Vol.2021-AVM-113 No.10
2021/6/24

Facial Mask aware Facial Expression Recognition Approaches and Application

Yang Bo[1,a)] Wu Jianming[1,b)] Hattori Gen[1,c)] Takishima Yasuhiro[1,d)]

Abstract: Facial mask aware facial expression recognition (FER) problem, which can be considered as a sub-task of FER under partial occlusion, has become a challenging task in the computer vision area, especially when wearing face masks is considered an effective means of preventing the transmission of coronavirus during the COVID-19 pandemic. To encourage communication of human beings, making it available that reading each other's emotions even when large facial areas are covered by face masks an urgent issue is needed to be solved. To stress this problem, we firstly propose a method that can add face masks to existing FER datasets automatically using differently shaped masks according to facial orientations. The FER models based on VGG19 and MobileNet are trained on public and private FER datasets added with mask, and the experimental results show that training a FER model based on a simulated masked FER dataset is feasible. Then we propose a two-stage attention model 1 to improve the accuracy of face-mask-aware FER: a pre-trained occlusion-aware facial landmark detection (FLD) model to roughly distinguish the masked facial parts from the unobstructed region in stage 1, a FER classifier to do facial expression classification considering both occluded and non-occluded regions in stage 2. To further improve the accuracy of face-mask-aware FER, we further propose an advanced two-stage attention model 2: a masked/unmasked binary deep binary deep classifier in stage 1 and a FER classifier in stage 2. Both models outperform other state-of-the-art occlusion-aware FER methods on face-mask-aware FER datasets, whether in the wild or in the laboratory. Finally, we develop a real-time FER application concerning facial masks based on our proposals, and a demonstration experiment is also being conducted, which is expected to attract attention from inside and outside the company.

Keywords: Facial expression recognition, occlusion, face mask, deep attentional network, generative adversarial network, facial landmark detection, COVID-19

1. Introduction

Facial expression recognition (FER) plays an important role in not only social communication in daily life but also artificial intelligence (AI) applications, such as human-computer interaction, driver safety, health care, and entertainment. In recent years, with the major boom of deep learning implementation, the FER technology has attracted increasing attention and achieved reasonable accuracy [1]. As the coronavirus disease 2019 (COVID-19) is spreading worldwide, governments and organizations like WHO advocated the wearing of face masks as a key strategy in reducing the spread of the coronavirus. However, the negative effect aspects of waring masks have been studied by the psychologists who report that it strongly confused counterparts in reading emotions, thus crucially affecting social interaction [2]. In the real world, partial occlusion in the face by random objects (hands, hairs, cups, etc.) and facial accessories (sunglasses, scarves, masks, etc.) is one of the major challenges for accurate FER [3]. Unlike other partial occlusion FER problems, wearing a mask covers half of a person's face, especially the mouth, which is highly informative in helping to distinguish between the emotions of sadness and disgust, or fear and surprise [4]. To address the issue of face masks in FER, we present a two-stage attention deep network for robust FER in this paper: the occlusion-aware facial landmark detection (FLD) stage and the masked facial expression recognition (MFER) stage. In the occlusion-aware facial landmark detection stage, the generative adversarial network (GAN) mechanism is conducted to reconstruct the masked facial parts, which improves the accuracy of FLD even when one person is wearing a face mask. Using the pre-trained GAN model, 68 landmarks of a facial image with a face mask could be obtained, using landmarks around the nose we could roughly distinguish the masked facial parts from the unobstructed region. In the masked facial expression recognition stage, the attention mechanism is conducted to guide the model to focus on the facial parts most important to the FER classification results and, meanwhile, pay more attention to the unmasked region but less to the masked region.

1 KDDI Research, Inc., Japan
a) bo-yang@kddi-research.jp
b) ji-wu@kddi-research.jp
c) ge-hattori@kddi-research.jp
d) takisima@kddi-research.jp

Further, we present an advanced two-stage attention deep network for robust FER: the masked/unmasked binary classification stage and the masked facial expression recognition stage. In the masked/unmasked binary classification stage, the attention mechanism is conducted in the classifier to generate attention heatmaps of the masked region and reverse attention heatmaps of the unmasked region. In the masked FER classifier stage, the attention mechanism is also conducted to guide the model to focus on the facial parts most important to the FER classification results and, meanwhile, pay more attention to the unmasked region but less to the masked region.

