

Facial expression recognition of masked faces using deep learning

Boutaina Hdioud¹, Mohammed El Haj Tirari²

¹Department of Computer Science, ENSIAS, Mohammed V University, Rabat, Morocco

²Department of Statistics, INSEA, Rabat, Morocco

Article Info

Article history:

Received Feb 4, 2022

Revised Oct 6, 2022

Accepted Nov 5, 2022

Keywords:

Deep neural networks
Facial expression recognition
Knowledge distillation
Masked face
Transfer learning

ABSTRACT

Facial expression recognition (FER) represents one of the most prevalent forms of interpersonal communication, which contains rich emotional information. But it became even more challenging during the times of COVID, where face masks became a mandatory protection measure, leading to the challenge of occluded lower-face during facial expression recognition. In this study, deep convolutional neural network (DCNN) represents the core of both our full-face FER system and our masked face FER model. The focus was on incorporating knowledge distillation in transfer learning between a teacher model, which is the full-face FER DCNN, and the student model, which is the masked face FER DCNN via the combination of both the loss from the teacher soft-labels vs the student soft labels and the loss from the dataset hard-labels vs the student hard-labels. The teacher-student architecture used FER2013 and a masked customized version of FER2013 as datasets to generate an accuracy of 69% and 61% respectively. Therefore, the study proves that the process of knowledge distillation may be used as a way for transfer learning and enhancing accuracy as a regular DCNN model (student only) would result in 46% accuracy compared to our approach (61% accuracy).

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Boutaina Hdioud

Department of computer Science, ENSIAS, Mohammed V University

Street Mohammed Ben Abdellah Regragui, Madinat Al Irfane, BP 713, Agdal, Rabat, Morocco

Email: boutaina.hdioud@um5.ac.ma

1. INTRODUCTION

In the field of computer vision, emotion recognition is a prominent area. Machine learning and deep learning techniques are now widely used, the possibility of developing intelligent systems capable of properly recognizing emotions became a reality. Facial expression recognition (FER) technology is important in artificial intelligence (AI) since it assists in understanding the internal states of people [1]–[6]. With the worldwide spread of coronavirus disease 2019 (COVID-19), wearing face masks while interaction in public is becoming a common behavior to protect against infection. Thus, how to improve effectiveness of existing FER technology on masked faces has become an urgent issue.

Previous research has demonstrated that the mouth may provide a wealth of information about our emotions, such as the difference between fear and surprise or between sadness and disgust [7]. Therefore, current FER models will be less effective when a person's face is partially covered by a mask to more than half, including the nose and mouth. As an example, happy facial expressions have a key function in reducing anxiety in a physician-patient relationship. As a result, wearing face masks has a negative impact on the physician-patient interaction. The physician's ability to assess the patient's sentiments and emotions will be compromised if the patient's face is covered. The patient may also miss the physician's expressions of empathy.

This paper aims to study facial expression recognition for masked faces in order to further assist humans in interpersonal communication during COVID times. The goal is to construct a deep learning model based on a teacher-student architecture that uses knowledge distillation in transfer learning to distinguish facial expressions when there is occlusion due to masks. The model would be able to classify masked human faces based on the main seven facial expressions : sadness, happy, disgust, fear, surprise, anger, and neutral expression.

The outline of this paper is structured : section 2 examine existing approaches used in FER in general and under partial occlusion. Section 3 explains the methodology used in the study. Section 4 details our experiments and the results. Finally, section 5 draws some conclusions.

2. RELATED WORK

In recent years, researchers proposed several models to solve the issue of FER. The key difference between each model is dependent on how the architecture is modeled and how spatial-temporal approaches to processing image sequences are used. In this section, we present literature that is closely related to our study.

2.1. Facial expression recognition

2.1.1. Machine learning based approaches

Machine learning methods are combinations of feature extraction techniques and classifiers [8]. It is possible to extract features using geometric or appearance-based techniques. Appearance based techniques characterize the texture of the face by looking at the entire face as a whole, or at individual parts like the eyes, nose, and mouth. (E.g., [9]–[12]). Geometry based techniques may determine the contour of a face's appearance, along with its landmarks (such the eyes and nose) by tracking facial points. (E.g., [13], [14]). To categorize mood from extracted feature values, a variety of approaches employed support vector machines (SVM), K-neighbors neighbors (KNN), and random forest [15]–[17].

