International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Improving Human Character Recognition Performance Based on Facial Images with the Addition of Channel Attention Module in ResNet50 Model

Nurul Khairina[1], Muhathir[1]*, Fadhillah Azmi[1], Ferawaty[2], Wenripin Chandra[2], Mega Puspita Sari[1], Tika Ermita Wulandari[1]

Submitted: 28/01/2024 **Revised**: 06/03/2024 **Accepted**: 14/03/2024

Abstract: This study investigates the performance of several convolutional neural network (CNN) architectures in classifying human characters based on facial images, considering the addition of the Channel Attention Module (CBAM) to the ResNet50 model. This research aims to evaluate and compare the capabilities of the ResNet50 model with and without the addition of CBAM, and to compare it with the EfficientNetB3 and GoogleNet architectures in recognizing human characters based on facial images. This research uses an experimental approach by utilizing a tagged facial image dataset. Accuracy, precision, recall, and F1-score metrics are used to quantitatively evaluate the performance of the models. The addition of CBAM to the ResNet50 model successfully improves its performance in classifying human characters, especially in identifying the Savory and Unsavory classes. ResNet50 with CBAM demonstrates higher accuracy compared to ResNet50 without CBAM, and outperforms EfficientNetB3 and GoogleNet. This research indicates that the addition of CBAM to the ResNet50 model can enhance accuracy in recognizing human characters based on facial images. These results provide valuable insights into the importance of integrating enrichment techniques into CNN architectures. However, this research has limitations in dataset variation and further research is needed with more varied datasets and additional experiments to understand the factors that affect model performance more deeply.

Keywords: CNN, CBAM, Human Characters, ResNet50

1. Introduction

The human face has become an important research subject in computer image processing [1]. Research on human faces covers various aspects, such as face detection [2], facial expression recognition [3][4], face shape analysis [5], and face-based human character identification [6]. All of this is an important part of the study of human character, which plays a crucial role in various applications such as security, identification [7], and human behaviour modelling [8]. In this context, the development of a classification model using a deep learning approach is becoming increasingly interesting.

A deep learning model is a machine learning algorithm consisting of multiple processing layers used to learn a hierarchical representation of data [9]. Various deep learning architectures have been developed, including Convolutional Neural Networks (CNNs) [10], Recurrent Neural Networks (RNNs) [11], and etc. In this study, we chose ResNet50 as the model to evaluate. ResNet50 is one of the most well-known CNN architectures and is proven to be effective [12] The reason for this is due to its ability to handle performance degradation issues when network depth is increased, as well as having a balance between reliability and computational efficiency [13].

Although ResNet50 has proven to be effective in various image processing tasks, it still has some weaknesses. One of the significant gaps is the lack of ability to effectively handle spatial relationships between features in the image [14]. This can result in the loss of important information and degrade the performance of the model. To overcome this weakness, this research aims to introduce The Convolutional Block Attention Module (CBAM) into the ResNet50 architecture. CBAM has been proven effective in improving model performance by accommodating spatial and channel-wise information [15].

Convolutional Block Attention Module (CBAM) is a mechanism that enables convolutional neural networks to adaptively select important channels and important locations in the image, improving the model's ability to capture relevant information [16]. CBAM has been integrated into various CNN architectures, including ResNet [17] [18], DenseNet [19], VGG [20] [21], MobileNetx[22] and etc. The application of CBAM to these models has been shown to improve performance in

[1]Universitas Medan Area, Medan, Indonesia
nurul@staff.uma.ac.id, muhathir@staff.uma.ac.id,*
fadhillah@staff.uma.ac.id, mega@staff.uma.ac.id,
tikaermita@staff.uma.ac.id
[2]Universitas Pelita Harapan, Medan, Indonesia
ferawaty.fik@uph.edu, wenripin@lecturer.uph.edu
** Corresponding Author Email: muhathir@staff.uma.ac.id*

various image processing tasks, such as object classification, segmentation, and others.

The gap of this study is to evaluate the effect of adding CBAM in the ResNet50 architecture on the performance of the model in classifying human characters based on facial images. Within the parameters of this research, the following specific objectives are stated:

a. Can the addition of The Convolutional Block Attention Module (CBAM) in ResNet50 architecture improve the accuracy of human character classification based on facial images?

b. What are the changes in the performance of the ResNet50 model after the addition of CBAM, especially in terms of precision, recall, and F1-score for each human character classification class?

2. Material & Method

2.1 Research Framework

The research framework of this human character recognition research can be seen in Figure 1 below.

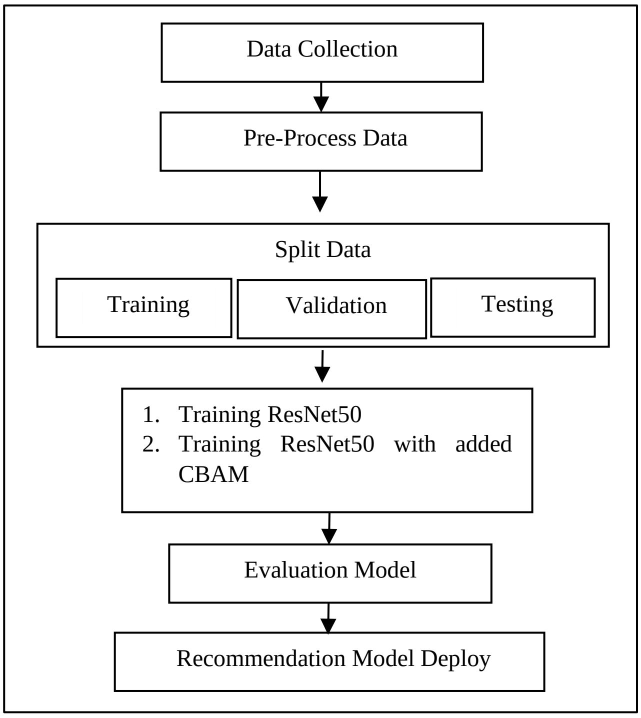


Fig. 1. Research Framework Model Utilized in This Study for Human Character Recognition

This research utilizes ResNet50 architecture as well as ResNet50 enhanced with CBAM (Convolutional Block Attention Module) for human character classification through facial images. Initially, the first step is data collection, followed by pre-processing to form a data set containing various facial images representing human characters. Next, the dataset is divided into three subsets: test, validation and training. The training subset is used to train the two models, ResNet50 and ResNet50 with CBAM. After training, the validation subset is used to evaluate the accuracy and performance of both models. The results of this evaluation will determine which model is the best in human character classification through facial images. Recommendations for the next stage of implementation: The model that gives the best results from the validation evaluation will be recommended for deployment in the next stage.