The main contributions of our works can be summarized as follows:

1. To the best of our knowledge, our works are the first to release FER datasets concerning face masks and propose a new approach to deal with the face-mask-aware FER, specifically.

2. In stage 1 of both face mask aware FER proposals, the pre-trained model is used to roughly distinguish the masked facial parts from the unobstructed region, thus focus on the unmasked regions rather than the masked regions.

3. In stage 2 of both face mask aware FER proposals, the proposed FER classifier utilizes reweighed combination loss to pay more attention to the unmasked region, but still takes the region around the face mask into consideration to promise robustness of the model.

4. Both proposed proposals show a better performance on face-mask-aware FER datasets, compared to the other approaches dealing with occlusion-aware FER.

5. To the best of our knowledge, our company conducts the first press release of face masks aware FER technologies. We also apply it in the product prototype and a demonstration experiment is also conducted.

2. Related Works

2.1 Partial Occlusion FER

Although the majority of existing FER studies are focusing on non-occluded faces, some approaches have also been proposed to solve the partial occlusion problem in two-dimensional FER, which can be roughly classified into three categories: reconstruction-based, holistic-based, and sub-region-based approaches. The reconstruction-based approach attempts to recover the features in the occluded regions, such as the approaches used in [5]. However, when more than half of a facial region is occluded by objects, such as a face mask, it becomes even harder to recover the whole face based on the non-occluded region. The holistic-based approach treats the face as a whole and uses a sparse representation of images and designated regularization, which is only robust for small occlusion [6] [7]. The sub-region-based approach treats the face as a combination of overlapped or non-overlapped patches that are small in size. The occluded patches are ignored, or the occluded and non-occluded patches are assigned with different weights.

The attention mechanism is also conducted to guide the model to pay more attention to the region that is essential for classification accuracy. Li et al. [8] proposed an ACNN framework, which consisted of patch-based ACNN(pACNN) and global-local-based ACNN (gACNN) with an attention net, to balance local representations at the patch-level and the global representations at the image-level. Wang et al. [9] proposed an RAN framework, which consisted of self-attention and relation-attention modules, to adaptively capture the importance of the facial regions for occlusion-and-pose variant FER. Ding et al. [10] argued that both of the aforementioned frameworks might not accurately locate large non-occluded facial regions because the self-attention-based methods lacked additional supervision. Ding proposed an OADN framework, which consisted of a landmark-guided attention branch and a facial region branch to improve the FER accuracy, even when a large non-occluded region existed. However, none of these approached were designed especially for FER concerning face masks, which have huge redundant regions that should be discarded or paid less attention to from the very beginning.

2.2 Partial Occlusion FER Dataset

Recently, many FER datasets are released for public study, some famous datasets among them are RAF-DB, AffectNet, and FERPlus. RAF-DB [11] contains 30,000 in-the-wild facial expression images, annotated with basic or compound expressions by forty independent human labelers. AffectNet [12] is currently the largest expression dataset, in which about 400,000 images manually annotated with seven discrete facial expressions and the intensity of valence and arousal. FERPlus [13] is a real-world facial expression dataset consists of 28,709 trainning images, 3,589 validation images and 3,589 test images, with each image labeled with one of the eight expressions by ten independent taggers.

There are several FER datasets concerning partial occlusion released for public study, such as Occlusion-AffectNet, Occlusion-FERPlus, and FED-RO. Occlusion-AffectNet [9] and Occlusion-FERPlus [9] are two datasets, where images under occlusion are collected from the aforementioned AffectNet and FERPlus, accordingly. FED-RO [8] is a facial expression dataset with real world occlusions, in which each image has natural occlusions including sunglasses, medical mask, hands and hair. And FER-RO dataset also contains 400 images labeled with seven expressions for testing. However, none of the existing FER datasets and occlusion-aware FER datasets are specified for face-mask-aware FER, even some of occlusion-aware FER datasets containing samples with face mask.

