

## Article

# Comparative Analysis of AI-Based Facial Identification and Expression Recognition Using Upper and Lower Facial Regions

Seunghyun Kim <sup>1,†</sup>, Byeong Seon An <sup>1,†</sup> and Eui Chul Lee <sup>2,\*</sup> 

<sup>1</sup> Department of AI & Informatics, Graduate School, Sangmyung University, Hongjimun 2-gil 20, Jongno-gu, Seoul 03016, Republic of Korea; 202134015@sangmyung.kr (S.K.); 202232033@sangmyung.kr (B.S.A.)

<sup>2</sup> Department of Human-Centered Artificial Intelligence, Graduate School, Sangmyung University, Hongjimun 2-gil 20, Jongno-gu, Seoul 03016, Republic of Korea

\* Correspondence: ecllee@smu.ac.kr

† These authors contributed equally to this work.

**Abstract:** The COVID-19 pandemic has significantly impacted society, having led to a lack of social skills in children who became used to interacting with others while wearing masks. To analyze this issue, we investigated the effects of masks on face identification and facial expression recognition, using deep learning models for these operations. The results showed that when using the upper or lower facial regions for face identification, the upper facial region allowed for an accuracy of 81.36%, and the lower facial region allowed for an accuracy of 55.52%. Regarding facial expression recognition, the upper facial region allowed for an accuracy of 39% compared to 49% for the lower facial region. Furthermore, our analysis was conducted for a number of facial expressions, and specific emotions such as happiness and contempt were difficult to distinguish using only the upper facial region. Because this study used a model trained on data generated from human labeling, it is assumed that the effects on humans would be similar. Therefore, this study is significant because it provides engineering evidence of a decline in facial expression recognition; however, wearing masks does not cause difficulties in identification.

**Keywords:** face identification; facial expression recognition; face cropping; emotion; masked face



**Citation:** Kim, S.; An, B.S.; Lee, E.C. Comparative Analysis of AI-Based Facial Identification and Expression Recognition Using Upper and Lower Facial Regions. *Appl. Sci.* **2023**, *13*, 6070. <https://doi.org/10.3390/app13106070>

Academic Editors: Mukesh Prasad, Ali Braytee, Xian Tao and Pang-jo Chun

Received: 26 April 2023

Revised: 10 May 2023

Accepted: 12 May 2023

Published: 15 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Owing to the COVID-19 pandemic, health policies have mandated the use of masks in many public institutions and workplaces. Children are particularly encouraged to wear masks because of their weak immune systems. However, the limited facial information visible in people wearing masks can affect children's social skills, which are heavily influenced by their ability to read facial expressions [1,2]. Recent studies have explored methods to identify faces and recognize facial expressions when a mask partially covers the face, particularly when divided into top and bottom parts. For example, one study focused on using the upper facial region for face identification [3]. Another study separated the upper and lower facial regions to derive action units for facial expression recognition [4].

As will be discussed in detail in Section 2, existing facial expression recognition studies only pursue technical improvements in facial expression recognition. The same goes for face identification technology. These studies on improving accuracy have been developed into studies on the recognition rate of the upper and lower parts of the face according to the COVID-19 pandemic. However, studies comparing the results are lacking. We felt the need to interpret the results of facial expression recognition studies and studies regarding face identification of the upper and lower parts of the face.

In this study, we aimed to specify the parts of the face that are most important for both face identification and facial expression recognition, as well as to analyze how masks affect the ability to recognize emotions. Specifically, we analyzed the accuracy of eight facial expressions (neutral, happy, sad, surprise, fear, disgust, anger, and contempt) affected

by the upper or lower parts of the face. We based our study on a deep learning model trained on a sufficiently large amount of data labeled by experts. Not only does this have the advantage of reducing dependence on the demographic pool of testers, but research [5] has shown that the areas on which deep learning models focus during facial expression recognition are not significantly different from those on which humans focus. Similarly, studies [6,7] have compared the performance of human and deep learning models in face identification and found them to be similar. Therefore, this study was conducted under the assumption that a sufficiently trained deep learning model can act as a sample of many different people.

The remainder of this paper is organized as follows: Section 2 provides an overview of related studies, including the models used for face identification and facial expression recognition tasks. Section 3 describes our proposed methods. Section 4 presents an analysis and discussion of the experimental results, and Section 5 concludes the study.

## 2. Related Works

### 2.1. Facial Expression Recognition of Upper and Lower Facial Regions

Several previous studies have analyzed the accuracy of facial expression recognition based on different regions of the face. A brief summary of these studies is presented in Table 1. One of the earliest studies was conducted by Che et al. [8]. They first selected two representative expressions: happiness and sadness; these can be seen as the two most opposing and representative expressions from the perspective of the valence coordinate axis, which is still widely used in facial expression recognition research. They then divided the level of happy and sad expressions into seven steps to share the degree of each expression. They took the same person's happy and sad images and morphed them into seven steps using FantaMorph 4.0 [9]. The test data consisted of the upper and lower facial regions of the generated image separated and randomly merged. Four people were used as data, and the experiment was conducted with 49 images per person. Subsequently, the results were derived from the standard normal and cumulative distribution functions through a survey. Only the upper and lower facial regions were observed to affect happiness [8]. In this study, we also compared the recognition rates for sadness and happiness in the upper and lower facial regions. Moreover, we used natural images of people, derived the results from a much larger dataset, and compared the results of eight different expressions. This clearly showed how the upper and lower facial regions affect facial expressions.

**Table 1.** Summary of facial expression recognition related works.

Method	Summary	Limitations
Che et al. [8]	Using the FantaMorph 4.0 dataset, the sad and happy facial expressions were divided into 7 steps.	Lack of test data set and classifying only happy and sad emotions.
Mika Itoh et al. [10]	The six emotions were tested by changing only the upper part of the face, and the relationship between each emotion was analyzed.	The lack of access to a diverse demographic pool based on the opinion of a female college student. The data used in the experiment were small and self-generated.
Seyedarabi et al. [11]	In the faces extracted from images and videos, the effect of Action Unit on upper and lower facial expression classification was analyzed.	The characteristics of action units were classified for the entire face. Characteristics by region of the face were not analyzed.
Khoeun et al. [12]	Using the CK+ and RAF-DB datasets, this study generated a face wearing a mask and calculated the accuracy through CNN.	Only the upper part of the face was used to analyze the effect of the upper part.

Mika Itoh et al. conducted an experiment using six emotional labels: anger, disgust, fear, happiness, sadness, and surprise, with natural expressions [10]. After setting the upper or lower facial region as neutral and the other face as one of the facial expression

labels, the participants were asked to evaluate the intensity of the emotion. The upper facial region was highly related to anger, surprise, and sadness, whereas the lower facial region was related to fear and happiness. Evaluation of the upper facial region, which is relatively important in facial expression recognition, was not significantly different from that of the entire face. Anger and fear are often confused with disgust and surprise, respectively [10]. The conclusions of this study were based on the opinions of 63 female university students, and the results may not be representative of all people because of the lack of access to a diverse demographic pool; additionally, another limitation is that the data used in the experiment were small and self-generated. Therefore, we used a model trained on a large amount of data to obtain the results.

Seyedarabi et al. examined facial features and extracted expressions from images and videos [11]. Expression classification was based on the Facial Action Coding System (FACS) and lower and upper facial action units (AU); discrimination was performed using probabilistic neural networks (PNNs) and rule-based systems. The experimental results showed an average recognition rate of 96.11% for six basic emotions in face image sequences and 94% for five basic emotions in static face images, detection, tracking, and reasonable classification. In previous studies, such as facial emotion recognition, AU captured features that affected facial expressions [11]. In this study, we conducted experiments to classify facial expressions from different facial regions rather than the characteristics of facial expressions composed of the whole face.

