compared to other SR methods. However, it is also evident that the computational time for EDSR is longer than the others. From these results, it is also found that in terms of speed, Bicubic Interpolation yields the best performance. Furthermore, the data reveals that SR Bicubic Interpolation provides higher SSIM results compared to SR U-Net and EDSR.

Table 9. Summary of Comparison of PSNR and SSIM in Each SR

No	Super Resolution	PSNR (db)	SSIM (db)	Conversion Time per Image (ms)
1	U-Net	31,04	0,94	98.000
2	EDSR	31,07	0,93	633.000
3	Bicubic interpolation	30,91	0,94	298

4.3. Face Recognition Evaluation Results

In this section, the researchers present the results of each facial recognition (FR) using both the combination of super resolution (SR) and without using SR.

4.3.1. ResNet50

4.2.1.1. ResNet50 with High-Resolution Images

From the training and testing results, an accuracy of 85%, precision of 87%, and recall of 84% can be observed. The training results can be seen in Table 9. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 10.

Table 10. ResNet50 Model Results Using Original Dataset.

Accuracy	Precision	Recall
85%	87%	84%

Table 11. Precision, Recall, and F1 score results from Original Dataset Image Results and Classified with ResNet50.

Class	Recall	Precision	f1 Score
0	82,35%	87,50%	84,85%
1	71,43%	100,00%	83,33%
2	91,49%	91,49%	91,49%
3	85,71%	75,00%	80,00%
4	96,77%	85,71%	90,91%
5	83,33%	89,29%	86,21%
6	62,50%	71,43%	66,67%
7	71,43%	55,56%	62,50%
8	80,00%	100,00%	88,89%
9	70,59%	92,31%	80,00%
10	54,55%	85,71%	66,67%
11	73,33%	84,62%	78,57%
12	88,89%	72,73%	80,00%
13	75,00%	50,00%	60,00%
14	7,00%	77,78%	73,68%
15	75,00%	100,00%	85,71%
16	81,82%	100,00%	90,00%
17	96,00%	80,00%	87,27%
18	80,00%	100,00%	88,89%

4.2.1.2. ResNet50 with Low-Resolution Images

From the training and testing results, an accuracy of 61%, precision of 64%, and recall of 60% can be observed. The training results can be seen in Table 11. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 12.

Table 12. ResNet50 Model Results Using Low Resolution Dataset.

Accuracy	Precision	Recall
61%	64%	60%

Table 13. Results of Precision, Recall, and F1 scores from Low Resolution Dataset Image Results and Classified with ResNet50.

Class	Recall	Precision	f1 Score
0	47,06%	66,67%	55,17%
1	14,29%	66,67%	23,53%

Class	Recall	Precision	f1 Score
2	78,72%	62,71%	69,81%
3	64,49%	72,00%	67,92%
4	84,95%	65,83%	74,18%
5	40,00%	54,55%	46,15%
6	75,00%	85,71%	80,00%
7	28,57%	14,29%	19,05%
8	60,00%	54,55%	57,14%
9	35,29%	66,67%	46,15%
10	27,27%	725,00%	40,00%
11	40,00%	66,67%	50,00%
12	88,89%	66,67%	76,19%
13	50,00%	28,50%	36,36%
14	70,00%	87,50%	77,78%
15	25,00%	40,00%	30,77%
16	81,82%	90,00%	85,71%
17	64,00%	53,33%	58,18%
18	10,00%	14,29%	11,76%

4.2.1.3. ResNet50 with U-Net

From the training and testing results, an accuracy of 85%, precision of 87%, and recall of 85% can be observed. The training results can be seen in Table 13. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 14.

Table 14. ResNet50 Model Results Using U-Net.

Accuracy	Precision	Recall
85%	87%	85%

Table 15. Precision, Recall, and F1 score results from U-Net Image Results and Classification with ResNet50.