2.1.2. Deep learning approaches

The deep learning approach is considered a novelty in FER and on which several studies have been done. Before tackling the details of some of the approaches, below is an overview of the works previously undertaken on deep learning in FER. Deep learning algorithms utilize feature extraction as a way to discover and extract different characteristics. It has a multi-layered data representation architecture in which the network's lower layers serve as low-level feature extraction, and its final layers act as high-level feature extraction [18]–[20]. For processing videos, recurrent convolutional networks (RCNs) have been proposed [21]. For the processing of temporal information, convolutional neural networks (CNNs) are applied to video frames before being input into a recurrent neural network (RNN). With little training data, these models function well when the target ideas are complicated, but they have drawbacks when using deep networks. With the help of DeXpression [22], robust facial recognition has been attempted to solve this difficulty. Each block has layers like convolutional, pooling, and rectified linear units (ReLU). It achieves greater performance by combining numerous features instead of using single features.

A multitask global-local network (MGLN) for facial expression identification is suggested Yu *et al.* [23], which includes two modules : a part-based module (PBM) that learns temporal data from the areas around the mouth, nose, and eyes, and a global face module (GFM) that extracts spatial characteristics from the frame with the peak expression. Using a CNN and an long short-term memory (LSTM) network, GFM and PBM features are combined to capture substantial facial expression variance [23], [24]. More work was also done on the deep neural network e.g. [25] worked with an 18-layered CNN paired with four pooling layers. Moreover, [26] along with [27] in both of their papers have further developed deep face recognition systems. Ding *et al.* [26] introduced their new architecture FaceNet2ExpNet, on which [22] have worked and added transfer learning. For instance, [28] have applied transfer learning using AlexNet - a very strong pre-trained model having eight layers in total (five convolutional and 3 fully connected) - combined with SVM classifiers. In the same context, knowledge distillation was previously designed to compress a collection of deep neural networks [29]. The goal is to build smaller models (student models) that fulfill the same function as larger models (teacher models) with the requirement that the student model outperform the baseline model [30].

2.2. Facial expression recognition under occlusion

Numerous psychosocial studies were conducted after it was discovered that facial occlusion significantly affected FER. These studies aimed to identify the facial features that were crucial for human perception and recognition of facial expressions from partially obscured faces. Recent research has concentrated on how to use deep neural networks to immediately execute FER on covered face images, without including procedures like facial occlusion detection, feature extraction by hand, or classifier development [31].

Recent research has incorporated more forms of occlusion and datasets, however the majority of existant researchs are largely founded on a restricted number of artificially created types of occlusions, and progress is still modest. Many more papers were produced within the same context with more techniques; nonetheless we will be focusing on CNNs and DCNNs.

3. METHODOLOGY

The goal of this work is to recognize facial expressions on masked faces. In this part, a solution to the problem of performing face expression recognition has been put forth as shwon in Figure 1. Our approach is divided into three steps. First, we apply Face detection method which is responsible for detecting face in the input frames. Second, we use facial recognition to identifying or verifying the identity of a person. Finally we use emotion detection to classify the expression as one of the seven fundamental human expressions (angry, surprise, fear, disgust, happiness, neutral, sadness).