3. Proposed AWFM and Masked FER Dataset

To solve the problem of the lack of an FER dataset with face masks, we proposed an automatic wearing face mask

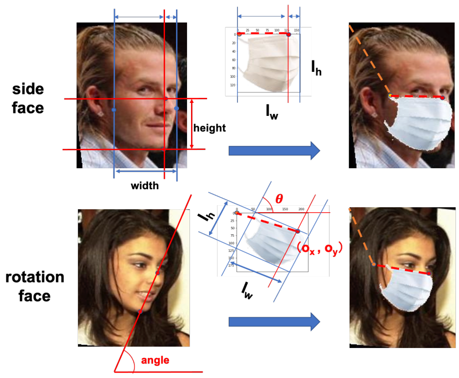

Fig. 1 Automatic wearing face mask (AWFM) approach. [14]

(AWFM) approach, which could automatically add face masks to existing FER datasets using differently shaped masks according to facial orientations. This work has already been published in ACM MUM 2020 [14]. As is shown in Figure 1, the mask is resized according to the ration between left side and right side when the side face is not rotated. Only white-colored front view/side view masks with white color are used in this paper and sample images are selected from the Labeled Faces in the Wild (LFW) dataset [15]. When the face is in a rotated position, the mask is first resized in the case of the side face. Then, the angle of the rotated face is calculated according to the line between nose and jaw. Finally, the mask is rotated according to the angle and the offset (o_x, o_y) between left-top point of the mask and nose is calculated. From result samples, we could easily tell that the performance of AWFM for the side face with the mask on using AWFM is apparently better than the results using most of current apps.

For the purpose of fair evaluation, we evaluated the performance of two famous deep learning models (MobileNet[16] and VGG19[17]) for facial emotion classification with three categories (positive, neutral, negative) on the M-LFW-FER and M-KDDI-FER datasets.*1 To prepare M-LFW-FER and M-KDDI-FER datasets, firstly the LFW dataset was manually annotated according to three types of facial expressions (positive, negative, neutral), which all contain five types of facial orientations (up, left, center, right, down). Some images, difficult to distinguish facial expressions, were removed and 10487 out of 13000 samples were selected from the LFW dataset to finally obtain the LFW-FER dataset. AWFM was then used to process all samples in the LFW-FER dataset by putting a mask on the faces to obtain a final M-LFW-FER dataset. We also constructed a private KDDI-FER dataset, in which the facial expressions (positive, neutral, negative) of 12 Asian subjects (5 females, 7 males) were photographed with five facial orientations (up, left, center, right, down) for each expression. Then, a total of 17236 samples were collected with 3447 samples for each orientation category and 1149 samples for each expression. AWFM was also used to process all the samples in the KDDI-FER dataset and finally a masked KDDI-FER dataset (M-KDDI-FER) was obtained. In order to test the FER models trained on the above-mentioned datasets, we also constructed a real-world masked FER test dataset for model evaluation. We manually crawled 562 masked facial expression figures (213 natural, 162 positive, 187 negative) from the Internet by searching keywords, such as "smile, face, with mask" or "angry, face, with mask". The obtained real-world masked FER dataset for the test was called the as M-FER-T dataset.

*1 https://github.com/KDDI-AI-Center/LFW-emotion-dataset

4. Proposed Approach 1

We proposed a two-stage model to solve the face mask aware FER problem, which was published in the 112th AVM meeting of IPSJ [18] and won the best paper award.

4.1 Stage 1: Face mask aware Face Inpainting

In stage 1, as is illustrated in Figure 2, unmasked images and masked images, which are worn with a mask using the aforementioned AWFM approach, are used to train a generative adversarial network (GAN) model. The pre-trained GAN model can be further utilized to generate attention weight maps for face mask regions and reverse attention weight maps for unmasked regions. Unlike the aforementioned ACNN, RAN, and OADN frameworks that have to deal with various kinds of occlusion objects, our proposed approach just pays less attention to the face mask, which is regular in shape and less informative. One simple idea is to treat face-mask-aware FER as an object detection task and discard the masked region to eliminate its effect on FER classification. Nevertheless, two problems may arise: 1. The object detection task is not a simple job and needs a lot of human power, such as annotation. 2. Face masks vary in shape and orientation, which makes them hard to be detected precisely. Thus, some regions around the face mask, informative for FER classification, are also discarded. Inspired by the lightweight attentional convolutional network proposed in [19] and the visualization technique proposed in [20], we to narrow down the face mask region by training the GAN model, consisting of the global discriminator and local discriminator.