Class	Recall	Precision	f1 Score
0	76,47%	86,67%	81,25%
1	71,43%	100,00%	83,33%
2	91,49%	91,49%	91,49%
3	92,86%	92,22%	81,25%
4	97,85%	85,85%	91,46%
5	80,00%	92,31%	85,71%
6	75,00%	85,71%	80,00%
7	85,71%	54,55%	66,67%

Class	Recall	Precision	f1 Score
8	80,00%	80,00%	80,00%
9	52,94%	100,00%	69,23%
10	54,55%	100,00%	70,59%
11	73,33%	100,00%	84,62%
12	88,89%	88,89%	88,89%
13	75,00%	60,00%	66,67%
14	80,00%	88,89%	84,21%
15	87,50%	100,00%	93,33%
16	81,92%	100,00%	90,00%
17	96,00%	75,00%	84,21%
18	60,00%	66,67%	6316,00%

Class	Recall	Precision	f1 Score
14	80,00%	80,00%	80,00%
15	87,50%	100,00%	93,33%
16	90,91%	100,00%	95,24%
17	96,00%	85,71%	90,57%
18	70,00%	77,78%	73,68%

4.2.1.5. ResNet50 with Bicubic Interpolation

From the training and testing results, an accuracy of 41%, precision of 44%, and recall of 40% can be observed. The training results can be seen in Table 17. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 18.

Table 18. ResNet50 Model Results Using Bicubic Interpolation.

Accuracy	Precision	Recall
41%	44%	40%

Table 19. Precision, Recall, and F1 score results from Bicubic Interpolation Image Results and Classification with ResNet50.

Class	Recall	Precision	f1 Score
0	23,53%	30,77%	26,67%
1	21,43%	50,00%	30,00%
2	48,94%	47,92%	48,42%
3	42,86%	32,43%	36,92%
4	64,52%	51,28%	57,14%
5	26,67%	33,33%	29,63%
6	87,50%	77,78%	82,35%
7	14,29%	7,14%	9,52%
8	30,00%	25,00%	27,27%
9	29,41%	50,00%	37,04%
10	18,18%	66,67%	28,57%
11	13,33%	33,33%	19,05%
12	33,33%	33,33%	33,33%
13	50,00%	66,67%	57,14%
14	40,00%	50,00%	44,44%
15	25,00%	25,00%	25,00%
16	45,45%	35,71%	40,00%
17	36,00%	31,03%	33,33%
18	0,00%	0,00%	100,00%

4.2.1.4. ResNet50 with EDSR

From the training and testing results, an accuracy of 86%, precision of 87%, and recall of 85% can be observed. The training results can be seen in Table 15. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 16.

Table 16. ResNet50 Model Results Using EDSR.

Accuracy	Precision	Recall
86%	87%	85%

Table 17. Precision, Recall, and F1 score results from EDSR Image Results and Classification with ResNet50.

Class	Recall	Precision	f1 Score
0	82,35%	82,35%	82,35%
1	64,29%	100,00%	78,26%
2	91,49%	9348,00%	92,47%
3	89,29%	75,76%	81,97%
4	97,85%	85,85%	91,46%
5	83,33%	89,29%	86,21%
6	75,00%	85,71%	80,00%
7	71,43%	62,50%	66,67%
8	80,00%	100,00%	88,89%
9	76,47%	92,86%	83,87%
10	84,55%	100,00%	70,59%
11	80,00%	92,31%	85,71%
12	77,78%	70,00%	73,68%
13	75,00%	60,00%	66,67%

4.3.2. MobileNetV2

4.2.2.1. MobileNetV2 with High Resolution Images

From the training and testing results, an accuracy of 43%, precision of 46%, and recall of 41% can be observed. The training results can be seen in Table 19. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 20.

Table 20. MobileNetV2 Model Results Using Original Dataset.

Accuracy	Precision	Recall
43%	46%	41%

Table 21. Precision, Recall, and F1 score results from Original Dataset Image Results and Classified with MobileNetV2.