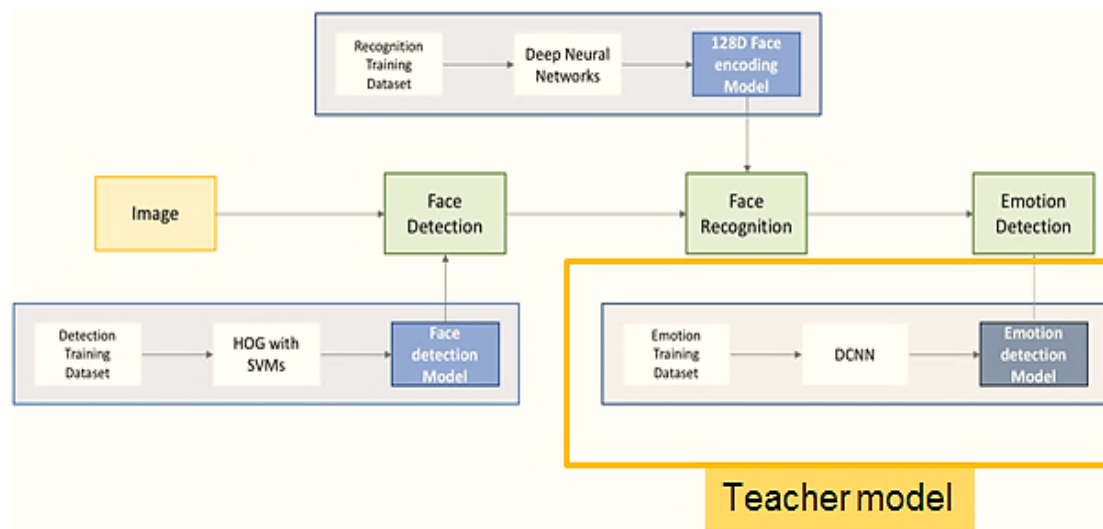


Figure 1. The stages of the proposed system for recognizing facial expressions

3.1. Face detection

Unprocessed face images contain enormous amounts of data, and feature extraction is necessary to condense this data into smaller sets known as features. Different techniques allow for face detection [32], we may use a histogram of oriented gradients (HOG) and linear SVM object detector, or even the built-in Haar cascades. Alternatively, we might also rely on deep learning-based algorithms to locate faces. Nonetheless, regardless of the method used to recognize the face in the image, it is more crucial to acquire the face bounding box. In fact, face detection using traditional feature descriptors and linear classifiers was very effective. In this paper we are going to be using dlib's face detection pre-trained model since the focus will be mainly on the facial expression detection model.

3.2. Face recognition

Face recognition is the second step after face detection. Once we retrieved the face boxes, we can move to matching them with our database. In order to perform this, we are using Dlib's face recognition module. This latter is a pretrained model that matches people's faces based on image embeddings. The face embedding of a person on a previously unseen image may be determined by calculating the distance between the face embedding and that of a known person, and if the face embedding is near enough to the embedding of person X, we can claim that the image includes the face of person X.

For Dlib face-recognition to be trained, a large number of pictures of people were used. A random vector is generated for each image at the beginning of training, so when the photos are plotted, they will be dispersed freely. Then, the model uses the triplet loss concept, in which it picks each time a reference image, a positive element (matching image) and a negative element (unmatching image). The ResNet network used in Dlib face-recognition has 29 convolutional layers, is based on the ResNet-34 network, but some layers have been eliminated, and the number of filters per layer has been reduced by half.

3.3. Facial expression recognition

3.3.1. Deep convolutional neural network

The strength of DCNNs is in their layering. A DCNN processes the image's red, green, and blue elements all at once using a three-dimensional neural network. Compared to standard feed forward neural networks, this drastically lowers the quantity of artificial neurons needed to process an image. The architecture of a convolutional network typically consists of four types of layers: Convolutional layers, pooling layers, non-linear layers, and fully connected layers. Convolution layer is used to extract low-level characteristics such as edges. It is composed of several convolutional filters (kernels). The output feature map is produced by convolving these filters with the input image, which is expressed as N-dimensional metrics. Second, the pooling layer decreases the feature resolution. Features are more resistant to noise and distortion as a result. You can pool in two different methods - the maximum pooling method and the average. Third, a non-linear "trigger" function is used to indicate different identification of probable features on each hidden layer of a neural network. This non-linear triggering may be easily implemented by using a range of specialized functions, such as ReLUs and continuous trigger functions (non linear). The last fully connected layers compute the class scores on the entire original image. Because there are so many variables to modify in DCNN, training a large DCNN model can be challenging. A vast network frequently needs a huge amount of training data as overfitting can occur while training with small or insufficient quantities of data. It can also be challenging to get enough data for the DCNN to be properly trained for some scenarios. Furthermore, in certain situations, a large volume of data is not easily available. Needless to add that network structures have grown nowadays from a dozen of layers to hundreds of layers, while trying to score high performance and compensate for the lack of data. For instance, DCNNs have moved from AlexNet with 8 layers to ResNet with 152 layers. So, training a deep network with numerous parameters would demand massive computational resources and would be impossible to run on smaller devices with smaller processors and memory. Nonetheless, transfer learning and knowledge distillation have proved to be able to solve both challenges.