2.2 Data Collection and Pre-Processing Data

For this research, data was collected from various secondary data sources openly available on Kaggle. These data sources include various facial images that represent human characters in different situations. Each image was labelled as "savory" or "unsavory" based on the observed facial characteristics. A total of 12420 images were collected, of which 6210 were categorized as "savory" and 6210 as "unsavory". Data division was done by allocating 11200 images for model training, 600 images for validation, and 600 images for testing. The training data was used to train the model to understand the patterns in the data, while the validation dataset was used to evaluate the performance of the model and prevent overfitting. The testing dataset is used to test the performance of the trained model with data that has never been seen before. All data used is taken from secondary data sources openly available on Kaggle, and the full dataset can be accessed via the link: https://www.kaggle.com/datasets/gpiosenka/good-guysbad-guys-image-data-set.

2.3 Performa Measure

The confusion matrix stands as a pivotal tool in evaluating the accuracy of an object estimation model, providing a comprehensive assessment of its performance through the comparison of predicted classification outcomes with actual class labels [23]. Central to this evaluation are key metrics such as accuracy, precision, recall, and the F1-Score, each offering unique insights into the model's predictive capabilities. Accuracy, delineated by Eq (1), gauges the proportion of correctly predicted instances to the total instances, serving as a fundamental measure of overall correctness. Precision, as delineated by Eq (2), assesses the accuracy of positive predictions by calculating the ratio of true positive instances to the total predicted positive instances, thereby emphasizing the model's ability to avoid false positives. Meanwhile, recall, as defined by Eq (3), quantifies the model's ability to identify true positive instances by computing the ratio of true positive instances to the total actual positive instances. The F1-Score, the harmonic mean of precision and recall illustrated by Eq (4), offers a balanced evaluation of the model's performance in both positive and negative predictions. The computation of these metrics necessitates the utilization of specific formulas incorporating true positive human character (TP_{hc}), true negative human character (TN_{hc}), false positive human character (FP_{hc}), and false negative human character (FN_{hc}) values, providing a nuanced understanding of the model's strengths and areas for improvement in object estimation tasks[24].

$$Accuracy = \frac{TN_{hc}+TP_{hc}}{TN_{hc}+FP_{hc}+TP_{hc}+FN_{hc}} \quad (1)$$

$$Precision = \frac{TP_{hc}}{TP_{hc}+FP_{hc}} \quad (2)$$

$$Recall = \frac{TP_{hc}}{TP_{hc}+FN_{hc}} \quad (3)$$

$$F1 = \frac{2*Presicion*Recall}{\text{Presicion+Recall}} \quad (4)$$

2.4 ResNet50 Framework

ResNet50 is a deep neural network architecture used for image recognition and classification tasks. It is part of the ResNet (Residual Network) architecture family developed by Kaiming He et al. in 2015[25]. ResNet50 has a deep structure consisting of 50 layers (denoted by the number 50), which include convolutional layers, batch normalization, ReLU activation functions, and pooling layers, and it uses skip connections (also known as shortcut connections) to address the vanishing gradient problem and enable training of deeper networks. The ResNet50 architecture is renowned for successfully addressing the problem of degradation that occurs when increasing the network depth, meaning adding layers leads to decreased model performance [26]. By employing skip connections, ResNet50 enables training of very deep networks with hundreds of layers while maintaining or even improving model performance. ResNet50 has been widely used in various image recognition applications, including object recognition, object detection, and image classification [27] [28] . Resnet50 Framework is illustrated in figure 2.

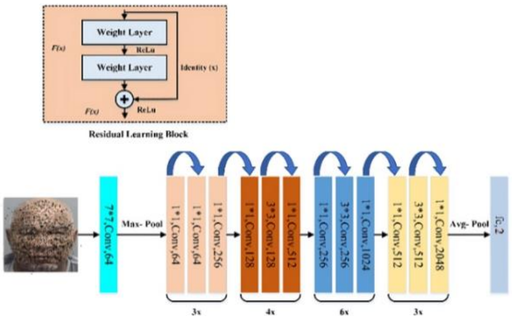


Fig. 2. Resnet50 Framework

2.5 Convolutional Block Attention Module (CBAM)

CBAM, or Convolutional Block Attention Module, represents a breakthrough in image processing pioneered by Woo et al. in 2018 [29]. This module is designed to empower a convolutional neural network (CNN) with the capability to dynamically focus on significant features within the processed image. Comprising two core modules - the Spatial Attention Module and the Channel Attention Module - CBAM functions to accentuate crucial features across both the spatial and channel dimensions of the image [30] [31].

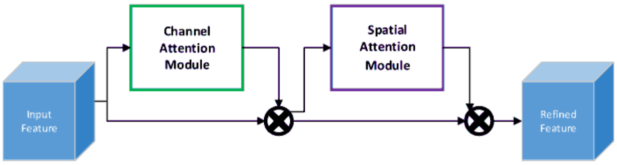


Fig. 3. Convolutional Block Attention Module (CBAM)

Firstly, the channel attention module compresses input features in the spatial dimension, employing global max pooling and global average pooling based on width and height, respectively. These two pooled one-dimensional vectors are then fed into the shared multilayer perceptron (MLP) model, where the corresponding elements of the MLP output features are summed individually. Secondly, employing the sigmoid activation function, an inner product operation is executed with the initial feature map [32]. The resulting output feature map serves as the required input feature for the spatial attention module, as illustrated in Figure 4

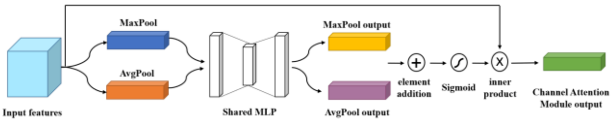


Fig. 4. Channel Attention Module (CAM)

The feature map generated by the channel attention module acts as the input feature map for this subsequent module. Initially, global max pooling and global average pooling are conducted, followed by merging the resulting feature maps along the channel-based dimension. Subsequently, a 7 × 7 convolution operation is applied [33]. Post sigmoid activation, an inner product operation occurs between the output feature map and the feature map provided by the spatial attention module, yielding the final generated features, as depicted in Figure 5.