Class	Recall	Precision	f1 Score
0	82,35%	8,75%	84,85%
1	71,43%	100,00%	83,33%
2	91,49%	91,49%	91,49%
3	85,71%	75,00%	80,00%
4	96,77%	85,71%	90,91%
5	83,33%	89,29%	86,21%
6	62,50%	71,43%	66,67%
7	71,43%	55,56%	62,50%
8	80,00%	88,89%	88,89%
9	70,59%	92,31%	80,00%
10	54,55%	85,71%	66,67%
11	73,33%	84,21%	78,57%
12	88,89%	72,73%	80,00%
13	75,00%	50,00%	60,00%
14	70,00%	77,78%	73,64%
15	75,00%	100,00%	85,71%
16	81,82%	100,00%	90,00%
17	96,80%	80,00%	87,27%
18	0,00%	100,00%	88,89%

4.2.2.2. MobileNetV2 with Low Resolution Images

From the training and testing results, an accuracy of 22%, precision of 10%, and recall of 2%

can be observed. The training results can be seen in Table 21. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 22.

Table 22. MobileNetV2 Model Results Using Low Resolution Dataset.

Accuracy	Precision	Recall
22%	10%	2%

Table 23. Results of Precision, Recall, and F1 scores from Low Resolution Dataset Image Results and Classified with MobileNetV2.

Class	Recall	Precision	f1 Score
0	0,00%	0,00%	100,00%
1	0,00%	0,00%	100,00%
2	63,80%	16,67%	9,23%
3	0,00%	0,00%	1,00%
4	83,87%	26,09%	39,80%
5	33,33%	1,00%	5,00%
6	0,00%	0,00%	100,00%
7	0,00%	0,00%	100,00%
8	0,00%	100,00%	0,00%
9	0,00%	0,00%	100,00%
10	0,00%	0,00%	100,00%
11	0,00%	0,00%	100,00%
12	0,00%	0,00%	100,00%
13	0,00%	0,00%	100,00%
14	0,00%	100,00%	0,00%
15	0,00%	0,00%	100,00%
16	0,00%	0,00%	100,00%
17	4,00%	16,67%	6,45%
18	0,00%	0,00%	100,00%

4.2.2.3. MobileNetV2 with U-Net

From the training and testing results, an accuracy of 42%, precision of 45%, and recall of 37% can be observed. The training results can be seen in Table 23. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 24.

Table 24. MobileNetV2 Model Results Using U-Net.

Accuracy	Precision	Recall
42%	45%	37%

Table 25. Precision, Recall, and F1 score results from U-Net Image Results and Classification with MobileNetV2.

Class	Recall	Precision	f1 Score
0	47,06%	33,33%	39,02%
1	14,29%	28,57%	19,05%
2	44,68%	43,75%	44,21%
3	35,71%	58,82%	44,44%
4	75,27%	43,48%	55,12%
5	23,33%	63,64%	34,15%
6	87,50%	100,00%	93,33%
7	0,00%	0,00%	100,00%
8	100,00%	20,00%	13,33%
9	5,88%	50,00%	10,53%
10	9,09%	33,33%	14,29%
11	13,33%	20,00%	16,00%
12	33,33%	30,00%	31,58%
13	50,00%	100,00%	66,67%
14	40,00%	66,67%	50,00%
15	12,50%	20,00%	15,38%
16	36,36%	36,36%	36,36%
17	36,00%	40,91%	38,30%
18	30,00%	30,00%	30,00%

4.2.2.4. MobileNetV2 with EDSR

From the training and testing results, an accuracy of 42%, precision of 44%, and recall of 39% can be observed. The training results can be seen in Table 25. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 26.

Table 26. MobileNetV2 Model Results Using EDSR.

Accuracy	Precision	Recall
85%	44%	39%

Table 27. Precision, Recall, and F1 score results from EDSR Image Results and Classification with MobileNetV2.