Khoeun et al. proposed a feature vector technique consisting of three steps for recognizing emotions in mask images [12]. First, a mask was applied by using the boundary and area expression techniques such that only the upper part of the image, including the eyes, eyebrows, part of the bridge of the nose, and forehead, was visible. To flexibly extract a set of feature vectors that can effectively represent the features of a masked face, they utilized a feature extraction technique based on the proposed rapid landmark detection method using an infinity shape. Finally, these features, including the positions of the detected landmarks and histograms of the oriented gradients, were introduced into the classification process by adopting convolutional neural networks (CNNs) and long short-term memory networks. Accuracies of 99.30% and 95.58% were achieved for CK+ and RAF-DB, respectively [12]. Khoeun's research focused on using the upper part of the face to improve facial expression recognition accuracy in the context of the COVID-19 pandemic. The study describes the influence of eye area and eyebrows on facial expressions. However, the facial action unit of facial expression is located not only in the upper facial region but also in the lower facial region. Our research does not aim to improve the accuracy of using only the upper facial region but aims to determine which part of the face is most affected when recognizing facial expressions. Therefore, we studied two cases: a dataset with the lower part of the face masked and another with the upper part of the face masked.

Previous studies have excessively focused on identifying the facial features necessary for facial expression analysis in well-known datasets. This approach only pursues technological improvements in facial expression recognition and does not consider the effect of this facial expression recognition technology on humans. Considering the recent social phenomenon of wearing masks, research is being conducted on the upper and lower facial regions. This study investigated the effect of facial expression recognition technology on people in the current situation of limited faces; the analysis proves that it is necessary at the right time. However, research has been conducted to specify the part that plays an important role in a specific emotion by limiting it to the face wearing a mask or by mixing the upper and lower parts. Expression recognition of the upper and lower facial regions is independently analyzed; studies that calculate the accuracy of each facial expression have not been conducted, and studies that analyze it are lacking. In this study, the accuracy of each of the eight facial expressions are calculated using only the upper and lower facial regions. We analyze how limited facial information affects facial expression recognition.

## 2.2. Face Identification of Upper and Lower Facial Regions

Research on face identification by dividing the face area has been conducted in many forms, especially after the COVID-19 pandemic; most studies have focused on masking the dataset. A brief summary of these studies is presented in Table 2. Deng et al. performed mask face recognition using large-margin cosine loss (MFCosface) [13]. Due to the lack of data for the model, a dataset was built using a mask-generating model. After finding a face using a multitask cascaded convolutional network (MTCNN), a mask was applied through feature-point extraction. Using the large-margin cosine function, we mapped a feature space with a smaller intraclass distance and a larger interclass distance. We designed the Att-Inception module, which combines the Inception-ResNet module and the convolutional block attention module, and performed face recognition. In this study, the face alignment of the dataset in face recognition maintained the masking effect but reduced the computation required [13]. In a previous study, the dataset was augmented by synthesizing various masks on the face of the same person. However, this cannot rule out the influence of the type of mask on accuracy. This study uses feature points for face segmentation in the dataset, as in the previous study. However, the upper and lower parts are used independently to eliminate mutual influence. This focuses on the analysis of each area of the face.

**Table 2.** Summary of face-identification-related works.

Method	Summary	Limitations
Deng et al. [13]	This uses MTCNN to synthesize the mask on the face. Face recognition is performed using the Att-Inception module.	An effect of mask type on accuracy cannot be ruled out.
Mukhiddinov et al. [14]	After converting the AffectNet data set into a high-contrast image, the experiment was conducted. After generating a mask with the MaskTheFace algorithm, the result was derived through CNN.	Face recognition accuracy cannot be guaranteed in a wild environment through high-contrast conversion.
Pann et al. [15]	The image preprocessed through CBAM and ArcFace was used for upper part recognition through mask synthesis.	Only the upper part of the face was studied.
Stajduhar et al. [16]	After presenting faces in forward and reverse directions, face recognition was performed on children.	Only male faces were used as the dataset.

Research on face identification by area is in progress, and analysis is being performed on well-known datasets. Mukhiddinov et al. used the AffectNet dataset. Their study went through the conversion of low-contrast images into high-contrast images for data preprocessing [14]. They also used the MaskTheFace algorithm to mask images. The upper part was defined by detecting the area of the eyes and eyebrows using landmark extraction, after which, a CNN was used to conduct learning. As a result of the learning, the accuracy of the dataset was 69.3% [14]. In this study, MS-Celeb-1M and CASIA-Webface, which are also well-known data sets, are used. However, in the data preprocessing process, the three original characteristics were not damaged. In previous studies, the accuracy of the recognition rate was obtained by changing the image quality; however, in this study, only the angle is preprocessed while preserving the properties of the images in the face recognition dataset. Therefore, we conducted an analysis using a wider dataset.

It is possible to conduct face identification studies by region on a larger dataset, and studies focusing on improving the accuracy have been conducted. Pann et al. proposed the convolutional block attention module (CBAM) and angular margin ArcFace loss to improve the accuracy of general facial recognition methods [15]. CBAM was integrated with a CNN to extract the input image feature map of the area around the eye. ArcFace was used to optimize feature embedding for masked face recognition and improve the discriminant



features. We also use a data augmentation method to generate masked facial images from a typical facial recognition dataset. Data generated from the LFW, AgeDB-30, and CFP-FP datasets are used as the MFR2 verification datasets; this confirms the improvement in accuracy [15]. As in a previous study, LFW and AgeDB-30 are used as test data. However, studies thus far have focused on the upper part of the face and the calculated accuracy, but not on the lower part. In this study, the recognition rates of the upper and lower parts of the face are analyzed simultaneously, focusing on the differences in the recognition rate according to the study, which explored the effect on people.

However, another study has analyzed the recognition rates of the upper and lower facial regions. Stajduhar et al. measured face recognition rates in children aged 6 to 14 [16]. After presenting faces in the forward and reverse directions, with or without a mask, face discrimination was performed. Children's ability to recognize faces deteriorated when presented with a face wearing a mask; this decline was greater in children than in adults. These results provide evidence of significant quantitative and qualitative changes in masked face processing in school-aged children. Similarly, this study also revealed the effect of the upper and lower facial regions of a person's face on the recognition rate and presented evidence of its influence on children's social problems [16]. As in the previous study, we also examine the effect of the difference in face identification according to the facial area on humans. In the previous study, only male faces were examined, and various faces were not studied. In other words, no study has performed an analysis while simultaneously satisfying the generality of the dataset and face area at the same time. To proceed with our research, we analyzed the effect of identification by facial region on a person in a reliable dataset.

Face recognition proceeds in various experimental environments and yields high accuracy; however, as in facial expression recognition research, only studies on faces wearing masks have been conducted. In addition, the effect of partial face recognition on children is investigated. Through this, it is possible to determine the effect of the upper part of the face on a child's psychology. However, there is a lack of research comparing the recognition rate by dividing the face into upper and lower parts, and there is no study that comparatively analyzes face and expression recognition. In this study, face recognition is performed by dividing the upper and lower parts of the face and analyzing the differences between the two parts. The analysis is performed together with facial expression recognition.

### 2.3. Models

Face identification and facial expression recognition are highly researched areas, and new models with state-of-the-art (SOTA) performance are constantly being developed. However, the focus of our study is not to create a new face identification or facial expression recognition model with the highest performance but to compare the impact of different facial areas. To achieve this, we utilized an existing high-performance model for face identification and facial expression recognition.