3.3.2. Transfer learning

Transfer learning is a machine learning approach in which a model that has been trained for a specific function is reused for a new one. In classic deep learning scenarios, a model can perform poorly if it was reused in a new context as the model is biased from the original data. Transfer learning for deep machine learning is a procedure of first trained on a benchmark dataset, and the best-learned network feature (the network's weights and structures) is then transferred to a second network that will be trained on a target dataset.

3.3.3. Knowledge distillation

The main goal of knowledge distillation (KD) is to produce a smaller model (student) that is capable of solving the same issue as a larger model (teacher), with the caveat that the student model must outperform the regular traditional model [33]. KD uses a teacher-student architecture. The architecture of a neural network is set up to allow us to apply a "softmax" activation function and obtain the probability of the classes that are being classified. According to [34], the general equation for such an output layer is

$$y_i = \exp(x_i/T) / \sum_j \exp(x_j/T)$$

Where j is the number of classes, x_i is the logit, and y_i is the class probability. The temperature value, T , is indicated here and is typically 1 [34]. A higher temperature value represents a softer probability distribution for the classes. The basic goal of distillation is to move this knowledge from the large model to the smaller model by transferring these probabilities. The student model was trained Bucilua *et al.* [34] to produce accurate labels in addition to the teacher's soft labels by using a weighted average of two functions. Cross entropy with correct labels is the first objective function, but cross entropy with soft labels is regarded as the second objective function. The distillation process is a spread by the custom loss function like in (1),

$$L_{\text{student}} = \alpha L_{\text{CL}} + (1 - \alpha) L_{\text{KD}} \quad (1)$$

$$L_{\text{KD}} = T^2 \text{KL}(y_s, y_t)$$

Here, T is the temperature value, y_s , y_t are the targets softened for the student model, and L_{KL} is the built-in Kullback-Leibler (KL) divergence loss, L_{CL} is the normal cross-entropy loss. The hyperparameter accentuates the difference between the two loss functions' weighted average. Different methods can be used to apply the knowledge distillation framework to networks; in this paper, we employed offline distillation.

3.3.4. Our approach

Figure 2 shows our proposed technique for FER for masked faces. Our contribution concerns to incorporate knowledge distillation method in transfer learning between a teacher model, which is the full face FER DCNN, and the student model, which is the masked face FER DCNN via the combination of both the loss from the teacher soft-labels vs. The student soft labels and the loss from the dataset hard-labels vs. The student hard-labels.

