A Performance Study on Emotion Models Detection Accuracy in a Pandemic Environment

Priyadashini Saravanan[1(✉)], Suvendran Ravindran[1], Leong Yeng Weng[2(✉)], Khairul Salleh Bin Mohamed Sahari[3], Adzly Bin Anuar[1], Muhammad Fairuz Bin Abdul Jalal[1], Zubaidi Faiesal Bin Mohamad Rafaai[1], Prashalini Naidu A/P Raventhran[1], Husni Mohd Radzi[2], and Salman Yussof[2]

[1] Collage of Engineering (COE), Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia
{priyadashini,Suvendran,Adzly,mfairuz,Zubaidi,Prashalini}@uniten.edu.my

[2] Institute of Informatics and Computing in Energy (IICE), Universiti Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia
{ywleong,husni,Salman}@uniten.edu.my

[3] MQA, Jalan Teknkotrat 7, Cyber 5, 63000 Cyberjaya, Selangor, Malaysia
khairulsalleh@mqa.gov.my

Abstract. This paper studies emotion detection using deep learning on the prevalent usage of face masks in the Covid-19 pandemic. Internet repository data Karolinska Directed Emotional Faces (KDEF) [1] was used as a base database, in which it was segmented into different portions of the face, such as forehead patch, eye patch, and skin patch to be representing segments of the face covered or exposed by the mask were transfer learned to an Inception v3 model. Results show that the full-face model had the highest accuracy 74.68% followed by the skin patch (area occluded by the mask) 65.09%. The models trained on full-face were then used to inference the different face segments/patches that showed poor inferencing results. However, certain emotions are more distinct around the eye region. Therefore, this paper concludes that upper segmented faces result in higher accuracy for training models over full faces, yet future research needs to be done on additional occlusion near the eye section.

Keywords: Deep learning · Covid19 · Pandemic · Emotion

1 Introduction

Emotions are intuitive feelings that have great influence by the individual's circumstances (different cultural and ethnic backgrounds) [2]. The psychologist Paul Eckman has identified six discrete states of emotions in humans such as anger, sadness, disgust, happiness, fear, and surprise [3]. Individuals express these emotions via body language, verbal communication, and facial expression. Little did we know, fifty-five percent of effective communication generally constitutes face expressions [4]. Hence,

H. Badioze Zaman et al. (Eds.): IVIC 2021, LNCS 13051, pp. 322–331, 2021.
https://doi.org/10.1007/978-3-030-90235-3_28

facial expressions are crucial for daily social interaction as it possesses great importance in non-verbal communication. The movements of facial muscles (both upper and lower areas of the face) contribute to one's facial expression, With the specific facial expression, the emotion of the individual is being interpreted.

However, in the light of the current global pandemic Covid19, most countries advised their citizens to wear facial masks for everyone's safety and to minimize the risk of spreading the virus. Due to the mask mandate, covering the lower areas of the face below the eyes can be challenging for successful emotion recognition since only the forehead, eyebrows, and eye muscles will be the sole contributor to one's emotion recognition.

Since the coronavirus outbreak and the new norm of wearing face masks to protect ourselves, many researchers have studied the effect of the mask on facial and emotional recognition using various methods. M. Grahlow et al. conducted two different studies on the mask effect by using an adapted version of the Validated Emotion Recognition Task (VERT-K). In the first study, they digitally added surgical masks to the original facial stimuli, and they observed that emotion recognition was difficult when faces were covered with masks especially for angry, sad, and disgusted expressions. The second study was adapting VERT-K to cropped faces and photos of skin-toned bubbles obscuring the mouth and nose area. Emotions of fear and happiness were recognized more accurately when the only upper face was presented, and disgust has higher recognition rate for bubbled faces [5]. Hence, presenting the exposed face to the training model shows a better accuracy but not all emotions were recognized accurately.

Another study was conducted on train-test strategies in which the researcher found the result shows a higher accuracy when both training and testing data are occluded than using non-occluded data for the strategy. However, in this study, various occlusions were approached and not specifically to the lower face and upper face region [6].