Face identification can be divided into two categories: close-set and open-set [17]. Close-set recognition involves recognizing people within a specific group, whereas open-set recognition involves matching people not in the training set. Learning proceeds to solve this problem by embedding facial images into a discriminative feature space using methods such as triplet loss. A typical example is FaceNet [18], released by Google in 2015, which uses triplet loss for learning. Instead of considering the absolute distance, the triplet loss learns the difference in the relative distance. It learns by minimizing the distance between images with the same ID as the center image and maximizing the distance between images with different IDs. This study has the disadvantage of using large amounts of private data and numerous computer sources; therefore, the results are not reproducible. VGGFace [19], released in 2015, achieves similar performance with less data than FaceNet and achieves 99% accuracy; it has also made its dataset public. Subsequent studies on face identification focused on datasets, preprocessing methods, and loss functions. ArcFace [20], released

in 2018, is a representative example of a study focusing on loss functions. It locates the feature-embedding values derived from the image data in a spherical surface space and minimizes the spherical angle of the data for the same person. This approach prevents the model from converging toward its origin and still exhibits state-of-the-art performance in face verification. The equation for calculating the ArcFace loss is shown in Equation (1).  $W_{y_i}$  represents the representative vector of the correct answer class and  $W_{j \neq y_i}$  represents the representative vector of the non-correct class.

$$L_{arcface} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s(\cos(\theta_{y_i} + m))}}{e^{s(\cos(\theta_{y_i} + m))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}} \quad (1)$$

Equation (1) can be explained using Figure 1, where we assume that there is a representative vector  $W_{y_i}$  for the correct answer class, a representative vector  $W_{j \neq y_i}$  for the non-correct answer class, and an image-embedding vector  $X_i$  that is being processed. The distance between the image and the correct answer is denoted as  $\theta_{y_i, i}$ , whereas the distance from the incorrect vector is denoted as  $\theta_{j, i}$ . ArcFace aims to minimize the distance  $\theta_{y_i, i}$  between the image-embedding and the correct answer classes while simultaneously maximizing the distance  $\theta_{j, i}$  between the image-embedding and the incorrect answer classes. This is done by introducing a margin  $m$ , such that learning is performed in a way that  $\theta_{j, i}$  is larger than  $\theta_{y_i, i} + m$ . This approach ensures that the intra-class distances are kept very small, whereas the inter-class distances are increased, thereby improving the overall performance of the model.

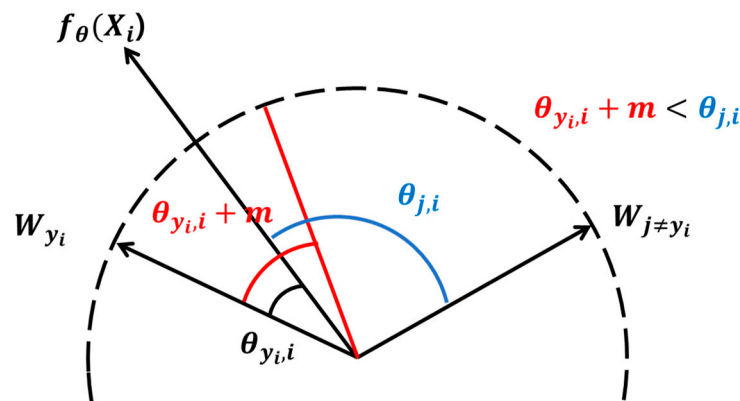


Figure 1. Explanation of ArcFace in image.

Therefore, our study applied the ArcFace model to ResNet50 [21] to obtain the face identification results. The overall architecture of face identification is shown in Figure 2. Figure 2a shows an example of the same person’s image; in contrast, Figure 2b shows an example of a different person’s images. The model attempts to update its weights to minimize angle  $\theta_1$  and maximize angle  $\theta_2$ .

The dataset used for training is MS-Celeb-1M [22], which consists of approximately 10 million photos of approximately 1 million celebrities. It is one of the most widely used datasets for facial identification training.

Face-expression recognition classifies N target expressions. The value of N depends on the model used; however, the most widely used dataset in recent years is AffectNet [23]. AffectNet is a large-scale dataset for facial expression recognition with approximately 0.4 million pieces of manually labeled data. Eight classes were used: neutral, happy, angry, sad, fear, surprise, disgust, and contempt; they include valence and arousal intensity. Currently, the best-performing SOTA model on the AffectNet dataset is EmoNet [24], with 75% accuracy for eight facial expressions. This is higher than the performance of Multitask EfficientNet-B2 [25], (63.06%) which is expressed as SOTA in paperswithcode [26]. In the case of EmoNet, when performing facial expression recognition, it classifies expressions and

learns regressions using valence and arousal together. It also learns face landmarks together, so that the features of each area of the face are used together and learned simultaneously to improve the overall performance. A brief structure of EmoNet is shown in Figure 3. We used the EmoNet model to reason about the results of the expression.

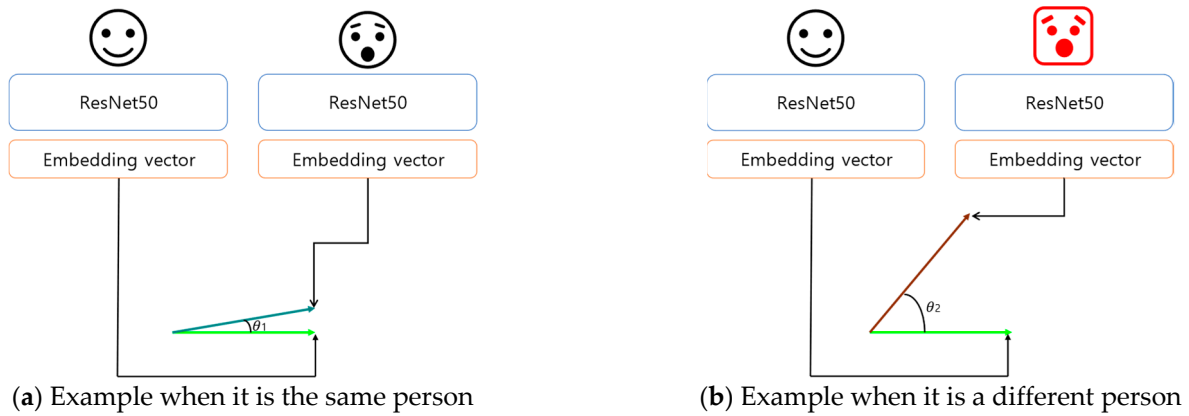


Figure 2. A brief architecture of face identification.