Existing facial landmark detection methods, such as feature extraction-based SVM classification methods or CNN-based methods, have already shown good performance in detecting facial important landmarks. However, the existing methods suffer from poor performance when the reconstruction error spreads over the whole face under occlusions and each of these approaches hardly reaches state-of-the-art performance on "in the wild" datasets. Liu et al. introduced an occlusion-aware facial landmark detection using the generative adversarial network with improved auto-encoders (GAN-IAs) and deep regression networks [21]. Inspired by GAN-IAs, we also propose a GAN model to reconstruct the facial parts covered by the face mask by a generator, in which the backbone of VGG19 (convolu-

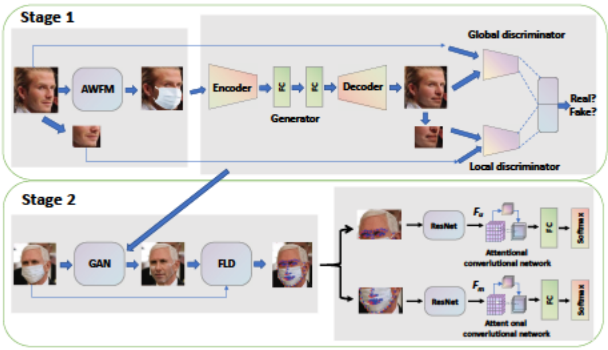

Fig. 2 Two-stages framework for face mask aware robust FER of approach 1. [18]

tional layers and pooling layers) is used. We firstly introduce the reconstruction loss L_r, which is the distance between the model output and the original image. If only the single loss L_r is used, the over-punishment problem might be aroused and the generated image tends to be blurred. To ensure the generator can recover the masked image realistically, two discriminators are introduced: global discriminator and local discriminator, along with the adversarial loss. Thus, it turns the standard optimization of a neural network into a min-max optimization problem. In each training iteration, the discriminators are updated with the generator together.

4.2 Stage 2: Attention mechanism based FER classifier

In stage 2, as is illustrated in Figure 2, we first put the masked facial image into a pre-trained GAN model from stage 1 to generate an unmasked image. Then using landmarks are detected by a deep regression network, which can be any facial landmark detection (FLD) deep model, for example, CNN-based FLD. In this way, we can finally obtain 68 facial landmarks of high accuracy on the original image with the face mask. In this study, we observed many samples and decided to use No.29 of the facial landmarks to divide the image into two parts: one part is with the facial regions almost without the mask and one part is mainly covered by the face mask. To extract facial features, the masked part and the unmasked part are fed into the convolutional feature extractor, for example, the backbone of ResNet or VGG19 without the fully connected layer and the average pooling layer. Employing this, the feature of the region around the masked region and the unmasked region can be manipulated separately.

After the rough separation, the F_m and F_u are respectively fed into the lightweight attentional convolutional networks, which is also mentioned in stage 1. For F_u, the attention mechanism can guide the model to learn the most important region in detecting a specific emotion, like the region around the eyes. For F_m, the approach in stage one can not guarantee that the masked region and the unmasked region are separated precisely. Suppose that the facial region around the face mask may also affect the FER accuracy, we still consider this part but will pay less attention to it. Furthermore, for both F_u and F_m, two fully connected layers are used to reduce the dimension of the feature maps, and finally, the output is fed into a softmax layer to predict the image with a certain expression category. Cross-entropy loss is utilized here to train the unmasked and masked regions for FER classification, where the losses are denoted as L_u and L_m, separately. Finally, the total loss of stage two is treated as the combination of L_u and L_m.

5. Proposed Approach 2

To improve the performance of face mask aware FER, we further proposed an advanced two-stage model to solve the face mask aware FER problem, which has been submitted to IEEE ICIP 2021 [22].