Furthermore, several deep learning methods were also used to carry out face recognition for better performance [7]. One of the studies proposed cropping and attention-based approach for partially occluded faces for face recognition. The attention mechanism can significantly improve the recognition performance using model trained with Masked-Web Face datasets, but the performance is limited when testing is done on routine face recognition. The cropping method that uses integration of optimal cropping and Convolutional Block Attention Module (CBAM) module in ResNet50 network has better recognition on face-masked images [8]. Another face recognition studied by W. Hariri in which he discarded the masked region for training has concluded on high recognition performance on Real-World-Masked-Face-Dataset [9].

Even though many studies have been conducted on occluded faces and face recognition, the best performing method of cropping and discarding approach using deep learning on emotional recognition with face masks is still very limited. Training and testing on segmented face regions are essential as every section of our face contributes to a certain percentage of the emotions. Therefore, evaluating the performance of each face part is crucial in recognizing the emotion accurately on full faces and occluded faces.

This paper tries to fill this gap left out by previous research by involving and contributing to cropping and segmenting methods for training and testing data to evaluate the effect of masks occlusion on emotion recognition. The next section describes the methodology, on the overall environmental set-up and the details of segmenting the full faces into three different parts to obtain better accuracy on the recognition. Sect. 3 demonstrates the result and in-depth analysis. We compare the detection rate on full faces and segmented faces, thus further analyse the contribution of the face sections to the respective emotions. Finally, our findings are concluded and suggestion for future research is proposed in Sect. 4.

2 Methodology

2.1 Experimental Set-Up

Hardware. The environment used for this training model is Windows OS i7-8700 CPU with a clock speed of 3.20 Ghz, RAM of 16 GB, and NVIDIA GeForce GTX 1080ti GPU with 11 Gb RAM as a resource limitation and time required to train the classification model.

Software. In this experiment, MATLAB® was the only programming software used with pre-trained face detection MTCNN [10] for cropping the faces into three segments. As of the transfer learning, pretrained Inception-v3 convolutional neural network was used for image classification [11]. Inception-v3 is a 48 layers deep network that has an image input size of 299 × 299 [12]. This specific pretrained network is opted based on its speed, accuracy, and the compatibility of the available hardware. This model supported the system and was large enough to do the transfer learning.

Database. Karolinska Directed Emotional Faces (KDEF) database was used to train the training model. KDEF is a set of totally 4900 pictures of seven human facial expressions (anger, sad, surprise, fear, happy, disgust and neutral). Since this paper only focuses on the basic six emotions, neutral was excluded and only 4200 pictures were used for training and testing. To simply describe these KDEF subjects, they were not allowed to have beards, mustaches, wear earrings or eyeglasses, or visible make-up during the photo-session [1].

2.2 Method of Training

(See Fig. 1).

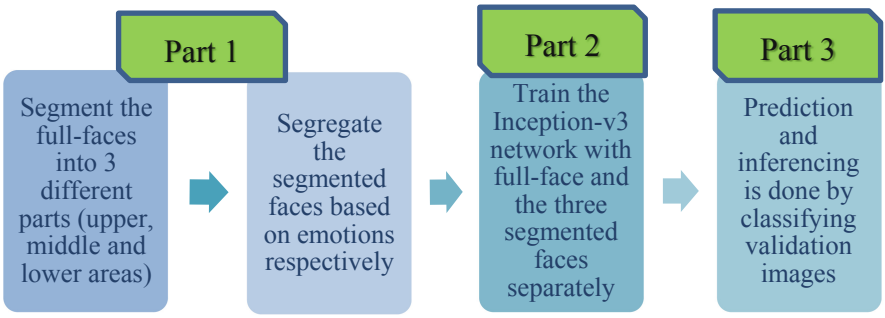


Fig. 1. The flowchart to represent the process of training the model and inferencing the validation images

Part 1. The full faces KDEF database was segmented into three different parts such as eye, forehead and skin patches using MTCNN [13]. Eye patch region is from forehead to nose which is the exposed part when face masks are being worn. Forehead region is from the top of the forehead till eyelids and the skin patch is from the nose till the lips. Figure 2 shows the examples of the KDEF full database that was cropped into three different parts. The three segmented faces are then segregated into six folders (emotions) respectively.