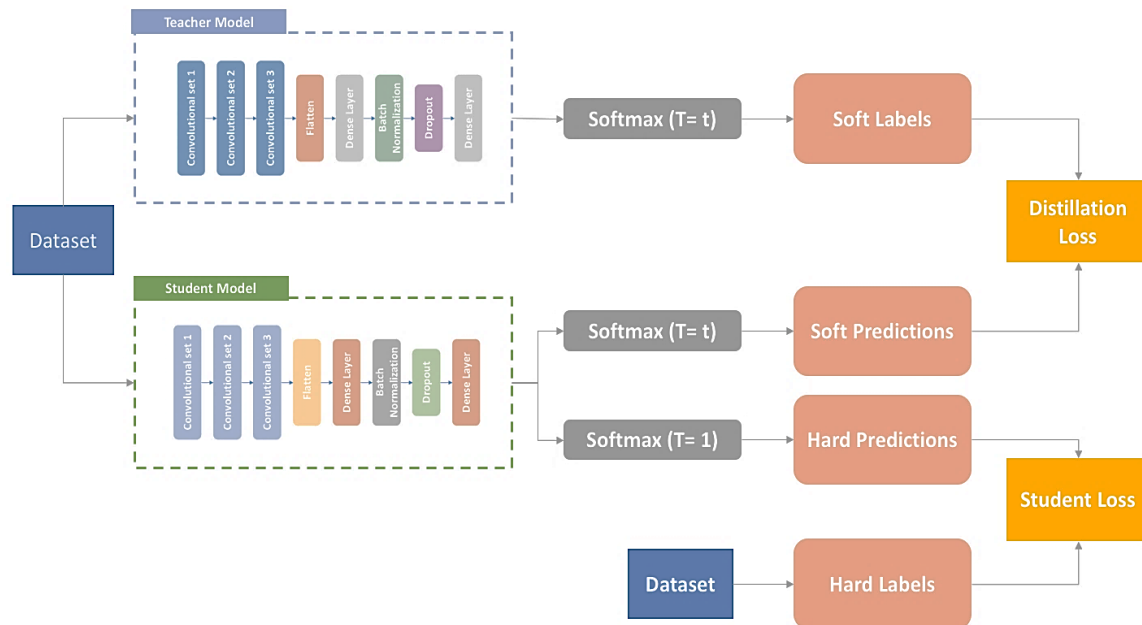


Figure 2. System architecture

A. Full face detection, recognition, and expression recognition

Once the dataset has been obtained, hybrid sampling is employed to ensure equality across all classes in the dataset. On the dataset, face detection is done in order to maintain the face region alone and remove any extraneous data. Unmasked faces can be found in this new dataset. Deep neural network algorithms are then utilized to extract the pertinent information prior to recognition. To identify FER in the full face, we are using a DCNN model. The suggested model is: a feature extraction part is comprised of 3 sets of convolutional layers, each followed by a pooling layer and a dropout layer. Each convolutional set has 2 cascading convolutional layers with their corresponding activation layers. For bigger and deeper networks, this is typically a good approach, because several stacked convolutional layers can create more complex features in the input volume before the pooling process. Then, the network ends with two fully connected layers. This means there are a total of 6 convolutional layers and 2 fully connected layers in our model. This approach greatly reduces the number of parameters that may be trained. A result of the successive convolutions in each layer is that the output becomes increasingly sensitive to tiny changes in the input, which ultimately helps classification.

B. Masked face detection, recognition, and expression recognition

Taking into consideration that masks are a type of occlusion of the face, face recognition under occlusions can be handled in two different ways: techniques that characterize the face despite the occlusions, and techniques that reconstruct the occluded portions of the face in order to recover the ideal analysis environment. In order to allow for FER, we will be using an offline knowledge distillation method with teacher-student architecture in which the student has the same size as the teacher. The teacher would be a 'homemade' pre-trained model that had learned FER on full faces. Nevertheless, the knowledge distillation would not be used in order to reduce the size of the teacher model but rather to transfer the learned knowledge that the teacher acquired from training on full faces to help guide the student in the masked faces learning. Hence, knowledge distillation would be used in transfer learning from teacher to student. The transfer learning would not be via attribution of teacher network's weights to the student, but rather through the soft labels' loss. The choice of this method was to allow the student to initialize and update its weights based on the hard labels from the dataset and the soft labels from the teacher model. We would not be initializing the model with the teacher

weights, since the new task to solve is different than the task solved by the teacher network and we only want the teacher to guide the student through the training. This is similar to how humans are attempting to recognize facial expressions now with masks. Finally, the model would be able to classify masked human faces according to the basic 7 facial expressions (disgust, fear, happiness, sadness, anger, surprise and neutral face).