5.1 Stage 1: Binary deep classifier for heatmaps generation

In stage 1, as is illustrated in Figure 3, masked and unmasked images are used to train an attentional convolutional network, which can be further utilized to generate attention weight maps for face mask regions and reverse attention weight maps for unmasked regions. Also inspired by the lightweight attentional convolutional network proposed in [19] and the visualization technique proposed in [20], we propose a novel idea to narrow down the face mask region by training a masked/unmasked binary classifier, which is also utilized for generating the attention heatmaps. The spatial transformer network is used as an attention mechanism to focus on the important facial regions. It warps the input from block $n-1$ to the output of block n, in which block n and block $n-1$ refer to adjacent blocks consisting of convolutional layers, max-pooling layers and the rectified linear unit (ReLU) activation function. In order to generate attention heatmaps focusing on the face mask region, each time we zero out a square patch of size $N \times N$ starting from the top-left corner of an image, and make a prediction using the trained model on the varied image. The patches that make the FER classifier predict masked as unmasked is considered as regions of importance in classifying masked and unmasked faces. In this way, we can get an attention heatmap S_m around the face mask region, which is also the most important region overall in distinguishing the masked face and unmasked faces. Then we up-sample the S_m and feed the result to the sigmoid function. The purpose of doing this is to convert each pixel in attention heatmaps to the range of $[0, 1]$. And we get an attention weight map for regions around the face mask.

5.2 Stage 2: Attention Deep Classifier for masked FER

In stage 2, as is illustrated in Figure 3, we firstly use the AWFM method to generate as much of the face-mask-aware FER data as possible based on existing FER datasets. In order to extract facial feature, images with the face mask are fed into the convolutional feature extractor, for example, the backbone of ResNet or VGG19

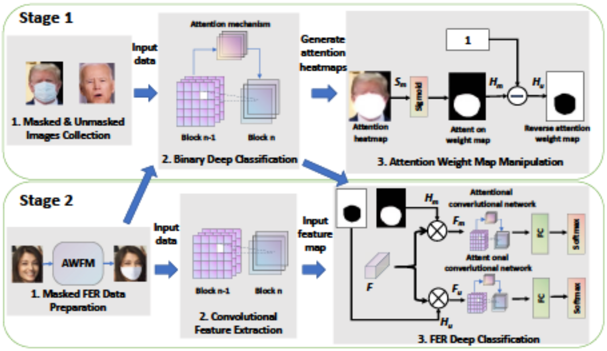

Fig. 3 Advanced two-stages framework for face mask aware robust FER of approach 2. [22]

without the fully-connected layer and the average pooling layer. Given the global feature maps F, along with the attention weight H_m and the reverse attention weight H_u, the reweighed features for the masked region F_m and the reweighed features for the unmasked region F_u are calculated. Utilizing this, the features of the region around the masked region and the unmasked region can be manipulated separately. After the rough separation, the F_m and F_u are respectively fed into the lightweight attentional convolutional networks as mentioned in the binary deep classifier of stage 1. For F_u, the attention mechanism can guide the model to learn the most important region in detecting a specific emotion, like the region around the eyes. For F_m, the approach in stage one can not guarantee that the masked region and the unmasked region are separated precisely. Suppose that the facial region around the face mask may also affect the FER accuracy, we still take this part into consideration but will pay less attention to it. Furthermore, for both F_u and F_m, two fully connected layers are used to reduce the dimension of the feature maps, and finally the output is fed into a softmax layer to predict the image with a certain expression category. Cross-entropy loss is utilized here to train the unmasked and masked regions for FER classification, where the losses are denoted as l_u and l_m, separately.

6. Face Mask aware FER System

By conducting performance evaluation, we found that approach 1 obtains the average detection accuracy of 87% [18] and approach 2 reaches around 90% [22], which outperforms other state-of-art occlusion-aware FER methods. Based on our previous work [23], we develop a real-time video-based FER system to detect the facial expression of a person and display the results by virtualized bars utilizing the probability value of positive, negative, and neutral, respectively. The real-time video-based FER system consists of three stages: In the first stage, the face of the user is detected using open-source tools such as OpenCV. After the region of the face is located, it is then cropped and resized to 96x96px and the manipulated image data is transferred to the server from the client side. In the second stage, the pre-trained FER classifier model processes the static face image and predicts it as probabilities of three