Class	Recall	Precision	f1 Score
0	29,41%	35,71%	32,26%
1	14,29%	33,33%	20,00%
2	55,32%	40,62%	46,85%
3	25,00%	31,82%	28,00%
4	76,34%	50,35%	60,68%
5	33,33%	40,00%	36,36%
6	87,50%	77,78%	82,35%
7	0,00%	0,00%	100,00%
8	30,00%	25,00%	27,27%
9	11,36%	40,00%	18,18%
10	9,09%	50,00%	15,38%
11	13,33%	40,00%	20,00%
12	33,33%	37,50%	3529,00%
13	50,00%	66,67%	57,14%
14	40,00%	50,00%	44,44%
15	12,50%	20,00%	15,38%
16	36,36%	33,33%	34,78%
17	28,00%	33,33%	30,43%
18	0,00%	0,00%	100,00%

4.2.2.5. MobileNetV2 with Bicubic Interpolation

From the training and testing results, an accuracy of 41%, precision of 44%, and recall of 40% can be observed. The training results can be seen in Table 27. The function of the confusion matrix is to calculate accuracy, precision, recall, and F1 score based on the classes. The calculation results can be seen in Table 28.

Table 28. MobileNetV2 Model Results Using Bicubic Interpolation.

Accuracy	Precision	Recall
41%	4%	40%

Table 29. Precision, Recall, and F1 score results from Bicubic Interpolation Image Results and Classification with MobileNetV2.

Class	Recall	Precision	f1 Score
0	35,29%	33,33%	34,29%
1	7,14%	33,33%	11,76%

Class	Recall	Precision	f1 Score
2	53,19%	42,37%	47,17%
3	46,63%	39,39%	42,62%
4	58,06%	52,43%	55,10%
5	36,67%	33,33%	34,92%
6	87,50%	77,78%	82,35%
7	28,57%	16,67%	21,05%
8	20,00%	28,57%	23,53%
9	25,53%	40,00%	29,63%
10	18,18%	66,67%	28,57%
11	20,00%	42,86%	27,27%
12	33,33%	33,33%	33,33%
13	50,00%	100,00%	66,67%
14	30,00%	60,00%	40,00%
15	12,50%	16,67%	14,29%
16	54,55%	33,33%	41,38%
17	36,00%	30,00%	32,93%
18	0,00%	0,00%	100,00%

4.4. Computation Speed Evaluation Results

4.4.1. U-Net

4.4.1.1. U-Net with ResNet50

From the results of 10 experiments with 100 testing images, the average processing time for U-Net was 101.24 ms, for ResNet50 it was 13.3 ms, and for their combination it was 114.54 ms. The research findings can be seen in Table 29.

Table 30. U-Net Computing Speed Average Time with ResNet50.

Experiment to	Processing Speed per Image (ms)		
	SR	FR	Combination
	U-Net	ResNet50	
1	101,7	12	113,7
2	97,5	11	108,5
3	116,7	16	132,7
4	94	13	107
5	100	11	111
6	99,7	17	116,7
7	103,3	11	114,3
8	96,2	11	107,2
9	100	20	120
10	103,3	11	114,3
Average	101.24	13.3	114.54

4.4.1.2. U-Net with MobileNetV2

From the results of 10 experiments with 100 testing images, the average processing time for U-Net was 96.6 ms, for MobileNetV2 it was 10.88 ms, and for their combination it was 107.48 ms. The research findings can be seen in Table 30.

Table 31. Time Average Computing Speed of U-Net and MobileNetV2.

Experiment to	Processing Speed per Image (ms)		
	SR	FR	Combination
	U-Net	MobileNetV2	
1	97	8,7	105,7
2	93	9,7	102,7
3	98	8,5	106,5
4	98	14,4	112,4
5	100	9	109
6	92	9,7	101,7
7	100	13,5	113,5
8	97,5	17,2	114,7
9	98,5	9,1	107,6
10	92	9	101
Average	96.6	10.88	107.48

4.4.2. EDSR

4.4.2.1. EDSR with ResNet50

From the results of 10 experiments with 100 testing images, the average processing time for EDSR was 1,426 ms, for ResNet50 it was 13.62 ms, and for their combination it was 1,439.62 ms. The research findings can be seen in Table 31.