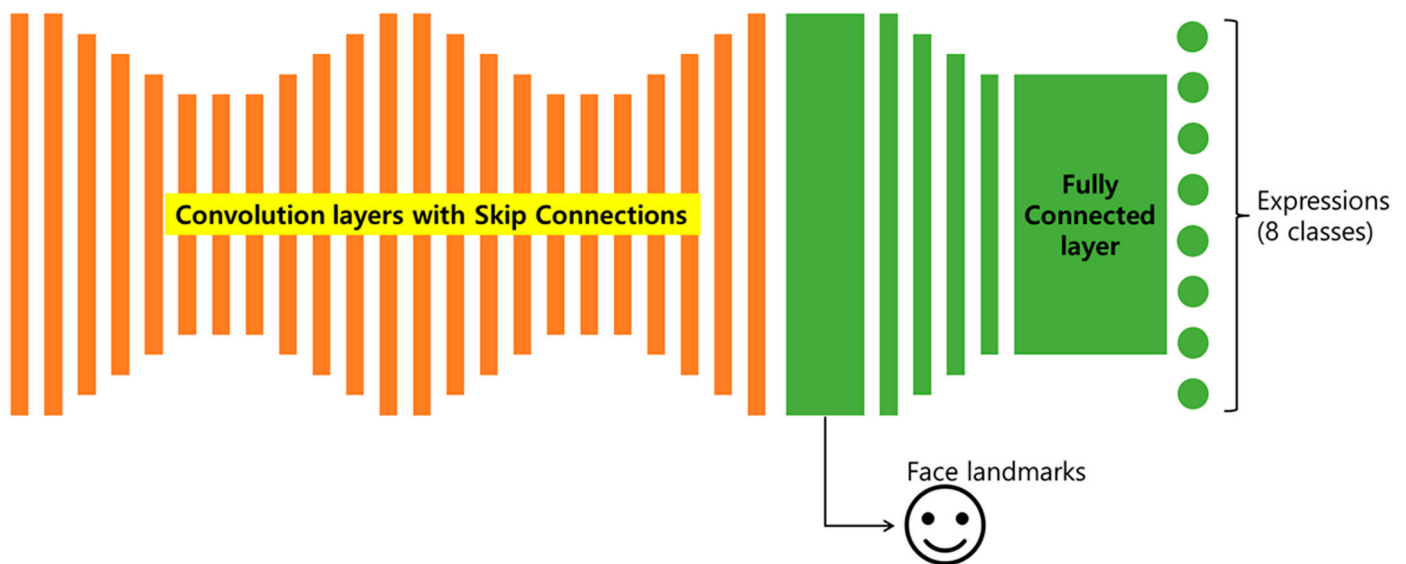


Figure 3. Brief architecture of EmoNet for facial expression recognition.

### 3. Methods

#### 3.1. Preprocessing Dataset

Facial alignment is necessary for facial identification and facial expression recognition by dividing the upper and lower facial regions.

When face alignment is not processed, an imbalance in the face area occurs during the face division process, as shown in Figure 4. Mounting the dataset, including the imbalanced face area, onto a learning model decreases the reliability of the results. Therefore, the face feature points were determined using the multitask cascaded convolutional network (MTCNN) model [27] to perform face alignment. MTCNN is a method of calculating the loss for classification, regression, and localization, and then adding them by weighting them. The feature points of the face obtained by the MTCNN are shown in Figure 5.

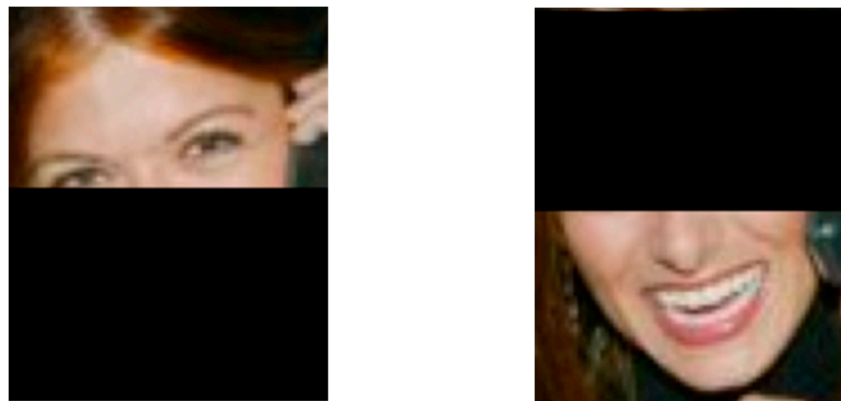


Figure 4. Example of cropped upper and lower facial regions without face alignment.



Figure 5. Facial feature points found by the MTCNN model (central point of each eye, tip of nose, end of mouth).

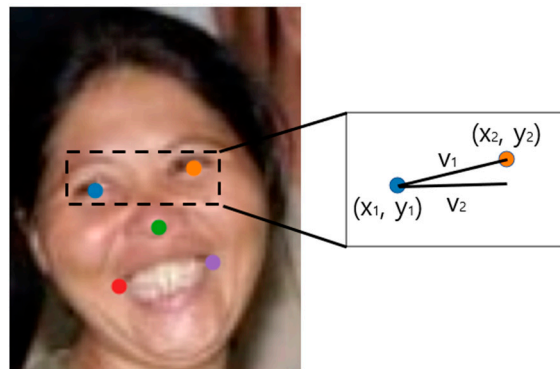
There are five facial characteristics: the center point of each eye, tip of the nose, and tip of the mouth. The face angle was adjusted using the coordinates of the center point of each eye.

As shown in Figure 6, the coordinates of the center points of the left and right eyes are defined as  $(x_1, y_1)$  and  $(x_2, y_2)$ , respectively. The vector calculated using Equation (2) is defined as  $V_1$ . We also define a vector parallel to the x-axis, calculated using Equation (3), as  $V_2$ . Finally, the angle between  $V_1$  and  $V_2$  is calculated using Equation (4): Face alignment is performed using the angle obtained from the following calculation:

$$V_1 = (|x_2 - x_1|, |y_2 - y_1|) \tag{2}$$

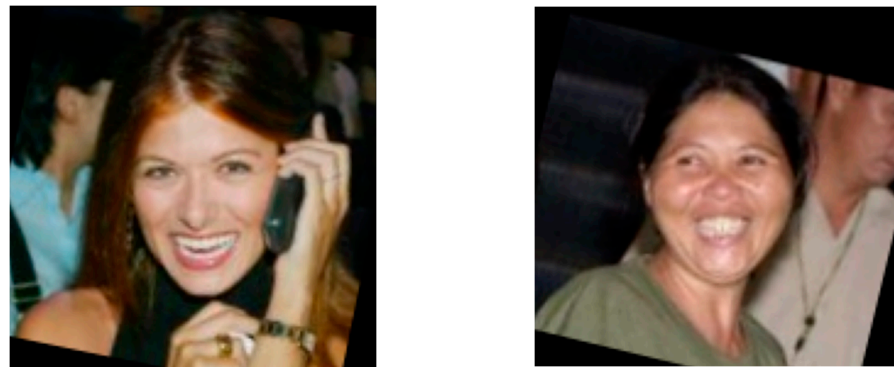
$$V_2 = \begin{cases} (x_1, 0), & (y_2 > y_1) \\ (x_2, 0), & (y_2 < y_1) \end{cases} \tag{3}$$

$$\begin{aligned} a &= |V_2| \\ b &= |V_1| \\ c &= |y_2 - y_1| \\ \cos a &= (b^2 + c^2 - a^2) / (2 * b * c) \\ \theta &= \begin{cases} 90 - \arccos(\cos a), & (y_2 > y_1) \\ \arccos(\cos a), & (y_2 < y_1) \end{cases} \end{aligned} \tag{4}$$



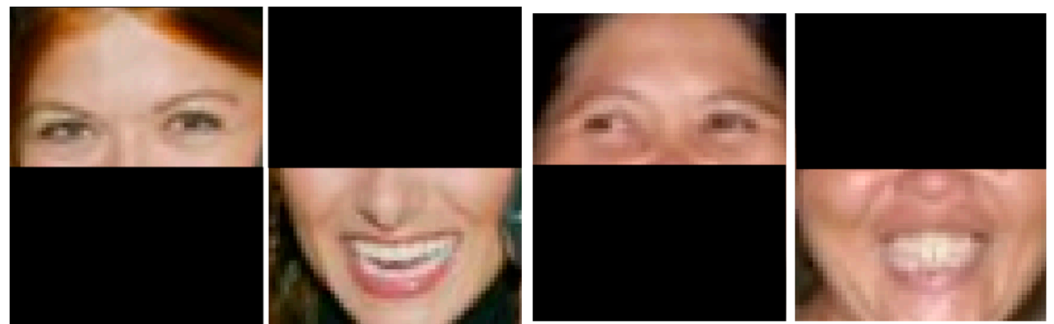
**Figure 6.** Set vectors and variables using the center point of the eyes.