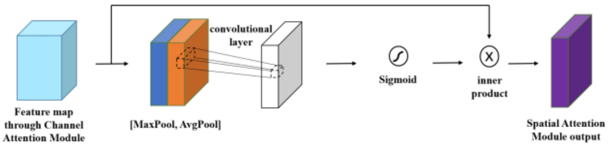

convolutional layer
Feature map through Channel Attention Module
[MaxPool, AvgPool]
Sigmoid
inner product
Spatial Attention Module output

Fig. 5. Spatial Attention Module (SAM)

2.6 ResNet50 with added Convolutional Block Attention Module (CBAM)

The Convolutional Block Attention Module (CBAM) has become a significant innovation in the field of image processing and has been widely applied in various deep learning architectures, such as DeepConvLSTM + CBAM

Attention [34], CBAM-VGGNet [35] [36], and XceptionCBAM-Dense [37].

The addition of the Convolutional Block Attention Module (CBAM) to the ResNet50 network represents a significant step in enhancing its ability to identify important features within images. By integrating CBAM, the ResNet50 network can adaptively select and attend to the most relevant features at each processing stage. This enables the network to better understand the spatial context and channel information within images, resulting in richer and more relevant feature representations.

At the convolutional block level, CBAM allows the network to highlight important spatial features using the Spatial Attention Module (SAM). By leveraging convolution layers and sigmoid activation, SAM generates spatial attention maps that emphasize crucial areas in the image. Then, at the channel level, the Channel Attention Module (CAM) enables the network to highlight important features within each channel using global average pooling (GAP) and fully connected layers. CAM produces channel attention maps that highlight the most informative channels in the image.

The integration of these attention modules at each ResNet50 block allows the network to adaptively attend to important features within images at each processing stage. The ultimate result of integrating CBAM with ResNet50 is a network that is more capable of identifying and leveraging crucial information within images, thereby enhancing the network's performance in image processing tasks such as classification and object detection. By embedding CBAM into the ResNet50 network, we can produce a more sophisticated and adaptive model in understanding the visual content of images.

Here is the ResNet50 framework that CBAM has added:

Input Layer: The input layer receives input images with appropriate dimensions, for example, RGB color images with dimensions of 224x224x3.

Convolutional Layer: The first convolutional layer has 64 filters with a kernel size of 7x7 and a stride of 2. It is followed by batch normalization and ReLU activation function. $\text{Conv}(7\times7,64,\text{stride}=2)$ $\text{BatchNormalization}()$ $\text{ReLU}()$

Max Pooling: Max pooling is performed with a kernel size of 3x3 and a stride of 2. $\text{MaxPool}(3\times3,\text{stride}=2)$

Residual Blocks with CBAM:

1) Shortcut Connections: $\text{Shortcut}(X)=X$

2) First Convolutional Layer:

   $F1(X)=\text{Conv}(3\times3,64)(X)$

   $F1(X)=\text{BatchNormalization}()(F1(X))$

   $F1(X)=\text{ReLU}()(F1(X))$

3) Convolutional Block Attention Module (CBAM):

   Spatial Attention Module – SAM: $Mspatial=\text{Sigmoid}(\text{Convspatial}(F1(X)))$

   Channel Attention Module – CAM: $Mchannel=\text{Sigmoid}(\text{FC}(\text{GAP}(F1(X))))$

   Integration of SAM and CAM: $Mcbam=Mspatial\otimes Mchannel$

   Where Mcbam is the output of CBAM, Mspatial is the output of SAM, and Mchannel is the output of CAM. $\otimes$ represents element-wise multiplication.

4) Second Convolutional Layer:

   $F2(F1(X)\otimes Mcbam)=\text{Conv}(3\times3,64)(F1(X)\otimes Mcbam)$

   $F2(F1(X)\otimes Mcbam)=\text{BatchNormalization}()(F2(F1(X)\otimes Mcbam))$

   $F2(F1(X)\otimes Mcbam)=\text{ReLU}()(F2(F1(X)\otimes Mcbam))$

5) Third Convolutional Layer:

   $F3(F2(X)\otimes Mcbam)=\text{Conv}(3\times3,128)(F2(X)\otimes Mcbam)$

   $F3(F2(X)\otimes Mcbam)=\text{BatchNormalization}()(F3(F2(X)\otimes Mcbam))$

   $F3(F2(X)\otimes Mcbam)=\text{ReLU}()(F3(F2(X)\otimes Mcbam))$

6) Fourth Convolutional Layer:

   $F4(F3(X)\otimes Mcbam)=\text{Conv}(3\times3,256)(F3(X)\otimes Mcbam)$

   $F4(F3(X)\otimes Mcbam)=\text{BatchNormalization}()(F4(F3(X)\otimes Mcbam))$

   $F4(F3(X)\otimes Mcbam)=\text{ReLU}()(F4(F3(X)\otimes Mcbam))$

7) Fifth Convolutional Layer:

   $F5(F4(X)\otimes Mcbam)=\text{Conv}(3\times3,512)(F4(X)\otimes Mcbam)$

   $F5(F4(X)\otimes Mcbam)=\text{BatchNormalization}()(F5(F4(X)\otimes Mcbam))$

   $F5(F4(X)\otimes Mcbam)=\text{ReLU}()(F5(F4(X)\otimes Mcbam))$

8) Global Average Pooling: Global average pooling is performed to generate a feature vector. $\text{GlobalAvgPool}()$