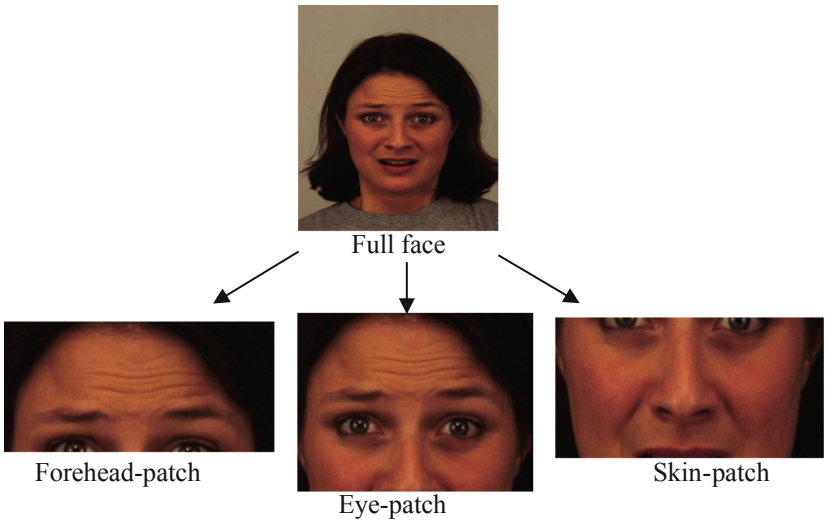


Fig. 2. Example of KDEF full-face that has been segmented into three parts

Part 2. To evaluate the full-face database on KDEF, training was done using the pre-trained Inception v3 network model (default augmentation settings) with 10 epochs, mini batch-size of 32 and the learning rate of 0.0001 using GPU for execution. The

cross-validation ratio for training and testing was 70:30. The same training method was done for all three patches respectively [14].

Part 3. To calculate the prediction accuracy, confusion matrix was computed. For classification, the validating images are resized to 299 × 299 to match the input size of the Inception-v3. The comparison is then made between patch trained vs full-face trained model on the respective patch datasets and full-face datasets via the confusion matrix.

3 Result and Discussion

The transfer learning performance for each training data is shown in Table 1.

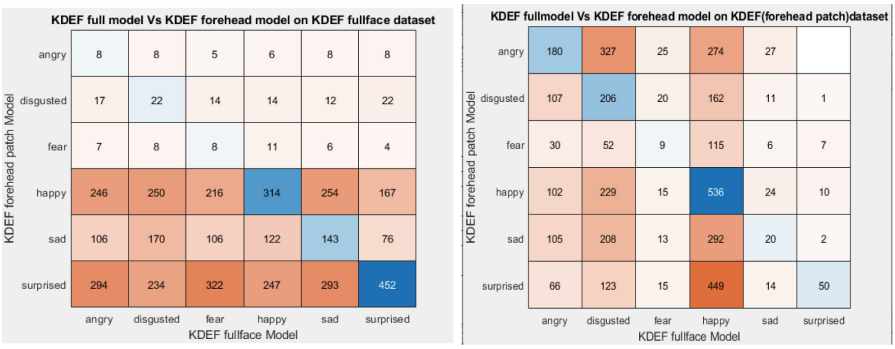
Table 1. Model transfer learning performance

KDEF model	Full face	Forehead patch	Eye patch	Skin patch
Performance (%)	74.68	44.09	51.65	65.09

Based on the above table, it can be clearly seen that the full face KDEF database has shown the highest performance over the patch model which is followed by skin patch, eye patch and the forehead patch. To further analyse its prediction on patch datasets and full-face datasets, these trained models are inferred by presenting various datasets and the true model (full face model) vs predicted model (patch model) on these datasets are populated using confusion matrix.

3.1 Discussion on Individual Patch Models vs Full-Face Model

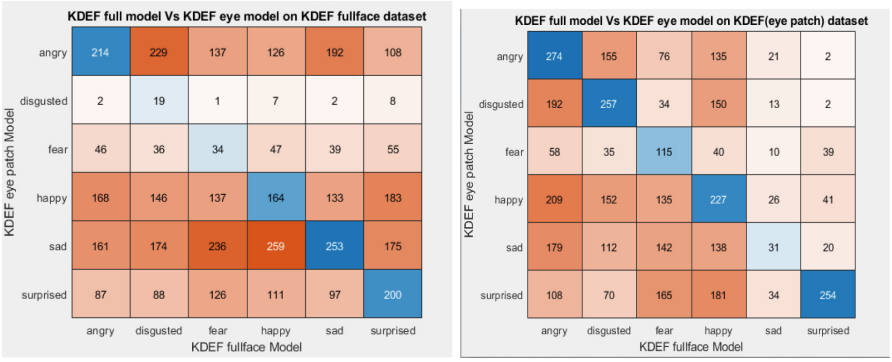
Confusion matrix is being plotted for the full-face trained model vs the patch models on the full-face datasets and its respective datasets of KDEF (Fig. 3).