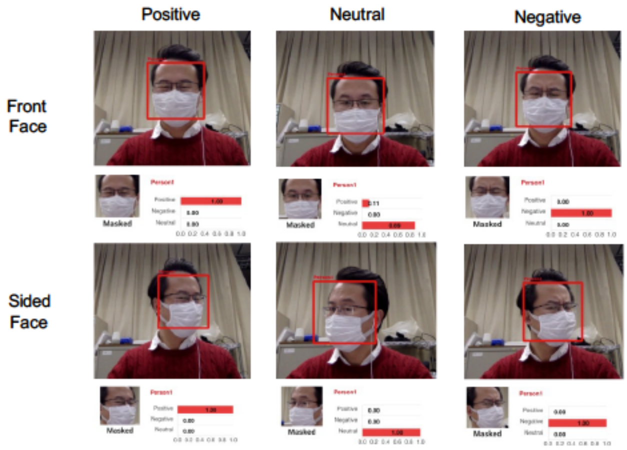

Fig. 4 Face mask aware FER technologies and application of KDDI Research, Inc.

emotions as is aforementioned, which add up to be integer one principally. In the final stage, the client-side gets the results from the server-side and uses three bars to display the probabilities of positive, neutral, and negative, respectively. The direction of the user's face is also taken into consideration, thus the face mask aware FER system is robust when the user turns his face to the left side or right side, as is shown in Figure 4. The system can also statistics the user's expression in an hour, which can reflect the user's inner state in a certain period.

Recently, KDDI Inc. and East Japan Railway Company have already conducted two press releases concerning "Space Free Workplace"[24] [25], considering to applicate the face mask aware facial expression recognition technologies and the demonstration experiment is undergoing at present. Our proposed FER technologies and application is also expected to apply to various fields, take one for example, training practitioners in the service industry to express their emotions clearly to the clients even when they have to wear face masks during the COVID-19 pandemic.

7. Conclusion

In this paper, we propose a two-stage attention model and an advanced two-stage attention model to improve the accuracy of face masks aware FER. In stage 1, we roughly distinguish the masked facial parts from the unobstructed region. In stage 2, we train a FER classifier, which is guided to pay more attention to the region that essential to the facial expression classification, and both occluded and un-occluded regions are taken into consideration but re-weighted. The proposed methods outperformed other state-of-the-art occlusion-aware facial expression recognition methods on masked facial expression datasets both in the wild and in the laboratory. We also applied the technologies in the product prototype and a demonstration experiment is also conducted. To further improve the performance and robustness, we are collecting more training data to make it more effective for various races, gender, expressions, colors of masks, etc.

References

[1] Li S. and Deng W.: Deep Facial Expression Recognition: A Survey, IEEE Transactions on Affective Computing, pp. 1–1, DOI: 10.1109/TAFFC.2020.2981446 (2020).

[2] Carbon C.C.: Wearing Face Masks Strongly Confuses Counterparts in Reading Emotions, Frontiers in Psychology, vol. 11, pp. 2526,
DOI: 10.1109/IJCB48548.2020.9304923 (2020).

[3] Zhang L. and Verma B. and Tjondronegoro D. and Chandran V.: Facial Expression Analysis under Partial Occlusion: A Survey, J. ACM Comput. Surv., vol. 51, pp. 1–49,
DOI: 10.1145/3158369 (2018).

[4] Schurgin M.W. and Nelson J. and Iida S. and Ohira H. and Chiao J.Y.: Eye movements during emotion recognition in faces, Journal of Vision, vol. 14, pp. 1–14,
DOI: 10.1167/14.13.14 (2014).

[5] Lu Y. and Wang S. and Zhao W. and Zhao Y.: GAN-Based Robust Occluded Facial Expression Recognition, J. IEEE Access, vol. 7, pp. 93594–93610,
DOI: 10.1109/ACCESS.2019.2928125 (2019).

[6] Wright J. and Yang A.Y. and Ganesh A. and Sastry S.S. and Ma Y.: Robust Face Recognition via Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, iss. 2, pp. 210–227,
DOI: 10.1109/TPAMI.2008.79 (2009).

[7] Osherov E. and Lindenbaum M.: Increasing CNN Robustness to Occlusions by Reducing Filter Support, 2017 IEEE International Conference on Computer Vision(ICCV) IEEE, pp. 550–561,
DOI: 10.1109/ICCV.2017.6 (2017).