The face angle aligned result is shown in Figure 7.



**Figure 7.** Result of face alignment.

Finally, after the face-area-cropping operation, each image is saved using only the upper and lower facial regions. Example images are shown in Figure 8.



**Figure 8.** Examples of data preprocessing for use on the upper and lower facial regions.

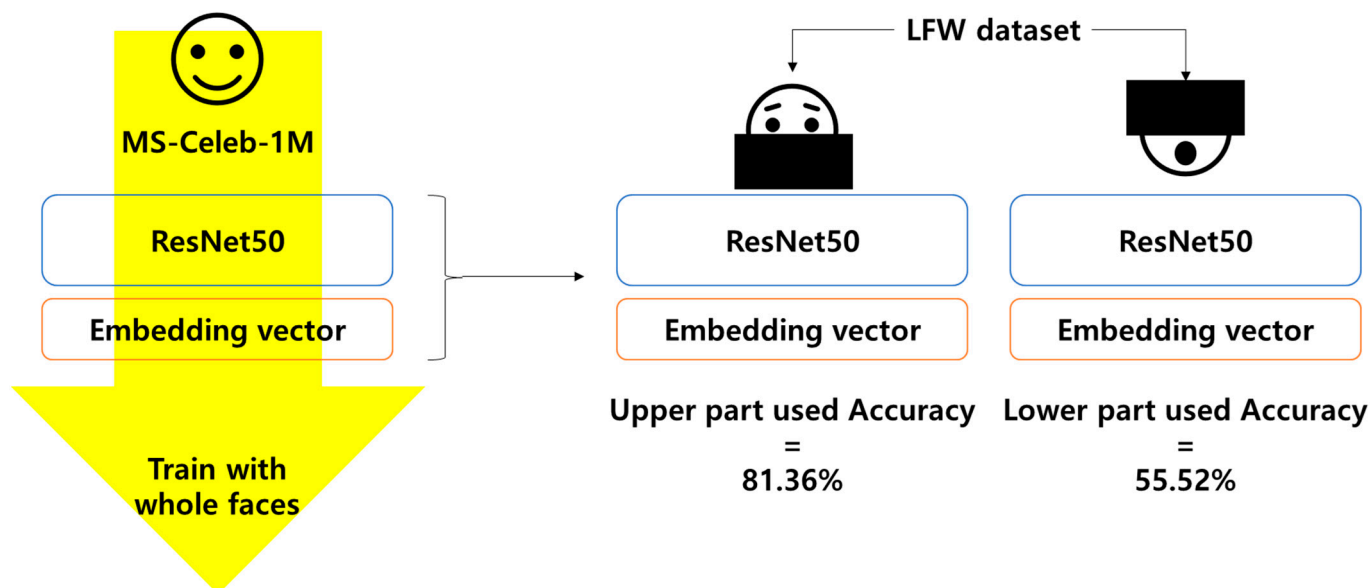
The above process was applied to the Labeled Faces in the Wild (LFW) [28] and AffectNet datasets used for facial recognition and facial expression recognition.

### 3.2. Face Identification

The datasets widely used for face identification include MS-Celeb-1M and CASIA-Webface [29] for training data and LFW and aged30db [30] for test data. Because this is a study concerning which part of a person's face affects the identification or recognition of facial expressions, the accuracy is derived from the test data divided into parts from the pretrained model of the whole face. First, the widely used ResNet50 model is used for face identification, followed by a model trained with MS-Celeb-1M data using ArcFace. The



data used for the upper facial region showed an accuracy of 81.36%, and the data used for the lower facial region showed an accuracy of 55.52%. The two resulting accuracies were extracted by controlling all conditions equally, except for the test dataset, and the face identification rate was confirmed based on the part of the face. The experimental setup is shown in Figure 9.



**Figure 9.** Overall experimental setup of face identification.

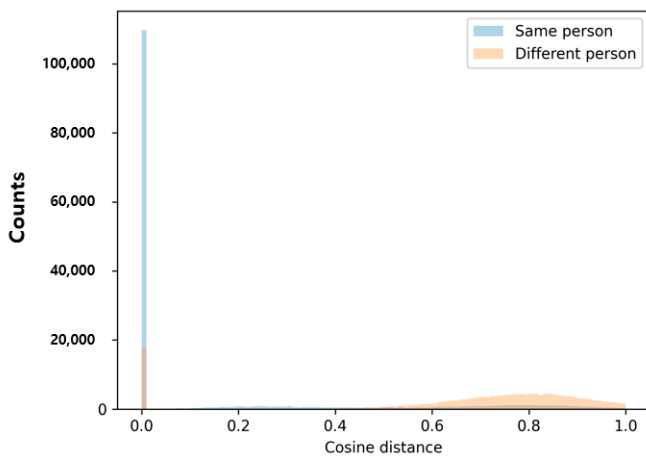
### 3.3. Facial Expression Recognition

Among the data used for facial expression recognition, AffectNet consists of 0.4 million manually labeled images. In the case of the data, eight expressions are labeled, and the intensity of the expression is expressed as positive and negative through valence and expressed as excitement and calm through arousal. The dataset was classified into training and test data, and the EmoNet model was used as the learning model. The accuracy of AffectNet, which is currently the highest-performing model based on studies with code, is approximately 63%, based on expressions. In this study, the resulting values of the top and bottom of the face were separately extracted from the test set among the datasets of the same AffectNet dataset using the same cropping process as mentioned in Section 3.1. When the upper half of the face was used, the performance was approximately 39%, and when the lower half was used, it was 49%, showing a significant difference of approximately 10%. Similarly, in the case of facial expression recognition, the experiment was conducted by controlling all conditions of the model, except for the test dataset, in the same way. Thus, it was confirmed that the model focused more on the lower half of the face when recognizing facial expressions.

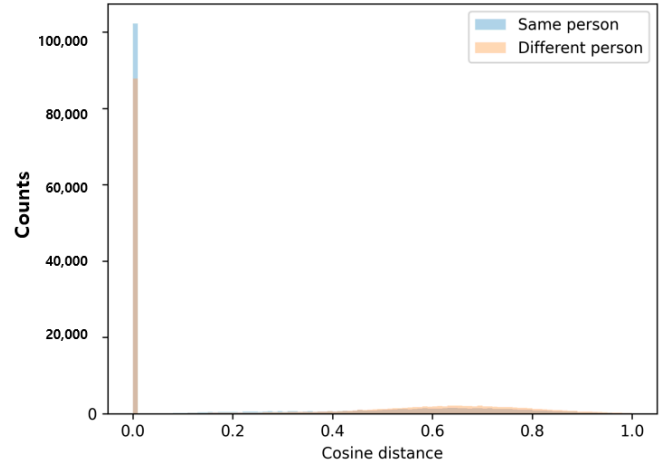
## 4. Results

For face identification, the results were extracted in the same manner as in LFW. The prediction of whether the ID of the two faces is the same or different is accurately expressed. Figure 10 shows the histograms for the two test situations, where the x-axis represents the cosine distance of the embedding vectors of the two face images. Blue bars indicate same-person comparisons, and red bars indicate different-person comparisons. It can be observed that the cosine distance is short for most same-person comparisons and long for most different-person comparisons, although there is some overlap. Interestingly, a significant difference was observed when the upper facial region was used. Specifically, the histograms in Figure 10c,d show a larger separation between the blue and red bars when using the upper facial region than when using the lower facial region. This suggests that

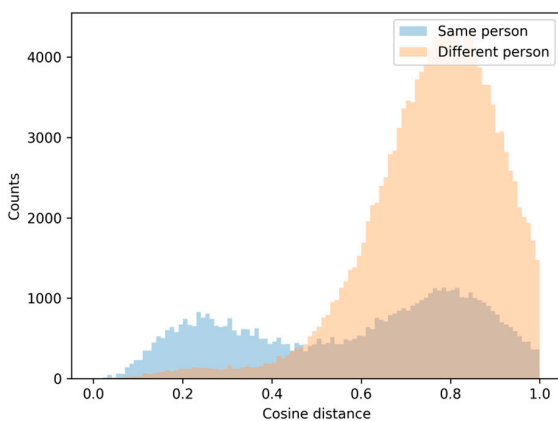
the upper facial region plays a more critical role in facial recognition than the lower facial region. Figure 10e,f show the receiver operating characteristics (ROC) curves for the two test situations. These graphs show that the upper facial region is more suitable for face identification.