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Dataset

In this study we used FER2013, the facial expression recognition (FER2013) [35] database consists 35,887 grayscale images of 48×48 resolution, split in 28,709 images of trains, and the validation and test sets include 3,589 images each. Each of the 35,887 data points in the FER-2013 dataset has been assigned a label based on seven emotions. We also tried to compensate for data challenges, hence we performed data augmentation to train model on invariances on small transformations. We used rotation up to 45 degrees, width shifts and height shifts up to 0.1, horizontal flips, zooming up to 0.2. We also normalized images (pixel values) as they can impact the neural network if not normalized. Figure 3 shows a part of dataset. Since datasets of masked faces for FER were not available, we decided to create our own dataset based on the same FER2013 dataset and the mask the face script created by Anwar and Raychowdhury [36]. It is a software that can be used to add occlusion to faces in images and mainly in the form of masks. For the application of the mask, it utilizes a Dlib-based face landmarks detector to find facial features. Figure 4 shows an example of the results obtained.



Figure 3. Images without mask



Figure 4. Images with mask

4.2. Experiments

4.2.1. Experiment 1: full face model

In this stage we start first by detecting all faces on images by using the default model of Dlib's face detection which is based on HOG and linear SVMs. Second, we are using the pre-trained model of Dlib, incorporated in the face_recognition library, to generate the 128-dimensioned embeddings for our dataset, i.e., the faces that can be recognized by the program. Next, the library's recognition function uses the KNN algorithm to do the matching and classification of faces. Once the faces detection and recognition phases are prepared, the next step is FER. We built a DCNN model. We used the kernel initializer 'he_normal' to initialize the network weights. We employed a NAdam optimizer for optimization. The chosen learning rate 0.001. Moreover, to face the main deep learning challenges presented above, we have added frequent dropouts in between each convolutional set to avoid overfitting and add more generalization to the model. We also used early stopping to avoid overfitting or underfitting caused by choosing the wrong epoch number. Overfitting of the training dataset can occur when there are too many epochs, and underfitting can occur when there are too few. This problem can be solved by early stopping because we may provide a high number of training epochs and the model will stop once the performance ceases to increase. We conditioned Early stopping on the validation accuracy metric.

Another key element is using the best weights from training. By default, the model weights obtained during the final training stage are adopted. Nonetheless, early stopping provides the option of restoring model weights from the epoch with the best value of the monitored quantity. So, we set restore_best_weights as true.

Moreover, we used 'ReduceLROnPlateau' to reduce the learning rate when validation_accuracy had stopped increasing. This callback tracks validation_accuracy and if no progress is made after 11 number of epochs (called 'patience'), the learning rate is decreased.

4.2.2. Experiment 2: masked face FER

For the masked faces, the teacher and student architecture are the same as the full-face model. Masked face FER without KD: in this experiment, we using the same steps as ran a model without knowledge distillation directly on masked faces dataset. The teacher model had learned FER on full faces. Masked face FER with KD: in this experiment, there is no need to train the teacher model again since we have already performed it. A new distiller (student, teacher) class is created, and it replaces the model methods compile (), train_step and test_step. The distiller employs the trained teacher model, the student model that will be trained, the student loss function, the distillation loss function, along with the selected temperature and an alpha factor to weight the student and distillation loss. In the train_step method, we perform a forward pass of both the teacher and student, calculate the loss with weighting of the student_loss and distillation_loss by alpha and 1 - alpha, respectively, and perform the backward pass. Note: due to the fact that only the student weights are updated, we only compute the gradients for the student weights. Using the provided dataset, we evaluate the student model using the test step method. Thus, we instantiate the distiller and compile it with the required losses, and hyperparameters. We set alpha to be 0.1 and Temperature to 10.

4.3. Results

In this section we given the results obtained after tested our models. FER models tested are evaluated with different evaluation metrics. Both the teacher and the student models were trained for 100 epochs and it took around 6–8 h to train them on the FER2013 dataset and the masked FER2013 respectively. Full face FER: teacher results: after evaluating our experiment, we obtained the results the Figure 5. The result demonstrates the behavior of model with accuracy and loss during training and testing on 100 epochs. The epochs' history shows that accuracy gradually increased and achieved 69% accuracy on both training and validation dataset, but the model stopped before reaching 100 to avoid overfitting.