[8] Li Y. and Zeng J. and Shan S. and Chen X.: Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism, IEEE Transactions on Image Processing, vol. 28, pp. 2439-2450,
DOI: 10.1109/TIP.2018.2886767 (2019).

[9] Wang K. and Peng X. and Yang J. and Meng D. and Qiao Y.: Region Attention Networks for Pose and Occlusion Robust Facial Expression Recognition, IEEE Transactions on Image Processing, vol. 29, pp. 4057–4069,
DOI: 10.1109/TIP.2019.2956143 (2020).

[10] Ding, H. and Zhou, P. and Chellappa, R.: Occlusion-Adaptive Deep Network for Robust Facial Expression Recognition, 2020 IEEE International Joint Conference on Biometrics(IJCB) IEEE, pp. 1–9,
DOI: 10.1109/IJCB48548.2020.9304923 (2020).

[11] Li S. and Deng W.: Reliable Crowdsourcing and Deep Locality-Preserving Learning for Unconstrained Facial Expression Recognition, IEEE Transactions on Image Processing, vol. 28, pp. 356-370,
DOI: 10.1109/TIP.2018.2868382 (2019).

[12] Mollahosseini A. and Hasani B. and Mahoor M. H.: AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild, IEEE Transactions on Affective Computing, vol. 10, pp. 18-31,
DOI: 10.1109/TAFFC.2017.2740923 (2019).

[13] Barsoum E. and Zhang C. and Ferrer C.C. and Zhang Z.: Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution, Proceedings of the 18th ACM International Conference on Multimodal Interaction(ICMI) ACM, pp. 279–283, Tokyo, Japan,
DOI: 10.1145/2993148.2993165 (2016).

[14] Yang, B. and Wu, J. and Hattori, G.: Facial Expression Recognition with the Advent of Face Masks, 19th International Conference on Mobile and Ubiquitous Multimedia(MUM) ACM, pp. 335–337, Essen, Germany,
DOI: 10.1145/3428361.3432075 (2020).

[15] Huang G.B. and Mattar M. and Berg T. and Learned-Miller E.: Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments, Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, pp. 1–14, Marseille, France (2008).

[16] Howard A. and Zhu M. and Chen B. and Kalenichenko D. and Wang W. and Weyand T. and Andreetto M. and Adam H.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, ArXiv, vol. abs/1704.0486 (2017).

[17] Simonyan K. and Zisserman A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, International Conference on Learning Representations(ICLR) (2015).

[18] Yang, B. and Wu, J. and Hattori, G.: Occlusion aware Facial Landmark Detection based Facial Expression Recognition with Face Mask, 112th AVM meeting of Information Processing Society of Japan (IPSJ), vol. 2021-AVM-112, num. 4, pp. 1–6, online, Japan (2021).

[19] Minaee S. and Abdolrashidi A. : Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network, ArXiv, vol. abs/1902.01019 (2019).

[20] Zeiler M.D. and Fergus R.: Visualizing and understanding convolutional networks, 13th European conference on computer vision(ECCV) Springer Verlag, pp. 818–833 (2014).

[21] Liu H. and Zheng W. and Xu C. and Liu T. and Zuo M.: Facial Landmark Detection Using Generative Adversarial Network Combined with Autoencoder for Occlusion, Mathematical Problems in Engineering, vol. 2020, num. 4589260, pp. 8,
DOI: 10.1155/2020/4589260 (2015).

[22] Yang, B. and Wu, J. and Hattori, G.: Face mask aware robust facial expression recognition during the COVID-19 Pandemic, 2021 IEEE International Conference on Image Processing (IEEE ICIP 2021), Anchorage, Alaska, USA (2021).

[23] Wu, J. and Hagiya, T. and Tang, Y. and Hoashi, K.: Effects of Objective Feedback of Facial Expression Recognition during Video Support Chat, Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia (ACM MUM 2017), doi = 10.1145/3152832.3152848 (2017).

[24] Press release 1 of KDDI Inc. about Space Free Workplace, available from ⟨https://news.kddi.com/kddi/corporate/newsrelease/2021/05/11/5095.html⟩ (2021).

[25] Press release 2 of KDDI Inc. about Space Free Consortium, available from ⟨https://news.kddi.com/kddi/corporate/newsrelease/2021/05/11/5094.html⟩ (2021).