Matrix 3(a) Matrix 3(b)

Fig. 3. Matrix 3(a) is the comparison between forehead-patch model and full model on the full-face dataset. Matrix 3(b) shows the result on validating forehead-patch datasets.

For the forehead patch model, the recognition on the forehead patch datasets is fairly better than the full-face datasets. The full-face model has least detection percentage and the forehead-patch model barely recognize angry, fear and disgust emotions when presented the full-face datasets, but has better detection rate for the sad, surprise and fear [Matrix 3a]. Meanwhile, the forehead-patch model recognizes the emotions better on forehead patch datasets, but the numbers are still not convincing since the full-face model solely recognizes more happy faces on forehead dataset [Matrix 3b] (Fig. 4).



Matrix 4(a) Matrix 4(b)

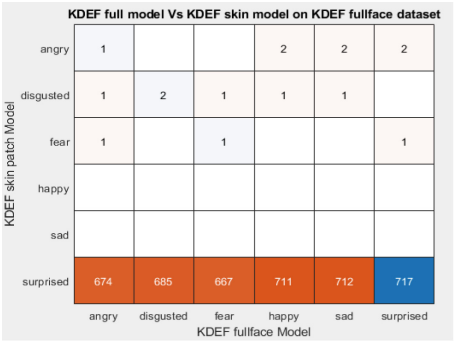
Fig. 4. Matrix 4(a) depicts the comparison between eye-patch model and full model on the full-face dataset. While Matrix 4(b) shows the result on validating eye-patch datasets.

When each emotion's prediction of the model is compared between the full-face datasets and the eye-patch datasets, the eye-patch model is able to predict five out of six emotions of the eye-patch datasets accurately whereas on the full-face datasets only four out of six are being more recognized. When we further analyze the matrix of eye-patch dataset, the full-face model was only able to detect 274/663 images (663 images are the total images identified by the full-face model) whereas the eye-patch model identified 663/700 as angry [Matrix 4b]. Further calculation using these values is done for detail analysis later.

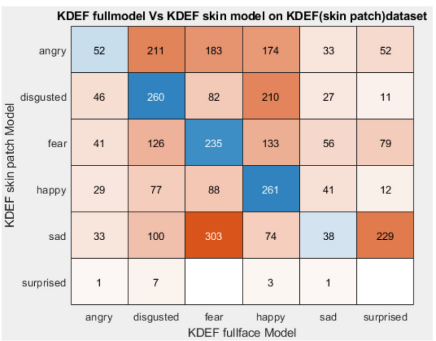
The eye patch model hardly captures disgust and fear on full face datasets [Matrix 4a] meanwhile it only had some difficulty on determining the sad emotion when presented the eye-patch dataset [Matrix 4b]. Overall, the performance of this eye-patch model on eye-patch datasets is quite convincing for five emotions (Fig. 5).

Based on Matrix 5a, it shows the full-face model on full–face dataset has the least detection on most emotions which is lesser than five percent. While the skin patch model on the full-face datasets is the worst in detection as it all the datasets prediction is saturated in one emotion but when the model is used to predict the skin-patch datasets, it seems to have a better detection rate but still has somewhat moderate [Matrix 5b].

There is no true value for the surprise emotion in Matrix 5b, but surprise is the only emotion that was detected by the skin-patch model when full-face dataset is presented. This is because surprise facial expression is highly perceived when the eyebrows elevate [14].



Matrix 5(a)



Matrix 5(b)

Fig. 5. Matrix 5(a) shows the comparison between skin-patch model and full model on the full-face dataset. Matrix 5(b) is the result on validating forehead-patch datasets.

3.2 Overall Discussion

By visual evaluation, it can be said among all six confusion matrices, the most convincing trained model is the eye-patch model as the model is able to capture all the six emotions without fail. Nevertheless, detail calculation is needed on each emotion categories to accept the hypothesis.

To further analyze this confusion matrix, a couple of simple calculations was made to look at the difference in the model's prediction.