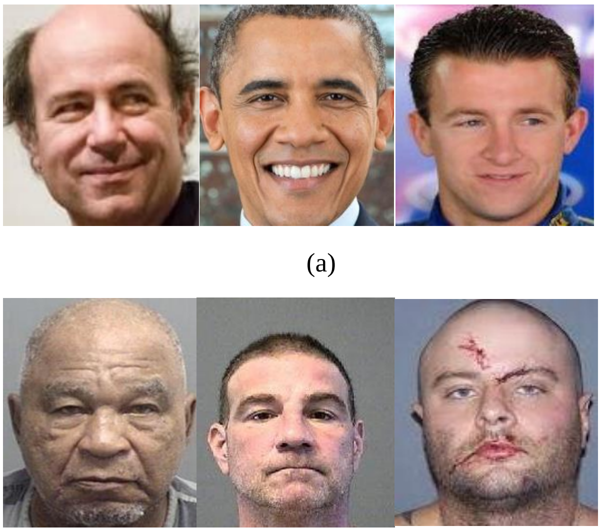
Fully Connected Layer: The fully connected layer consists of the number of output neurons corresponding to the desired number of classes. FC(2)

Softmax Activation: The softmax activation function is used at the output layer to generate the probability distribution over possible classes. Softmax()

3. Results and Discussion

3.1 Sample of Human Character

This research focuses on classifying human characters based on facial images, distinguishing between two types of characters, namely Savory and Unsavory. The sample data used in this study is presented in Figure 6, reflecting the variation of human characters observed and evaluated using the ResNet50 model enriched with the addition of the Convolutional Block Attention Module (CBAM).



(a)

(b)

Fig. 6. Sample of Human Character from a Facial Image

(a) Savory (b) Unsavory.

3.2 Training and Evaluation ResNet50

This study presents the ResNet50 model developed to classify human characters based on facial images. ResNet50 is one of the convolutional neural network architectures proven effective in handling complex tasks in image processing. Visualization of the training model (Figure 7) will display the performance progress of the model over training iterations or epochs. The classification results will be further elucidated through the confusion matrix (Figure 8). The confusion matrix provides a visual depiction of how well the model can classify various categories of human characters. Additionally, the model's performance evaluation will be presented in the form of a classification report structured in Table 1. This report will provide in-depth analysis of accuracy, precision, recall, and F1 scores for each classification class. With the provided information, we can evaluate how well the model can classify human characters based on facial images.

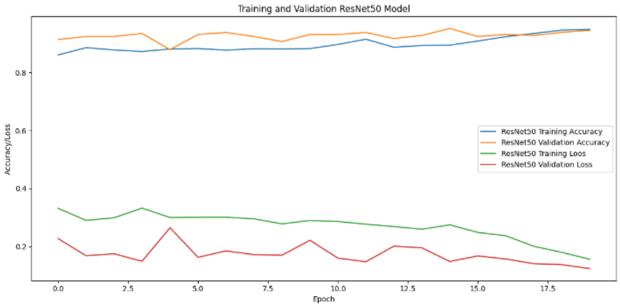


Fig. 7. Training ResNet50

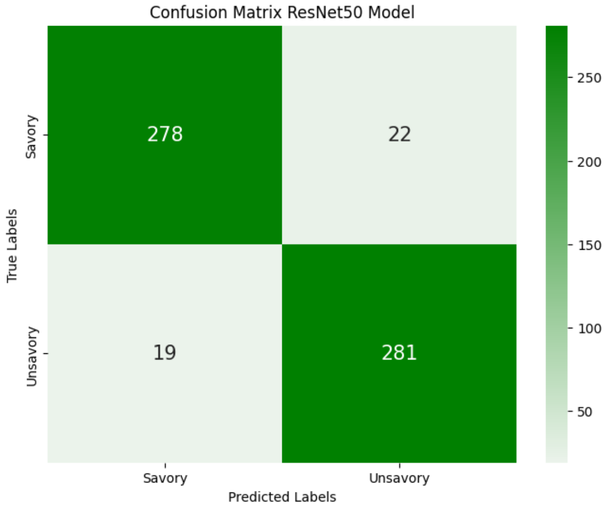


Fig. 8. Confusion Matrix ResNet50

Table 1. Classification Report ResNet50

Class	Precision	Recall	F1-Score	Support
Savory	0.9360	0.9267	0.9313	300
Unsavory	0.9274	0.9367	0.9320	300
Accuracy		0.913		
Average	0.9317	0.9317	0.9317	600

The training process of the ResNet50 model for classifying human characters based on facial images resulted in several important evaluation metrics. During training, the model's validation accuracy varied between 87.85% to 95.14%, with an average accuracy of approximately 92.88%. Meanwhile, the training accuracy ranged from 86.00% to 94.80%, with an average accuracy of around 89.91%. Analysis of the validation loss indicated variations between 0.124 to 0.2648, with an average loss of about 0.175, while the training loss ranged from 0.156 to 0.3317, with an average loss of approximately 0.267. These data provide a comprehensive overview of the model's performance during the training process, where increasing accuracy and decreasing loss indicate an improvement in the model's ability to classify human characters based on facial images.

The confusion matrix analysis of the ResNet50 model in classifying human characters based on facial images showed the following results: The model correctly predicted 278 human characters as "Savory" and 281 human characters as "Unsavory." However, 22 human characters that were actually "Savory" were incorrectly predicted as "Unsavory", and 19 human characters that were actually "Unsavory" were incorrectly predicted as "Savory." The analysis of the confusion matrix on the ResNet50 model indicates a fairly good performance in classifying human characters based on facial images. Although there were some prediction errors, the number is relatively small compared to the correct predictions. Thus, while there is still room for improvement, these results indicate that the ResNet50 model is capable of identifying and distinguishing human characters with a satisfactory level of accuracy.

In the classification results using the ResNet50 model for human character recognition, the Savory and Unsavory classes were observed. It was found that the model was able to classify human characters into both classes with a high level of accuracy. For the Savory class, a precision of 0.9360 indicates that about 93.60% of all predictions predicted as Savory are actually Savory. A recall of 0.9267 indicates that about 92.67% of all actual Savory instances were found by the model. The F1-score of 0.9313 is the harmonic average of precision and recall for the Savory class. Meanwhile, for the Unsavory class, a precision of 0.9274 indicates that about 92.74% of all predictions predicted as Unsavory are actually Unsavory. A recall of 0.9367 indicates that about 93.67% of all actual Unsavory instances were found by the model. The F1-score of 0.9320 is the harmonic average of precision and recall for the Unsavory class. The overall accuracy value of 0.9317 indicates that the model can make predictions correctly to the extent of 93.17% for all classes. The average values of precision, recall, and F1-score (shown as "Average") are 0.9317, indicating the model's consistent performance across all classification classes. Thus, these findings affirm that the ResNet50 model has good capabilities in recognizing and distinguishing human characters based on facial images, as well as being able to classify with high accuracy between the two classes.