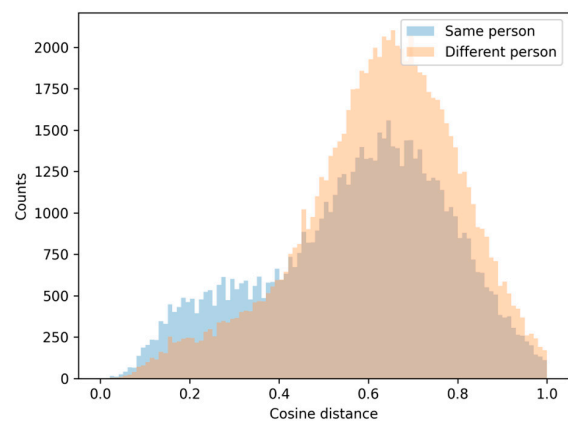
(a) Genuine and imposter graph of upper part of face [0.0~1.0]



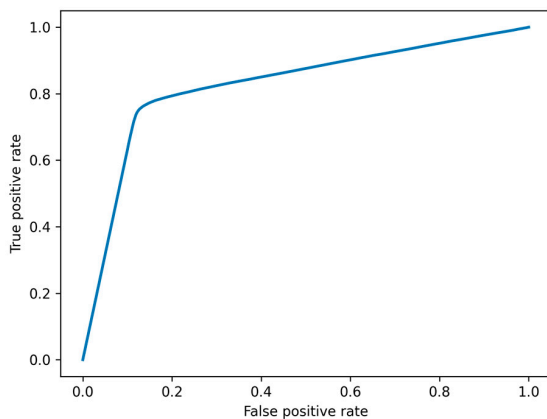
(b) Genuine and imposter graph of lower part of face [0.0~1.0]



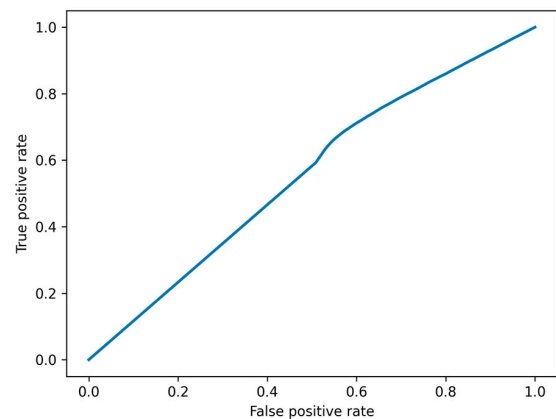
(c) Genuine and imposter graph of upper part of face (0.0~1.0)



(d) Genuine and imposter graph of lower part of face (0.0~1.0)



(e) ROC curve of upper part of face



(f) ROC curve of lower part of face

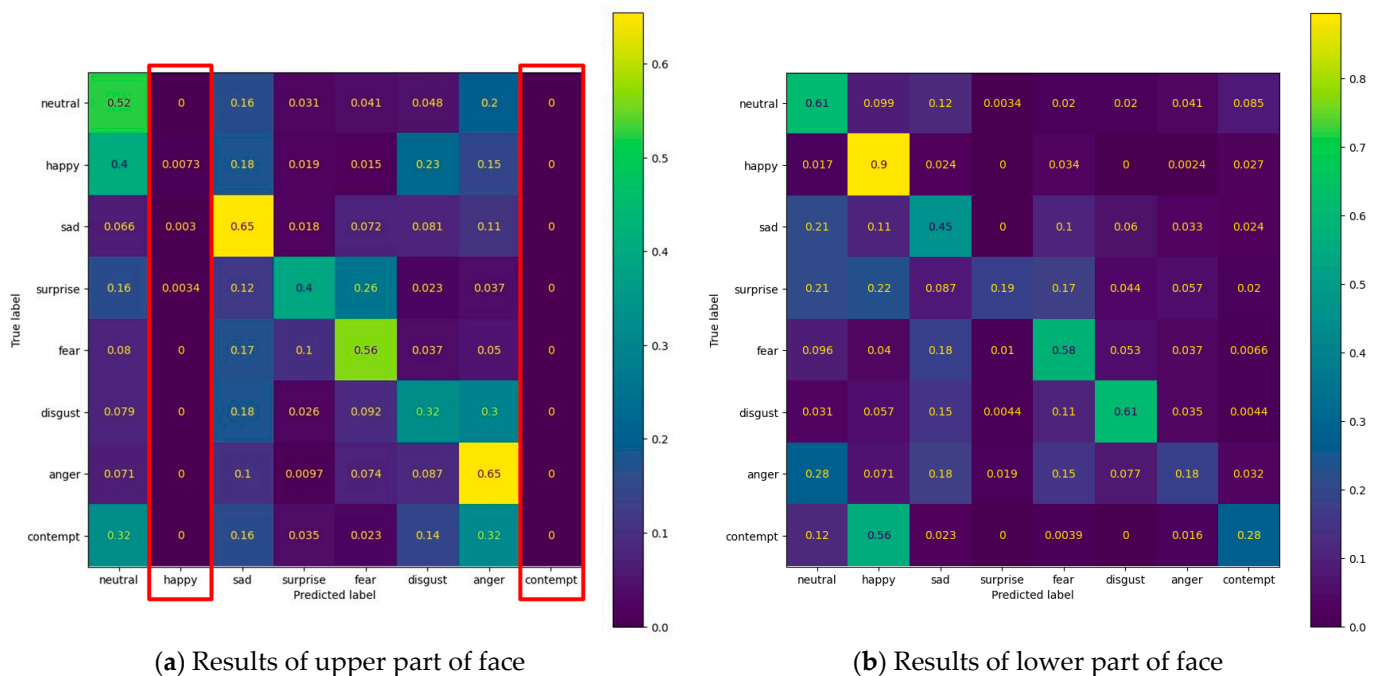
Figure 10. Accuracy of detailed facial identification by part of the face.

Table 3 shows that the face recognition rate is much higher when using the upper facial region than when using the lower facial region. Accordingly, we can see why people are able to easily recognize one another during COVID-19, even while wearing masks. In the case of facial expression recognition, the results were contrary to those of facial recognition. The recognition rate was approximately 10% higher when the lower facial region was used than when the upper facial region was used. In other words, the accuracy of using the upper facial region was approximately 1,465 times that of using the lower facial region.

**Table 3.** Accuracy of face identification and facial expression recognition according to the upper and lower facial regions.