True values for each emotion from the matrices were obtained to access the accuracy for the patch model vs the full-face model using Eq. 1.

$$\begin{aligned} \text{Angry eye} - \text{patch} &= \frac{\text{True value of angry}/}{\text{Total images detected as angry by eyepatch model}} \\ &= 274/663 \\ &= 0.4133 \end{aligned} \tag{1}$$

The above equation is applied for all the true values in the confusion matrix to yield Table 2 and Table 3.

Even though the transfer learning performance of the full-face model is high, when the patch datasets have been presented to the full-face model the performance on recognition certain emotions are skewed to sad and surprise. In the previous research, M. Grahlow states on poor detection rate on anger, fear and sad when occluded faces are presented to full face models [5]. Similar result can also be seen in this experiment on Table 2 as the model has poor recognition on fear.

The full-face model's detection rate on other patch datasets is convincing since the transfer learning of full-face model is 74.68%. It can be clearly seen, the full-face model, seem to be better in detecting the sad and surprise and highly skewed to those emotions. When we further analyze the patched model on patch datasets, better accuracy is obtained for a number of emotions as derived in Table 3 based on the confusion matrix.

Table 2. Accuracy assessment for full face model to patch datasets

		Emotions accuracy (%)					
	Datasets	Angry	Disgust	Fear	Happy	Sad	Surprise
Full face model	Eye-patch	26.86	32.91	17.24	26.06	**22.96**	**70.95**
	Forehead-patch	30.51	17.99	9.28	29.32	**19.61**	**71.43**
	Skin-patch	25.74	33.29	26.37	30.53	**19.39**	0.00

Table 3. Accuracy assessment for patch models to patch and full-face datasets

		Emotions accuracy (%)					
	Datasets	Angry	Disgust	Fear	Happy	Sad	Surprise
Forehead-patch model	Forehead-patch	21.61	40.63	4.11	58.52	3.13	6.97
	Full-face	18.60	21.78	18.18	21.70	19.78	24.54
Eye-patch model	Eye-patch	**41.33**	**39.66**	**38.72**	28.73	4.98	**31.28**
	Full-face	**21.27**	**48.72**	**13.23**	17.62	20.11	**28.21**
Skin-patch model	Skin-patch	7.38	**40.88**	**35.23**	**51.38**	4.89	0.00
	Full-face	14.29	**33.33**	**33.33**	0.00	0.00	17.21

Table 4. Differences between VERT-K model and Inception v3

Emotion accuracy on eye-patch datasets	VERT-K model M. Grahlow [5]	Inception-v3
Angry (%)	56	41.33
Disgust (%)	19	39.66
Surprise (%)	(not evaluated)	31.28

From the accuracy Table 3, we can evaluate further into specific emotions for each dataset and model. As mentioned by M. Grahlow the cropping method used on full face models gives different detection rate for different emotions [5]. Based on Table 4, we can see that, the Inception-v3 eye patch models on eye patch datasets are able to recognize more emotions accurately when compared to other research. Then looking at Table 3, angry emotion has the highest recognition rate using eye-patch model on eye patch datasets. This is because eyes play crucial role in determining anger as the facial expression of anger mostly emphasizes the central and downward movement of eyebrows and glaring eyes.

Moreover, surprise and fear emotions also can be expressed by movement of eyebrows and mouth muscles. Distinct elevated eyebrows show surprise and it can be seen that eye patch model have high detection and accurate on eye-patch dataset among other patch models. As for fear, eyebrow, forehead, and more lips movements are involved.

When Table 2 is observed, the eye patch and skin patch have better accuracy for fear and skin patch gives better result in both datasets as fear involves more lips muscles [15].

On the other hand, happiness is determined by smiles and wrinkles at edge of eye. That is why the skin patch seems to be more accurate. While, disgust is perceived by nose wrinkles, eyebrows pulled down and squint eyes [15]. Due to the involvement of middle part of the face, eye patch and skin patch model have higher detection while skin patch model on skin patch dataset has higher accuracy.

From comparing Table 2 and 3, the full-face model has quite convincing detection rate on segmented faces but to gain higher accuracy retraining the model with segmented model is highly encouraged.