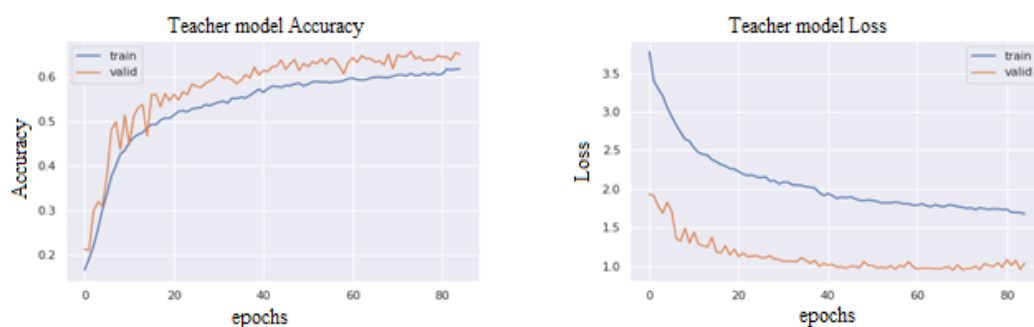


Figure 5. Model loss and accuracy graph for training and validation of teacher model

The confusion matrix as shows in Figure 6 shows that our model is performing well on the Happiness class, followed by surprise, neutral, sadness, anger, fear, and the last one is disgust. One of the reasons for disgust coming last is that the class has less data. For fear, as we can see on the confusion matrix, major parts of the test dataset are misclassified as surprise, or sadness, or even anger. This in fact makes sense since fear can exhibit a mixture of all these emotions at once, thus having each human react differently.

The classification report as shown in Table 1, confirms as well what the confusion matrix states. Happiness has the highest precision, recall, and F1-score. The average accuracy of the model is 69%. We ran a model without knowledge distillation to test the impact our approach will have. The student had the same structure as the teacher model and ran directly on masked faces dataset. The model reached an accuracy of 46%. Results are the Table 2.

Masked face FER with KD: student results: after evaluating our experiment, we obtained the results the Figure 7. The result demonstrate the behavior of model with accuracy and loss during training and testing on 160 epochs. The epochs' history shows that accuracy gradually increased and achieved 61% accuracy on both training and validation dataset, but the model stopped before reaching 100 to avoid overfitting.

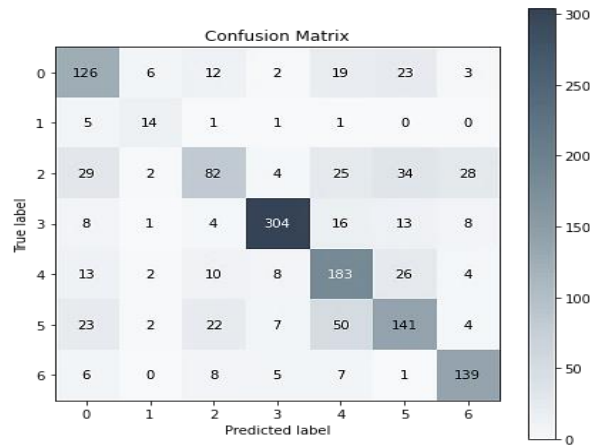


Figure 6. Teacher model confusion matrix of 7-class FER

Table 1. Results for our approach

	Classification		Report	
	Precision	Recall	F1Score	Support
Anger	0.60	0.66	0.63	191
Disgust	0.52	0.64	0.57	22
FEAR	0.59	0.40	0.48	204
Happiness	0.92	0.86	0.89	354
Neutral	0.61	0.74	0.67	246
Sadness	0.59	0.56	0.58	246
Surprise	0.75	0.84	0.79	166
Accuracy			0.69	1,432
Macro avg	0.65	0.67	0.66	1,432
Weighted avg	0.69	0.69	0.69	1,432

Table 2. Results for the model without knowledge distillation

	Classification		Report	
	Precision	Recall	F1Score	Support
Anger	0.34	0.43	0.38	641
Disgust	0.17	0.44	0.25	85
FEAR	0.34	0.26	0.30	645
Happiness	0.73	0.46	0.56	1,380
Neutral	0.41	0.53	0.46	930
Sadness	0.40	0.25	0.31	664
Surprise	0.52	0.80	0.63	610
Accuracy			0.46	4,955