	Upper Parts	Lower Parts
Face identification	81.36%	55.52%
Facial Expression Recognition (8 classes)	39%	49%
Facial Expression Recognition (2 classes)	51%	88%

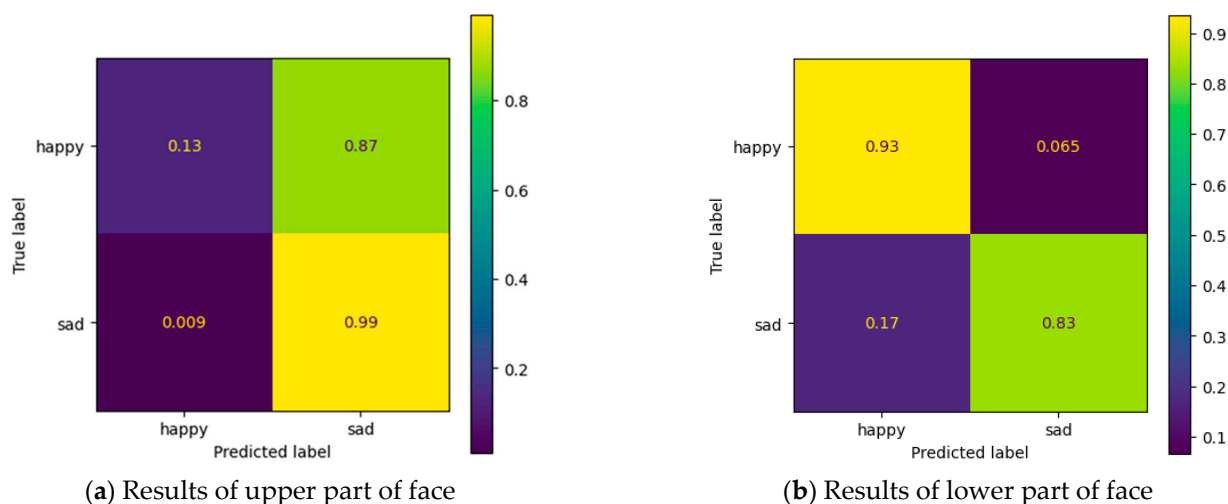
To analyze the results in more detail, the recognition rate for each facial expression was extracted. Figure 11a shows the recognition rate for each facial expression when using the upper facial region, and Figure 11b shows the recognition rate for each facial expression when using the lower facial region. In the case of the two red-boxed facial expressions in Figure 11a, it can be seen that they converge to almost zero. When the learned model recognizes only the upper facial region, it tends to fail to predict the facial expressions of happiness and contempt. Conversely, in the case of sadness and anger, it can be seen that the accuracy is higher when the upper facial region was used compared to the lower facial region. We can observe that a person is more conscious of the upper facial region when determining whether an emotion is sadness or anger.



**Figure 11.** Accuracy of detailed facial expression recognition by part of the face. Red boxed expressions represent expressions that can not be recognized with upper part of the face.

Because facial identification is a binary classification, and facial expression recognition is a problem in classifying eight classes, we assumed that a one-to-one comparison of the two fields would also be necessary. Performance was derived by defining sadness, the most representative facial expression, as a binary classification problem. The accuracy of

the upper facial region was approximately 51.54%, whereas when using the lower facial region, it was 88.85%. This indicates that using the lower facial region had approximately 1.256 times better performance than using the upper facial region. The figure is divided into true-positive, true-negative, false-positive, and false-negative areas, as shown in Figure 12.



**Figure 12.** Accuracy of detailed facial expression recognition in binary classification.

## 5. Discussion

The results of this study can be analyzed from two perspectives. From the perspective of facial identification, the overall performance of the upper facial region was approximately 1.465 times that of the lower facial region. The cosine distance histograms of the embedding vectors for the same person and for different people are shown in Figure 10c,d, respectively. These graphs show that the largest differences occur when the embedding vectors are from different people. The red histogram shows that the highest amplitude when using the upper face appears at 0.8 and shows a magnitude of approximately 4000. In addition, it can be confirmed that most of them are distributed over 0.5. However, as shown in Figure 10b,d, when using the lower face, the embedding vectors from different people are also most heavily skewed towards zero. Thus, we can conclude that using only the upper facial region is more advantageous for facial identification than using only the lower facial region. In fact, during the 2020 COVID-19 pandemic, most people were able to recognize each other despite wearing masks, and various studies have shown that face identification is possible even if a mask covers the lower facial region. This also applies to Face ID on Apple devices, which shows that the most important information for face identification is distributed in the upper facial region.

As mentioned in previous related studies, facial expressions have a common face shape. These are represented by action units: when most people are happy, their cheeks rise and the corners of their mouths turn up. When people are angry, their eyebrows are furrowed and their lips are pursed [31]. Thus, we can understand a person's expressions through several action units that appear on their face. In the cases of contempt and happiness, because the action unit identified from the mouth is larger, it cannot be observed from the upper facial region but only from the lower part. In contrast, in the cases of sadness and anger, prominent features such as squinting or watering eyes are more easily observed. Therefore, Figure 11 shows that in the cases of sadness and anger, the performance was better when the upper facial region was used than when the lower facial region was used. Several emotions exist for each facial expression where the lower facial region is advantageous and others where the upper part is advantageous. However, in our society, it is important to quickly understand whether an expression is positive or negative, such as happy or sad, to make interpersonal connections. Therefore, we performed a binary classification for happiness, which represents positive emotions, and sadness, which represents negative emotions.

Interestingly, the results were 51% accurate when the upper facial region was used and 88% accurate when the lower facial region was used. This can be considered as the opposite of the difference in degrees, which is similar to the value of face identification. From this, we can see that the upper facial region contains a large amount of information related to human identification; in contrast, the lower facial region contains a large amount of information about facial expressions, with a significant difference. In other words, owing to the bias of facial expression information, children are confused when judging the emotions of another person wearing a mask. As mentioned in the introduction, we assumed that our deep learning models were representative of a diverse sample of people. It is true that we did not train the models with data labeled by children, which can make it difficult to draw a clear correlation with the social development of children. However, as with deep learning models, children learn from the socially constructed rules and labeled data of adults. This suggests that deep learning models and young children go through a similar learning process. Therefore, we can present an analysis of children's behavior through the results of the deep learning model trained in this study. Thus, due to the masking phenomenon, children developed problems with empathy, which in turn led to problems in social development. Through the analysis of face identification and facial expression of masked faces, it can be concluded that problems related to children's social development during the COVID-19 pandemic do not concern confusion regarding the identification of the subject of an encounter, but rather, they concern the lack of facial expressions.

## 6. Conclusions

In this study, the upper and lower facial regions were divided, and the identification and expression recognition of the face included in each region were investigated. The LFW and AffectNet datasets were preprocessed through face alignment. Face identification was performed using ResNet50 and ArcFace models, and facial expression recognition was performed using EmoNet. In the case of face identification, the accuracy was 81.36% when only the upper facial region was used and 55.52% when only the lower region was used. This proves that there are no problems in distinguishing a person from a face when wearing a mask. However, in facial expression recognition, an accuracy of 39% was derived when only the upper part was used and 49% when only the lower part was used; in facial expression recognition, the lower facial region is more advantageous than the upper part. When the recognition rate of each of the eight facial expressions was derived, happiness and contempt could not be distinguished solely by the upper facial region. Although previous studies classified the results of facial expressions into eight classes, it was thought that the results of the binary classification of facial expressions into positive and negative facial expressions were also necessary for the basic emotional development of children. Therefore, we performed binary classification with two expressions—happy, the most representative positive expression, and sad, the contrasting expression—and derived the corresponding performance. Consequently, 88.85 percent accuracy was obtained when only the lower portion was used, whereas 51.54 percent accuracy was obtained when only the upper portion was used. As mentioned earlier, the results of this study can be seen as a technological underpinning for identifying trends in society, as deep learning models focus on the human-like aspects of facial identification and facial expression recognition. Therefore, based on this result, we can assume that children who encounter a mask-wearing person are confused about the distinction between the negative and positive expressions on their face, suggesting grounds that can affect emotional development. Further research can be conducted on this topic, such as analyzing in detail which parts of the face are most affected by each facial expression, and a facial expression recognition model with attention that focuses on the parts of the face that contain the least amount of human identity information.