Table 32. Time Average Compute Speed of EDSR and ResNet50.

Experiment to	Processing Speed per Image (ms)		
	SR	FR	Combination
	EDSR	ResNet50	
1	1440	11,3	1451,3
2	1425	15,3	1440,3
3	1410	11,1	1421,1
4	1425	18,7	1443,7
5	1440	17,4	1457,4
6	1440	11,2	1451,2
7	1435	10,9	1445,9
8	1410	15,8	1425,8
9	1415	12,9	1427,9
10	1420	11,6	1431,6
Average	1,426	13.62	1,439,62

4.4.2.2. EDSR with MobileNetV2

From the results of 10 experiments with 100 testing images, the average processing time for EDSR was 1,410.2 ms, for MobileNetV2 it was 10.69 ms, and for their combination it was 1,420.89 ms. The research findings can be seen in Table 32.

Table 33. Time Average Computing Speed of EDSR and MobileNetV2.

Experiment to	Processing Speed per Image (ms)		
	SR	FR	Combination
	EDSR	MobileNet v2	
1	1430	9,5	1439,5
2	1420	9,6	1429,6
3	1420	9,5	1429,5
4	1400	9,3	1409,3
5	1412	21,9	1433,9
6	1400	9,6	1409,6
7	1385	9,4	1394,4
8	1405	9,8	1414,8
9	1425	9,3	1434,3
10	1405	9	1414
Average	1,410.2	10.69	1,420.89

4.4.3.2. Bicubic Interpretation with MobileNetV2

From the results of 10 experiments with 100 testing images, the average processing time for Bicubic Interpolation was 0.0255 ms, for MobileNetV2 it was 2.557 ms, and for their combination it was 2.5825 ms. The research findings can be seen in Table 34.

Table 35. Bicubic Interpolation and MobileNetV2 Average Computing Speed Time.

Experiment to	Processing Speed per Image (ms)		
	SR	FR	Combination
	Bicubic	MobileNet V2	
1	0.0225	8,73	8.7525
2	0.0236	2,03	2.0536
3	0.0311	2,08	2.1111
4	0.0239	1,37	1.3939
5	0.0261	2,1	2.1261
6	0.0326	2,11	2.1426
7	0.0211	2,03	2.0511
8	0.0235	1,3	1.3235
9	0.0205	1,47	1.4905
10	0.0301	2,35	2.3801
Average	0.0255	2,557	2.5825

4.4.3. Bicubic Interpretation

4.4.3.1. Bicubic Interpolation with ResNet50

From the results of 10 experiments with 100 testing images, the average processing time for Bicubic Interpolation was 0.1896 ms, for ResNet50 it was 3.189 ms, and for their combination it was 3.3786 ms. The research findings can be seen in Table 33.

Table 34. Bicubic Interpolation and ResNet50 Average Computing Speed Time.

Experiment to	Processing Speed per Image (ms)		
	SR	FR	Combination
	Bicubic	ResNet50	
1	0.163	3.64	3.803
2	0.207	3.68	3.887
3	0.179	2.59	2.769
4	0.2005	2.2	2.4005
5	0.1725	3.93	4.1025
6	0.1705	3.64	3.8105
7	0.172	2.33	2.502
8	0.178	2.45	2.628
9	0.2805	3.77	4.0505
10	0.173	3.66	3.833
Average	0.1896	3.189	3.3786

4.5. Discussion of Face Recognition Results

From the above results, it can be concluded that the best face recognition (FR) method is ResNet50 with an accuracy rate of 85%. The FR results obtained using ResNet50 also match the results obtained without reducing the resolution. A comparison table can be seen in Table 36.