3.3 Training and Evaluation ResNet50 with CBAM

This study presents a ResNet50 model enriched with the addition of CBAM for classifying human characters based on facial images. ResNet50 enriched with CBAM represents a more advanced application of convolutional neural network architecture that has proven effective in handling complex tasks in image processing. The visualization of the training model (Figure 9) will demonstrate the evolution of the model's performance throughout the iteration or epoch training process. Classification analysis will be further explored through the confusion matrix presented in Figure 10. The confusion matrix provides a visual illustration of the model's ability to classify various categories of human characters. Additionally, the model's performance evaluation will be presented through a structured classification report (Table 2). This report will provide in-depth analysis regarding accuracy, precision, recall, and F1 scores for each classification class. With the information available, we can comprehensively evaluate how well the model can classify human characters based on facial images.

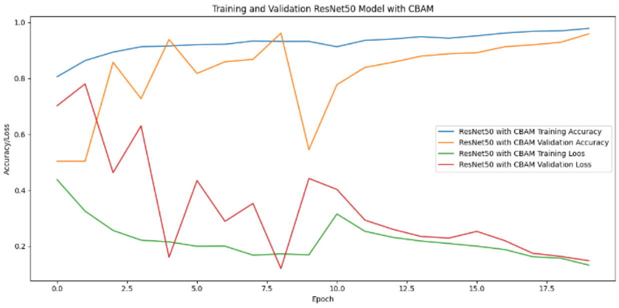


Fig. 9. Training ResNet50 with CBAM

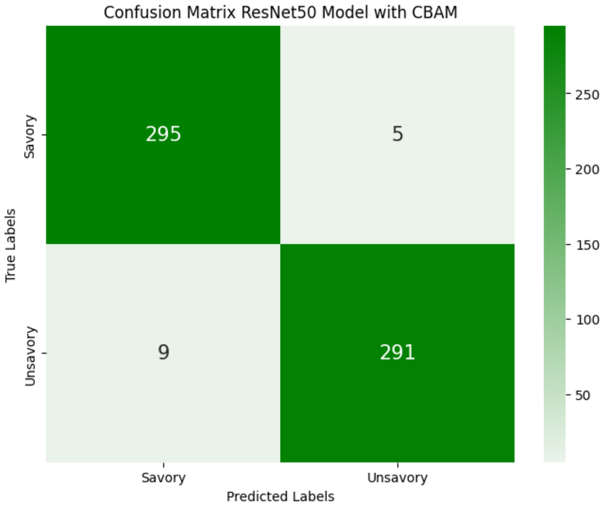


Fig. 10. Confusion Matrix ResNet50 with CBAM

Table 2. Classification Report ResNet50 with CBAM

Class	Precision	Recall	F1-Score	Support
Savory	0.9704	0.9833	0.9768	300
Unsavory	0.9831	0.9700	0.9768	300
Accuracy		0.9767		
Average	0.9768	0.9767	0.9767	600

The training process of the ResNet50 model enriched with the addition of CBAM in classifying human characters based on facial images resulted in several important evaluation metrics. During training, the model's validation accuracy ranged from 50.35% to 95.92%, with an average accuracy of about 92.88%. On the other hand, the training accuracy ranged from 80.60% to 97.87%, with an average accuracy of about 91.72%. Analysis of the validation data loss showed variation between 0.1195 and 0.7802, with an average loss of around 0.303. Meanwhile, the training loss ranged from 0.132 to 0.4375, with an average loss of about 0.237. This data provides a comprehensive overview of the model's performance during training, where increased accuracy and decreased loss indicate an improvement in the model's ability to classify human characters based on facial images.

The confusion matrix analysis of the ResNet50 model enriched with the addition of CBAM in classifying human characters based on facial images showed the following results: The model successfully predicted 295 human characters as "Savory" and 291 human characters as "Unsavory" correctly. However, 5 human characters that were actually "Savory" were misclassified as "Unsavory," and 9 human characters that were actually "Unsavory" were misclassified as "Savory." The results of the confusion matrix analysis on the ResNet50 model enriched with the addition of CBAM showed very good performance in classifying human characters based on facial images. Although there were some prediction errors, the number was relatively small compared to the number of correct predictions. Thus, although there is still room for improvement, these results indicate that the ResNet50 model enriched with the addition of CBAM is capable of identifying and distinguishing human characters with a very satisfactory level of accuracy.

In the classification results using the ResNet50 model with the addition of CBAM for human character recognition cases, both Savory and Unsavory classes were observed. The results showed that this model was able to classify human characters into both classes with significant accuracy. For the Savory class, a precision of 0.9704 indicates that 97.04% of predictions classified as Savory are indeed Savory. A recall of 0.9833 indicates that the model successfully detects 98.33% of the total human characters that are actually Savory. An F1-score of 0.9768 reflects the balance between precision and recall for the Savory class. Meanwhile, for the Unsavory class, a precision of 0.9831 indicates that 98.31% of predictions classified as Unsavory are indeed Unsavory. A recall of 0.97 indicates that the model successfully detects 97.0% of the total human characters that are actually Unsavory. An F1-score of 0.9768 reflects the balance between precision and recall for the Unsavory class. The model's accuracy of 0.9767 indicates the overall percentage of correct predictions out of all predictions made. With an average precision, recall, and F1-score of around 0.9768, this model demonstrates excellent performance in classifying human characters based on facial images, with high accuracy and a balance between precision and recall for both classes.