When comparing overall models, the accuracy table clearly shows that the patch model has higher accuracy on its own patch and the full-face datasets among the three trained models. When the accuracy values are compared between each dataset for the respective models, the eye-patch model has the greatest number of emotions recognized accurately among the three patched models. From this table, we can set the trained eye-patch model's prediction as the benchmark for this experiment. Hence, it can be evaluated that, the training the segmented regions gives better accuracy and emotion recognition for occluded datasets [6].

4 Conclusion

The current global pandemic has created a new norm of wearing facial mask in public to reduce the spread of Covid19. With this mask mandate, emotion recognition has been severely challenged as most of the face is occluded. Nevertheless, deep learning method for recognition is a big help to overcome that hurdle and different methods of training the model to maintain the high accuracy rate in emotion recognition is being deeply studied. This paper has proven that training the model with the segmented patch model is the best to recognize and predict the emotion accurately. Hence, this illustrates the significance of informing/training the learning model the presence of occlusion patterns for better recognition. Even though we ought to prove this method is resulting the best outcome, there are still several issues that need to be considered for future research on the additional occlusion of wearing sunglasses or a cap since the KDEF database is only focused on bare faces.

Acknowledgements. This research is supported by a TNB SEED grant managed by UNITEN R&D U-TD-TD-19–28.

References

1. Lundqvist, D.E., Flykt, A., Öhman, A.: The karolinska directed emotional faces - KDEF, CD ROM from Department of Clinical Neuroscience. Psychology section, Karolinska Institutet (1998). ISBN 91–630–7164–9
2. Fasel, B., Luettin, J.: Automatic facial expression analysis: a survey. PR **36**(1), 259–275 (2003). https://doi.org/10.1016/S0031-3203(02)00052-3

3. Ekman, P., Friesen, W.V.: Constants across cultures in the face and emotion. J. Pers. Soc. Psychol. **17**(2), 124 (1971)
4. Mehrabian, A.: Silent Messages - A Wealth of Information About Nonverbal Communication (Body Language). Silent Messages, Belmont (1981)
5. Grahlow, M., Rupp, C., Dernt, B.: The impact of face masks on emotion recognition performance and perception of threat (2021). https://doi.org/10.31234/osf.io/6msz8, Accessed 27 July 2021
6. Ranzato,M., Susskind, J., Mnih, V., Hinton, G.: On deep generative models with applications to recognition. In: CVPR 2011, pp. 2857–2864 (2011). https://doi.org/10.1109/CVPR.2011.5995710
7. Teoh, K.H., et al.: Face recognition and identification using deep learning approach. J. Phys. Conf. Ser. **1755**(1), 012006 (2021). https://doi.org/10.1088/1742-6596/1755/1/012006
8. Li, Y., Guo, K., Lu, Y., Liu, L.: Cropping and attention based approach for masked face recognition. Appl. Intell. **51**(5), 3012–3025 (2021). https://doi.org/10.1007/s10489-020-02100-9
9. Hariri, W.: Efficient masked face recognition method during the COVID-19 pandemic (2020). PREPRINT (Version 1) available at Research Square, https://doi.org/10.21203/rs.3.rs-39289/v1
10. Pinkney, J.: MTCNN face detection. https://github.com/matlab-deep-learning/mtcnn-face-detection/releases/tag/v1.2.4, GitHub. Accessed 26 July 2021
11. Pretrained Deep Neural Network. https://www.mathworks.com/help/deeplearning/ug/pretrained-convolutional-neural-networks.html#References, Accessed 25 July 2021
12. Inceptionv3. https://www.mathworks.com/help/deeplearning/ref/inceptionv3.html, Accessed 25 July 2021
13. Sajjanhar, A., Wu, Z., Wen, Q.: Deep learning models for facial expression recognition. In: 2018 Digital Image Computing: Techniques and Applications (DICTA), pp. 1–6 (2018). https://doi.org/10.1109/DICTA.2018.8615843
14. Salvador, R.C., Bandala, A.A., Javel, I.M., Bedruz, R.A.R., Dadios, E.P., Vicerra, R.R.P.: DeepTronic: an electronic device classification model using deep convolutional neural networks. In: 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM), pp. 1–5 (2018). https://doi.org/10.1109/HNICEM.2018.8666303
15. Zhang, L., Tjondronegoro, D.: Facial expression recognition using facial movement features. IEEE Trans. Affect. Comput. **2**(4), 219–229 (2011). https://doi.org/10.1109/T-AFFC.2011.13