Table 36. Comparison Table of Each FR Model

SR Model	FR Model	Accuracy	Precision	Recall
Original	ResNet50	85,00%	87,00%	84,00%
	MobileNetV2	43,00%	46,00%	41,00%
Low Res	ResNet50	61,00%	64,00%	60,00%
	MobileNetV2	22,00%	10,00%	2,00%
U-Net	ResNet50	85,00%	87,00%	85,00%
	MobileNetV2	42,00%	45,00%	37,00%
EDSR	ResNet50	86,00%	87,00%	85,00%
	MobileNetV2	42,00%	44,00%	39,00%
Bicubic	ResNet50	41,00%	44,00%	40,00%
	MobileNetV2	41,00%	44,00%	39,00%

5. CONCLUSION AND FUTURE WORKS

5.1. Conclusion

Based on the conducted research and the evaluation using PSNR and SSIM, it is concluded that the best-performing SR method is EDSR, with a PSNR value of 31.07 dB and an SSIM value of 0.93 dB. However, the drawback of EDSR is its longer conversion time compared to other SR models, which amounts to 633,000 ms. Furthermore, the best combination of SR and FR models is found to be U-Net and ResNet50, achieving an accuracy of 85%. In terms of processing speed, the combination with the fastest data processing is bicubic interpolation with MobileNetV2, with a recorded time of 2.5825 ms.

Nevertheless, it is important to note that this combination may not be the best in terms of accuracy. The limitations of the study include the omission of factors related to varying light intensities due to the study's specific focus on testing image quality with respect to the accuracy and speed of face recognition using different SRFR combinations. Additionally, the research primarily concentrated on ideal face images, without considering the impact of facial accessories, which might affect detection times. These findings highlight the trade-off between processing speed and accuracy in the context of super-resolution and face recognition, offering valuable insights for future research directions.

5.2. Future Works

Based on the results of this research, there is a tendency for suboptimal models that perform computation at a very slow pace. To optimize the models in terms of both accuracy and speed, it is suggested to consider model simplification strategies.

REFERENCES

- [1] A. Ahda, "Analisa Perbandingan Kinerja Cctv Dvr Dengan Cctv Portable Menggunakan Smartphone Android Secara Online," *Jurnal Perencanaan, Sians, Teknologi, Dan Komputer*, Vol. 1, Pp. 114–120, Dec. 2018.
- [2] A. Rohmadi, S. Tinggi, M. Informatika, D. Komputer, N. Mandiri, And (Stmik, "Monitoring Cctv Digital Secara Online Melalui Internet & Mobile Phone Pada Jaringan Wireless Lan: Studi Kasus Padapt Tiga Sinar Mandiri," *Jurnal Cki On Spot*, Vol. 9, No. 1, 2016.
- [3] A. M. Hasan, "Cctv: Antara Fungsi, Privasi, Dan Kontroversi," 2016, Accessed: Nov. 01, 2022. [Online]. Available: <https://Tirto.Id/Cctv-Antara-Fungsi-Privasi-Dan-Kontroversi-Bwmj>
- [4] M. Lupitha And H. Santoso, "Super Resolution Generative Adversarial Networks For Image Supervise Learning," *Sinkron*, Vol. 7, No. 2, Pp. 455–463, Apr. 2022, Doi: 10.33395/Sinkron.V7i2.11373.
- [5] Yulianto, "Penerapan Metode Super Resolusi Berbasis Generative Adversarial Networks Pada Pengenalan Citra Wajah Beresolusi Rendah," 2021.
- [6] K. Zhang *Et Al.*, "Soup-Gan: Super-Resolution Mri Using Generative Adversarial Networks," Jun. 2021, [Online]. Available: [Http://Arxiv.Org/Abs/2106.02599](http://Arxiv.Org/Abs/2106.02599)
- [7] G. Felipe And T. Vanegas, "A Deep Learning Framework For Video Temporal Super-Resolution," 2020.
- [8] X. Chen, X. Wang, J. Zhou, And C. Dong, "Activating More Pixels In Image Super-Resolution Transformer," May 2022, [Online]. Available: [Http://Arxiv.Org/Abs/2205.04437](http://Arxiv.Org/Abs/2205.04437)
- [9] A. Saarimäki, "Single Image Super-Resolution Using Convolutional Neural Networks," 2018.
- [10] Y. Makwana, "Single Image Super Resolution Using Deep Learning," 2020.
- [11] M. E. Abdulfattah, L. Novamizanti, And S. Rizal, "Super Resolution Pada Citra Udara Menggunakan Convolutional Neural Network," *Elkomika: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, & Teknik Elektronika*, Vol. 9, No. 1, P. 71, Jan. 2021, Doi: 10.26760/Elkomika.V9i1.71.
- [12] A. L. S. De Medeiros, "Deep Learning For Single Image Super-Resolution Using Residual Image Learning And Multiple Degradations," 2019.
- [13] W. Sun And Z. Chen, "Learned Image Downscaling For Upscaling Using Content Adaptive Resampler," Jul. 2019, Doi: 10.1109/Tip.2020.2970248.
- [14] V. Mazzia, F. Salvetti, M. Ahmed, And E. Awouda, "Face Recognition System For Service Robotics Applications With Deep Learning At The Edge," 2021.