Since our study only distinguished between upper and lower faces, it can be seen that we used a broader subset than action units. This will be the subject of future research. Additionally, we tested with cropped faces in a model trained on the whole face. While



this explains the initial situation of the COVID-19 pandemic, it is possible to extend the study by fine-tuning the model with cropped faces in the case of a prolonged pandemic, or by comparing the results of the model trained on new data. This will be the subject of our future work.

**Author Contributions:** Conceptualization, E.C.L.; methodology, S.K., B.S.A. and E.C.L.; validation, S.K. and B.S.A.; investigation, S.K. and E.C.L.; data curation, B.S.A.; writing—original draft preparation, S.K. and B.S.A.; writing—review and editing, E.C.L.; visualization, S.K. and B.S.A.; supervision, E.C.L.; project administration, E.C.L.; funding acquisition, E.C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a 2022 Research Grant from Sangmyung University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Since the data used in this study is a public open dataset, it can be used by contacting the data holder.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Schneider, J.; Sandoz, V.; Equey, L.; Williams-Smith, J.; Horsch, A.; Graz, M.B. The Role of Face Masks in the Recognition of Emotions by Preschool Children. *JAMA Pediatr.* **2022**, *176*, 96. [CrossRef] [PubMed]
2. Philippot, P.; Feldman, R.S. Age and social competence in preschoolers' decoding of facial expression. *Br. J. Soc. Psychol.* **1990**, *29*, 43–54. [CrossRef] [PubMed]
3. Ejaz, S.; Islam, R.; Sifatullah; Sarker, A. Implementation of Principal Component Analysis on Masked and Non-masked Face Recognition. In Proceedings of the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology, Dhaka, Bangladesh, 3–5 May 2019; pp. 1–5. [CrossRef]
4. Tian, Y.-I.; Kanade, T.; Cohn, J. Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 97–115. [CrossRef] [PubMed]
5. Park, S.; Wallraven, C. Comparing Facial Expression Recognition in Humans and Machines: Using CAM, GradCAM, and Extremal Perturbation. In Proceedings of the Pattern Recognition: 6th Asian Conference, ACPR 2021, Jeju Island, Republic of Korea, 9–12 November 2021. [CrossRef]
6. Nam, H.-H.; Kang, B.-J.; Park, K.-R. Comparison of Computer and Human Face Recognition According to Facial Components. *J. Korea Multimedia Soc.* **2012**, *15*, 40–50. [CrossRef]
7. Comparison of Human and Computer Performance across Face Recognition Experiments—ScienceDirect. Available online: <https://www.sciencedirect.com/science/article/pii/S0262885613001741> (accessed on 21 April 2023).
8. Chen, M.-Y.; Chen, C.-C. The contribution of the upper and lower face in happy and sad facial expression classification. *Vis. Res.* **2010**, *50*, 1814–1823. [CrossRef] [PubMed]
9. Abrosoft FantaMorph—Photo Morphing Software for Creating Morphing Photos and Animations. Available online: <https://www.fantamorph.com/> (accessed on 21 April 2023).
10. Itoh, M.; Yoshikawa, S. Relative importance of upper and lower parts of the face in recognizing facial expressions of emotion. *J. Hum. Environ. Stud.* **2011**, *9*, 89–95. [CrossRef]
11. Seyedarabi, H.; Lee, W.-S.; Aghagolzadeh, A.; Khanmohammadi, S. Facial Expressions Recognition in a Single Static as well as Dynamic Facial Images Using Tracking and Probabilistic Neural Networks. *Adv. Image Video Technol.* **2006**, *4319*, 292–304. [CrossRef]
12. Khoeun, R.; Chophuk, P.; Chinnasarn, K. Emotion Recognition for Partial Faces Using a Feature Vector Technique. *Sensors* **2022**, *22*, 4633. [CrossRef] [PubMed]
13. Deng, H.; Feng, Z.; Qian, G.; Lv, X.; Li, H.; Li, G. MFCosface: A Masked-Face Recognition Algorithm Based on Large Margin Cosine Loss. *Appl. Sci.* **2021**, *11*, 7310. [CrossRef]
14. Mukhiddinov, M.; Djuraev, O.; Akhmedov, F.; Mukhamadiyev, A.; Cho, J. Masked Face Emotion Recognition Based on Facial Landmarks and Deep Learning Approaches for Visually Impaired People. *Sensors* **2023**, *23*, 1080. [CrossRef] [PubMed]
15. Pann, V.; Lee, H.J. Effective Attention-Based Mechanism for Masked Face Recognition. *Appl. Sci.* **2022**, *12*, 5590. [CrossRef]
16. Stajduhar, A.; Ganel, T.; Avidan, G.; Rosenbaum, R.S.; Freud, E. Face masks disrupt holistic processing and face perception in school-age children. *Cogn. Res. Princ. Implic.* **2022**, *7*, 9. [CrossRef] [PubMed]
17. Wang, M.; Deng, W. Deep face recognition: A survey. *Neurocomputing* **2020**, *429*, 215–244. [CrossRef]
18. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A Unified Embedding for Face Recognition and Clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823.

19. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the British Machine Vision Conference 2015, Swansea, UK, 7–10 September 2015; British Machine Vision Association: Swansea, UK, 2015; p. 41.
20. Deng, J.; Guo, J.; Yang, J.; Xue, N.; Kotsia, I.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 5962–5979. [[CrossRef](#)] [[PubMed](#)]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
22. Guo, Y.; Zhang, L.; Hu, Y.; He, X.; Gao, J. MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Leibe, B., Matas, J., Sebe, N., Eds.; Max Welling Springer: Cham, Switzerland, 2016; pp. 87–102.
23. Mollahosseini, A.; Hasani, B.; Mahoor, M.H. AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. *IEEE Trans. Affect. Comput.* **2017**, *10*, 18–31. [[CrossRef](#)]
24. Antoine, T.; Kossaifi, J.; Bulat, A.; Tzimiropoulos, G.; Pantic, M. Estimation of Continuous Valence and Arousal Levels from Faces in Naturalistic Conditions. *Nat. Mach. Intell.* **2021**, *3*, 42–50. [[CrossRef](#)]
25. Savchenko, A.V.; Savchenko, L.V.; Makarov, I. Classifying Emotions and Engagement in Online Learning Based on a Single Facial Expression Recognition Neural Network. *IEEE Trans. Affect. Comput.* **2022**, *13*, 2132–2143. [[CrossRef](#)]
26. The Latest in Machine Learning | Papers with Code. Available online: <https://paperswithcode.com/> (accessed on 11 April 2023).
27. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [[CrossRef](#)]
28. Huang, G.B.; Ramesh, M.; Berg, T.; Learned-Miller, E. *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments*; University of Massachusetts: Amherst, MA, USA, 2007.
29. Chen, X.; Duan, Y.; Houthoofd, R.; Schulman, J.; Sutskever, I.; Abbeel, P. InfoGAN: Interpretable Representation Learning by In-formation Maximizing Generative Adversarial Nets. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 2180–2188. [[CrossRef](#)]
30. Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; Zafeiriou, S. AgeDB: The First Manually Collected, In-the-Wild Age Database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1997–2005. [[CrossRef](#)]
31. Ekman, P.; Friesen, W. Facial Action Coding System: A Technique for the Measurement of Facial Movement. 1978. Available online: <https://www.paulekman.com/facial-action-coding-system/> (accessed on 23 April 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.