3.4 Other Framework

In the context of developing convolutional neural network (CNN) models for image processing tasks, particularly human character classification, there are several other frameworks that serve as alternatives, such as EfficientNetB3 and GoogleNet, each with its own characteristics and advantages in representation and classification. EfficientNet is a series of convolutional neural network architectures developed by Google Brain. Compared to other architectures, EfficientNet is designed to achieve higher efficiency levels in image processing. EfficientNetB3 is one variant of the EfficientNet series that balances efficiency and accuracy. Its main advantage lies in its ability to achieve performance equivalent to or even better than other architectures with fewer parameters. GoogleNet, also known as Inception-v1, is one of the convolutional neural network architectures developed by Google's research team. This architecture is renowned for introducing the Inception module, which utilizes convolutions of various sizes on the same input layer. GoogleNet excels in reducing the number of parameters compared to previous architectures while maintaining good performance in classification.

The selection of EfficientNetB3 and GoogleNet as benchmarks for ResNet50 is based on several reasons. Firstly, all three architectures have proven successful in image processing and classification tasks. However, each has a different approach in architectural design, which can provide additional insights into the efficiency and reliability of the model. Additionally, comparing ResNet50 with EfficientNetB3 and GoogleNet can help understand the trade-off between efficiency, accuracy, and model complexity. Thus, this comparison can provide a deeper understanding of the most suitable architecture choice for the case of human character recognition based on facial images.

The classification results from EfficientNetB3 and GoogleNet will be presented in the form of a confusion matrix in Figure 11. The confusion matrix provides a visual representation of how well the model can classify various categories of human characters. Additionally, the model's performance evaluation will be presented in the form of a classification report in Tables 3 and 4. This report will provide in-depth analysis of accuracy, precision, recall, and F1 score for each classification class, enabling comprehensive comparison among the two models.

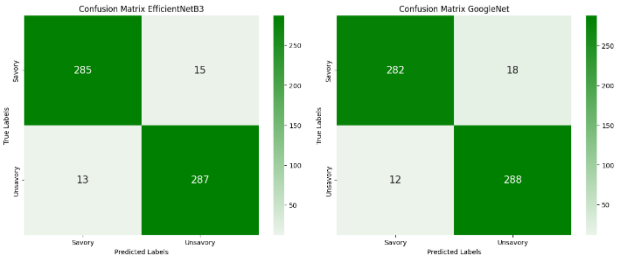


Fig.11. Confusion Matrix other Network (Left) EfficientNetB3 (Right) GoogleNet

Table 3. Classification Report EfficientNetB3

Class	Precision	Recall	F1-Score	Support
Savory	0.9564	0.9500	0.9532	300
Unsavory	0.9503	0.9567	0.9535	300
Accuracy		0.9533		
Average	0.9534	0.9533	0.9533	600

Table 4. Classification Report GoogleNet

Class	Precision	Recall	F1-Score	Support
Savory	0.9592	0.9400	0.9495	300
Unsavory	0.9412	0.9600	0.9505	300
Accuracy		0.9500		
Average	0.9502	0.9500	0.9500	600

EfficientNetB3 and GoogleNet, as comparisons to ResNet50 enriched with CBAM, provide insights into the performance of both models in classifying human characters based on facial images. EfficientNetB3 shows high precision for both Savory and Unsavory classes, with values of 0.9564 and 0.9503 respectively. This indicates that the majority of predictions classified as Savory or Unsavory by this model are correct. High recall for both classes (0.95 for Savory and 0.9567 for Unsavory) suggests that the EfficientNetB3 model is capable of detecting most of the human characters actually belonging to the Savory or Unsavory classes. The high F1-scores for both classes (0.9532 for Savory and 0.9535 for Unsavory) indicate a balance between precision and recall. Meanwhile, GoogleNet also demonstrates good results with high precision for both classes, namely 0.9592 for Savory and 0.9412 for Unsavory. The recall for both classes is also high, with values of 0.94 for Savory and 0.96 for Unsavory. The good F1-scores for both classes (0.9495 for Savory and 0.9505 for Unsavory) indicate that GoogleNet maintains a good balance between precision and recall. Both models, EfficientNetB3 and GoogleNet, show high levels of accuracy, namely 0.9533 and 0.95 respectively.

3.5. Dicussion

The Comparative classification report between ResNet50 and Other Framework can be seen in the table below :

Table 5. Comparison Classification Report ResNet50 and Other Framework

Class	Precision	Recall	F1-Score	Accuracy
Efficient NetB3	0.9534	0.9533	0.9533	0.9533
GoogleNet	0.9502	0.9500	0.9500	0.9500
ResNet	0.9317	0.9317	0.9317	0.9317
ResNet with CBAM (Purpose Methode)	**0.9768**	**0.9767**	**0.9767**	**0.9767**

This study provides accurate insights into the capabilities of several convolutional neural network (CNN) architectures in classifying human characters based on facial images. Evaluation results using accuracy, precision, recall, and F1-score metrics provide in-depth understanding of the performance of each model.

The comparison between ResNet50 with and without the addition of CBAM shows a significant improvement in the model's performance in classifying human characters. ResNet50 enriched with CBAM achieves higher accuracy and demonstrates better performance in identifying Savory and Unsavory classes. Similarly, compared to EfficientNetB3 and GoogleNet, ResNet50 with CBAM still outperforms. However, these research findings can effectively be associated with the body of knowledge on image processing and CNN model development. The discovery that ResNet50 with the addition of CBAM can enhance performance in human character recognition contributes significantly to understanding how to improve classification accuracy in specific use cases.

The main limitation of this study is the limited focus on classifying human characters based on facial images, without considering other variations in the dataset such as facial orientation, lighting, or expression. Additionally, the use of datasets that may not reflect the diversity of human characters broadly could limit the generalization of results. Threats to validity include the possibility of bias in the data collection process or pre-processing policies that may affect classification outcomes.