- [15] M. Abuzneid, "Improving Human Face Recognition Using Deep Learning Based Image Registration And Multi-Classifer Approaches," 2018.
- [16] M. Farid Naufal And S. Ferdiana Kusuma, "Pendeteksi Citra Masker Wajah Menggunakan Cnn Dan Transfer Learning," Vol. 8, No. 6, Pp. 1293–1300, 2021, Doi: 10.25126/Jtiik.202185201.
- [17] S. Dhanny, D. P. Andikha, S. Kezia, And F. Jenisa, "Implementasi Convolutional Neural Network Untuk Facial Recognition Implementasi Covolutional Neural Network Untuk Facial Recognition," *Media Informatika*, Vol. 2, 2021.
- [18] A. L. Akbar, C. Fatichah, And A. Saikhu, "Face Recognition Using Deep Neural Networks With The Combination Of Discrete Wavelet Transform, Stationary Wavelet Transform, And Discrete Cosine Transform Methods," *Juti: Jurnal Ilmiah Teknologi Informasi*, Vol. 18, No. 2, P. 158, Jul. 2020, Doi: 10.12962/J24068535.V18i2.A1000.
- [19] A. Almatarneh And S. John', "Evaluation Of Face Recognition Algorithms Under Noise," 2019.
- [20] R. E. Saragih And Q. H. To, "A Survey Of Face Recognition Based On Convolutional Neural Network," 2022.
- [21] M. N. Ridha, "Identifikasi Foto Wanita Berhijab Dari Majalah Untuk Pembuatan Katalog Busana Muslim Otomatis Memanfaatkan Convolutional Neural Network," *Journal Of Intelligent Systems And Computation*, 2022.
- [22] F. Firdaus And R. Munir, "Masked Face Recognition Using Deep Learning Based On Unmasked Area," In *Icpc2t 2022 - 2nd International Conference On Power, Control And Computing Technologies, Proceedings*, Institute Of Electrical And Electronics Engineers Inc., 2022. Doi: 10.1109/Icpc2t53885.2022.9776651.
- [23] M. Nowak-Trzos And A. Horzyk Cracow, "A Comparative Analysis And Evaluation Of Various Machine Learning Algorithms For Facial Recognition Analiza Porównawcza Wybranych Algorytmów Uczenia Maszynowego Do Rozpoznawania Twarzy," 2018.
- [24] K. Zhang, Z. Zhang, Z. Li, And Y. Qiao, "Joint Face Detection And Alignment Using Multitask Cascaded Convolutional Networks," *Ieee Signal Process Lett*, Vol. 23, No. 10, Pp. 1499–1503, Oct. 2016, Doi: 10.1109/Lsp.2016.2603342.
- [25] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, And Y. Fu, "Image Super-Resolution Using Very Deep Residual Channel Attention Networks," In *European Conference On Computer Vision*, Jul. 2018. [Online]. Available: [Http://Arxiv.Org/Abs/1807.02758](http://arxiv.org/abs/1807.02758)
- [26] Z. Wang, A. C. Bovik, H. R. Sheikh, And E. P. Simoncelli, "Image Quality Assessment: From Error Visibility To Structural Similarity," *Ieee Transactions On Image Processing*, Vol. 13, No. 4, Pp. 600–612, Apr. 2004, Doi: 10.1109/Tip.2003.819861.
- [27] D. M. W. Powers And Ailab, "Evaluation: From Precision, Recall And F-Measure To Roc, Informedness, Markedness & Correlation," *Journal Of Machine Learning Technologies*, Vol. 2, No. 1, Pp. 37–63, 2011, Accessed: Jun. 28, 2023. [Online]. Available: [Https://Arxiv.Org/Abs/2010.16061](https://arxiv.org/abs/2010.16061)
- [28] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, And E. Learned-Miller, "Labeled Faces In The Wild: A Database Forstudying Face Recognition In Unconstrained Environments," 2008, Accessed: Nov. 04, 2022. [Online]. Available: [Https://Hal.Inria.Fr/Inria-00321923](https://hal.inria.fr/inria-00321923)
- [29] J. Thorpe, "Exclusive: Changing The Narrative Around Facial Recognition - International Security Journal (Isj)," 2020. [Https://Internationalsecurityjournal.Com/E xclusive-Facial-Recognition/](https://internationalsecurityjournal.com/exclusive-facial-recognition/) (Accessed Nov. 04, 2022).
- [30] O. Ronneberger, P. Fischer, And T. Brox, "U-Net: Convolutional Networks For Biomedical Image Segmentation," In *International Conference On Medical Image Computing And Computer-Assisted Intervention*, 2015, Pp. 234–241.
- [31] K. He, X. Zhang, S. Ren, And J. Sun, "Deep Residual Learning For Image Recognition," *Proceedings Of The Ieee Computer Society Conference On Computer Vision And Pattern Recognition*, Vol. 2016-December, Pp. 770–778, Dec. 2016, Doi: 10.1109/Cvpr.2016.90.
- [32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, And L.-C. Chen, "Mobilenetv2: Inverted Residuals And Linear Bottlenecks."