Recommendations for further research include involving more diverse and representative datasets, as well as conducting additional experiments to understand the impact of factors such as facial orientation, lighting, and expression on model performance. Furthermore, incorporating data enrichment techniques and advanced processing such as image augmentation or transfer learning techniques can also enhance model performance in more complex cases. Lastly, further evaluation of CBAM integration into other architectures or the development of other enrichment methods can also be an interesting direction for further exploration

4. Conclusion

This study provides a deep understanding of the performance of several convolutional neural network (CNN) architectures in classifying human characters based on facial images. Evaluation results using accuracy, precision, recall, and F1-score metrics provide an accurate overview of the capabilities of each model. The comparison between ResNet50 with and without the addition of CBAM shows a significant improvement in the model's performance in classifying human characters. For future learning, it is necessary to enhance the dataset variation to reflect the diversity of human characters more broadly, as well as consider factors such as facial orientation, lighting, and expression in the classification process. The use of more varied and representative datasets can improve the generalization of results, while data enrichment techniques such as image augmentation and transfer learning can help enhance model performance in more complex cases. The benefit of this research is to provide valuable insights into how to improve the accuracy of human character classification based on facial images. The research findings can be applied in various practical applications such as facial recognition for security systems, character identification in videos, or emotion recognition based on facial expressions. However, this research has limitations, such as the limited focus on classifying human characters based on facial images without considering other variations in the dataset such as facial orientation, lighting, or expression. Threats to validity include the possibility of bias in the data collection process or preprocessing policies that may affect classification outcomes. Therefore, recommendations for further research include involving more varied and representative datasets, as well as conducting additional experiments to understand the impact of these factors on model performance. Lastly, further evaluation of CBAM integration into other architectures or the development of other enrichment methods can also be an interesting direction for further exploration.

Author contributions

All Authors contributed equally to this work.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] H. Vasudevan, A. Michalas, N. Shekokar, and M. Narvekar, Eds., Advanced Computing Technologies and Applications: Proceedings of 2nd International Conference on Advanced Computing Technologies and Applications—ICACTA 2020. in Algorithms for Intelligent Systems. Singapore: Springer Singapore, 2020. doi: 10.1007/978-981-15-3242-9.

[2] R. R. Koli and T. I. Bagban, “Human Action Recognition Using Deep Neural Networks,” in 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), London, United Kingdom: IEEE, Jul. 2020, pp. 376–380. doi: 10.1109/WorldS450073.2020.9210345.

[3] B. R. Ilyas, B. Mohammed, M. Khaled, and K. Miloud, “Enhanced Face Recognition System Based on Deep CNN,” in 2019 6th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria: IEEE, Nov. 2019, pp. 1–6. doi: 10.1109/ISPA48434.2019.8966797.

[4] V. R. R. Chirra, S. R. Uyyala, and V. K. K. Kolli, “Virtual facial expression recognition using deep CNN with ensemble learning,” J Ambient Intell Human Comput, vol. 12, no. 12, pp. 10581–10599, Dec. 2021, doi: 10.1007/s12652-020-02866-3.

[5] T. J. Iyer, R. K., R. Nersisson, Z. Zhuang, A. N. Joseph Raj, and I. Refayee, “Machine Learning-Based Facial Beauty Prediction and Analysis of Frontal Facial Images Using Facial Landmarks and Traditional Image Descriptors,” Computational Intelligence and Neuroscience, vol. 2021, pp. 1–14, Aug. 2021, doi: 10.1155/2021/4423407.

[6] V. D. A. Kumar, V. D. A. Kumar, G. K. Rajeswari, and M. Anitha, “Human Character Identification Based On a New Biometric Pattern – A Contemporary Approach,” Procedia Computer Science, vol. 133, pp. 99–107, 2018, doi: 10.1016/j.procs.2018.07.013.

[7] B. Li and D. Lima, “Facial expression recognition via ResNet-50,” International Journal of Cognitive Computing in Engineering, vol. 2, pp. 57–64, Jun. 2021, doi: 10.1016/j.ijcce.2021.02.002.

[8] D. Setyadi, T. Harsono, and S. Wasista, “Human character recognition application based on facial feature using face detection,” in 2015 International Electronics Symposium (IES), Surabaya, Indonesia:

IEEE, Sep. 2015, pp. 263–267. doi: 10.1109/ELECSYM.2015.7380852.

[9] E. A. Moh. Iqbal, R. Kusumawati, and I. B. Santoso, “Hybrid Model Transfer Learning ResNet50 and Support Vector Machine for Face Mask Detection,” International Journal of Advances in Data and Information Systems, vol. 4, no. 2, pp. 125–134, Sep. 2023, doi: 10.25008/ijadis.v4i2.1297.

[10] J. Y. I. Alzamily, S. B. Ariffin, and S. S. A. Naser, “Classification of Encrypted Images using Deep Learning – ResNet50,” . Vol., no. 21, 2022.

[11] K. Shivam, J.-C. Tzou, and S.-C. Wu, “Multi-Step Short-Term Wind Speed Prediction Using a Residual Dilated Causal Convolutional Network with Nonlinear Attention,” Energies, vol. 13, no. 7, p. 1772, Apr. 2020, doi: 10.3390/en13071772.

[12] P. Chen, S. Liu, and S. Kolmanič, “Research on Vehicle Re-Identification Algorithm Based on Fusion Attention Method,” Applied Sciences, vol. 13, no. 7, p. 4107, Mar. 2023, doi: 10.3390/app13074107.

[13] Z. Zhang, H. Mamat, X. Xu, A. Aysa, and K. Ubul, “FAS-Res2net: An Improved Res2net-Based Script Identification Method for Natural Scenes,” Applied Sciences, vol. 13, no. 7, p. 4434, Mar. 2023, doi: 10.3390/app13074434.

[14] G. Wang, Y. Yuan, X. Chen, J. Li, and X. Zhou, “Learning Discriminative Features with Multiple Granularities for Person Re-Identification,” in Proceedings of the 26th ACM international conference on Multimedia, Seoul Republic of Korea: ACM, Oct. 2018, pp. 274–282. doi: 10.1145/3240508.3240552.

[15] W. Islam, M. Jones, R. Faiz, N. Sadeghipour, Y. Qiu, and B. Zheng, “Improving Performance of Breast Lesion Classification Using a ResNet50 Model Optimized with a Novel Attention Mechanism,” Tomography, vol. 8, no. 5, pp. 2411–2425, Sep. 2022, doi: 10.3390/tomography8050200.

[16] L. Yang, C. Wang, J. Yu, N. Xu, and D. Wang, “Method of Peanut Pod Quality Detection Based on Improved ResNet,” Agriculture, vol. 13, no. 7, p. 1352, Jul. 2023, doi: 10.3390/agriculture13071352.