- [33] A. Rai, V. Chudasama, K. Upla, K. Raja, R. Ramachandra, And C. Busch, "Comsupresnet: A Compact Super-Resolution Network For Low-Resolution Face Images," In *2020 8th International Workshop On Biometrics And Forensics, Iwbf 2020 - Proceedings*, Institute Of Electrical And Electronics Engineers Inc., Apr. 2020. Doi: 10.1109/Iwbf49977.2020.9107946.
- [34] H. Mittal, A. C. Pandey, M. Saraswat, S. Kumar, R. Pal, And G. Modwel, "A Comprehensive Survey Of Image Segmentation: Clustering Methods, Performance Parameters, And Benchmark Datasets," *Multimed Tools Appl*, Vol. 81, No. 24, Pp. 35001–35026, Oct. 2021, Doi: 10.1007/S11042-021-10594-9/Tables/6.
- [35] K. Grm, W. J. Scheirer, And V. Štruc, "Face Hallucination Using Cascaded Super-Resolution And Identity Priors," *Ieee Transactions On Image Processing*, Vol. 29, No. 1, Pp. 2150–2165, 2020, Doi: 10.1109/Tip.2019.2945835.
- [36] M. Grandini, E. Bagli, And G. Visani, "Metrics For Multi-Class Classification: An Overview," Aug. 2020, [Online]. Available: [Http://Arxiv.Org/Abs/2008.05756](http://Arxiv.Org/Abs/2008.05756)