[17] J. Zhang, Y. Xie, Y. Xia, and C. Shen, “Attention Residual Learning for Skin Lesion Classification,” IEEE Trans. Med. Imaging, vol. 38, no. 9, pp. 2092–2103, Sep. 2019, doi: 10.1109/TMI.2019.2893944.

[18] Rujito, Muhathir, N. Khairina, V. Ilhadi, M. Ula, and I. Sahputra, “Enhancing Larval Classification Accuracy Through Hyperparameter Optimization in ResNet50 with Three Different Optimizers,” in 2023 International Conference on Modeling &amp; E-Information Research, Artificial Learning and Digital Applications (ICMERALDA), Karawang, Indonesia: IEEE, Nov. 2023, pp. 56–61. doi: 10.1109/ICMERALDA60125.2023.10458194.

[19] J. Hemalatha, S. Roseline, S. Geetha, S. Kadry, and R. Damaševičius, “An Efficient DenseNet-Based Deep Learning Model for Malware Detection,” Entropy, vol. 23, no. 3, p. 344, Mar. 2021, doi: 10.3390/e23030344.

[20] S. Tammina, “Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images,” IJSRP, vol. 9, no. 10, p. p9420, Oct. 2019, doi: 10.29322/IJSRP.9.10.2019.p9420.

A. Michele, V. Colin, and D. D. Santika, “MobileNet Convolutional Neural Networks and Support Vector Machines for Palmprint Recognition,” Procedia Computer Science, vol. 157, pp. 110–117, 2019, doi: 10.1016/j.procs.2019.08.147.

[21] R. Syuhada, Muhathir, N. Khairina, R. Muliono, Susilawati, and Z. Sembiring, “Analyzing the Effectiveness of VGG Deep Learning Architecture for Mushroom Type Classification,” in 2023 International Conference of Computer Science and Information Technology (ICOSNIKOM), 2023, pp. 1–6. doi: 10.1109/ICoSNIKOM60230.2023.10364551.

[22] M. Muhathir, N. Khairina, R. K. I. Barus, M. Ula, and I. Sahputra, “Preserving Cultural Heritage Through AI: Developing LeNet Architecture for Wayang Image Classification,” IJACSA, vol. 14, no. 9, 2023, doi: 10.14569/IJACSA.2023.0140919.

A. Indrawati, A. Rahman, E. Pane, and Muhathir, “Classification of Diseases in Oil Palm Leaves Using the GoogLeNet Model,” Baghdad Sci.J, vol. 20, no. 6(Suppl.), p. 2508, Dec. 2023, doi: 10.21123/bsj.2023.8547.

[23] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[24] M. Azwan, Muhathir, D. Noviandri, A. Indrawati, and R. Aziz, “Robust Classification of Red Chili Plant Leaves Using Smartphone Camera Data and ResNet Model in Noisy Environments,” in 2023 International Conference on Modeling & E-Information Research, Artificial Learning and Digital Applications (ICMERALDA), 2023, pp. 191–196. doi: 10.1109/ICMERALDA60125.2023.10458175.

[25] M. Govindan, V. K. Dhakshnamurthy, K. Sreerangan, M. D. Nagarajan, and S. K. Rajamanickam, “A Framework for Early Detection of Glaucoma in Retinal Fundus Images Using Deep Learning,” in CC 2023, MDPI, Feb. 2024, p. 3. doi: 10.3390/engproc2024062003.

[26] M. Muhathir, M. F. D. Ryandra, R. B. Y. Syah, N. Khairina, and R. Muliono, “Convolutional Neural Network (CNN) of Resnet-50 with Inceptionv3 Architecture in Classification on X-Ray Image,” in Artificial Intelligence Application in Networks and Systems, P. Silhavy Radek and Silhavy, Ed., Cham: Springer International Publishing, 2023, pp. 208–221.

[27] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in Computer Vision – ECCV 2018, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11211. , Cham: Springer International Publishing, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.

[28] S. Dou, L. Wang, D. Fan, L. Miao, J. Yan, and H. He, “Classification of Citrus Huanglongbing Degree Based on CBAM-MobileNetV2 and Transfer Learning,” Sensors, vol. 23, no. 12, p. 5587, Jun. 2023, doi: 10.3390/s23125587.

[29] Z. Bai, R. Zhu, D. He, S. Wang, and Z. Huang, “Adulteration Detection of Pork in Mutton Using Smart Phone with the CBAM-Invert-ResNet and Multiple Parts Feature Fusion,” Foods, vol. 12, no. 19, p. 3594, Sep. 2023, doi: 10.3390/foods12193594.

[30] W. Sheng, X. Yu, J. Lin, and X. Chen, “Faster RCNN Target Detection Algorithm Integrating CBAM and FPN,” Applied Sciences, vol. 13, no. 12, p. 6913, Jun. 2023, doi: 10.3390/app13126913.

[31] Y. Muhtar, M. Muhammat, N. Yadikar, A. Aysa, and K. Ubul, “FC-ResNet: A Multilingual Handwritten Signature Verification Model Using an Improved ResNet with CBAM,” Applied Sciences, vol. 13, no. 14, p. 8022, Jul. 2023, doi: 10.3390/app13148022.

[32] S. Agac and O. Durmaz Incel, “On the Use of a Convolutional Block Attention Module in Deep Learning-Based Human Activity Recognition with Motion Sensors,” Diagnostics, vol. 13, no. 11, p. 1861, May 2023, doi: 10.3390/diagnostics13111861.

[33] E. Haque and R. Ahmed, “Classification of Human Monkeypox Disease Using Deep Learning Models and Attention Mechanisms”.

[34] Areesha Ijaz et al., “Modality Specific CBAM-VGGNet Model for the Classification of Breast Histopathology Images via Transfer Learning,” IEEE, pp. 15750–15762, 2023, doi: https://doi.org/10.1109/ACCESS.2023.3245023.

[35] Sheng Yu, Shangzhu Jin, Jun Peng, Haiyang Liu, and Yuanyuan He, “Application of a new deep learning method with CBAM in clothing image classification,” presented at the 2021 IEEE International Conference on Emergency Science and Information Technology (ICESIT), Chongqing, China: IEEE, 2021. doi: https://doi.org/10.1109/ICESIT53460.2021.9696783.