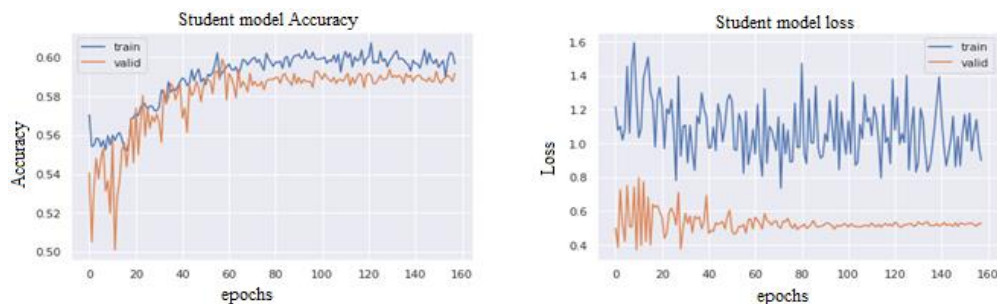


Figure 7. Model loss and accuracy graph for training and validation of student model

4.4. Analysis of results

The FER-2013 dataset presents extra difficulties compared to the other FER datasets we used. Another significant issue in this dataset, in addition to the intra-class variation of FER, is the imbalanced nature of the various emotional classes. There are much more examples for certain classes than others, such as neutral and

happiness. The model was trained using all 28,709 images in the training set, which was then tested against 3,500 validation images and reported on for accuracy using 3,589 test images. On the full face dataset and the masked face dataset, we were able to attain accuracy rates of about 69% and 61%, respectively. The table as shows in Table 3 compares the output of our model with a few earlier studies on the FER2013 full face.

Table 3. The accuracy of each method

Method	Accuracy
GoogleNet - full face	65.2%
Unsupervised domain adaptation - full face	65.3%
FER on SoC - full face	66%
Bag of Words - full face	67.4%
Our approach - full face	69%
Our approach - Masked face	61%
Aff-wild2 (based on VGG) - full face	75%
Human performance	65%

5. CONCLUSION

In this article, we proposed a method to detect FER with masks based on a teacher-student architecture using knowledge distillation in transfer learning. The study proved that using both the distillation loss and the student loss as a form of transfer learning can be beneficial as it upgraded the model accuracy from a regular DCNN with 46% accuracy to a student model with 61%. Hence, based on a 69% accuracy teacher that learned facial expressions on full faces, the student managed to score 61% on masked faces (lower face occluded). Moreover, comparing the teacher model with other major models ran on the same dataset, our model showed significant accuracy, knowing that the FER2013 dataset pose a greater challenge than they do in other FER datasets as it encompasses the difficult naturalistic conditions. In fact, the human performance on this dataset is estimated to be 65.5%. For future work, we plan to explore the context to improve our performance in occluded FER.




REFERENCES

- [1] S. Li and W. Deng, "Deep facial expression recognition: a survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 3, pp. 1195–1215, Jul. 2022, doi: 10.1109/TAFFC.2020.2981446.
- [2] E. G. Moung, "Face recognition state-of-the-art, enablers, challenges and solutions: a review," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 1.2, pp. 96–105, Apr. 2020, doi: 10.30534/ijatcse/2020/1691.22020.
- [3] Y. Pratama, L. M. Ginting, E. H. Laurencia Nainggolan, and A. E. Rismanda, "Face recognition for presence system by using residual networks-50 architecture," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 6, p. 5488, Dec. 2021, doi: 10.11591/ijece.v11i6.pp5488-5496.
- [4] F. A. Abd Almuhsen and Z. A. Khalaf, "Review of different combinations of facial expression recognition system," *Journal of Physics: Conference Series*, vol. 1591, no. 1, p. 12020, Jul. 2020, doi: 10.1088/1742-6596/1591/1/012020.
- [5] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: a comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, May 2009, doi: 10.1016/j.imavis.2008.08.005.
- [6] J. S. Nayak, P. G. M. Vatsa, M. Reddy Kadiri, and S. S., "Facial expression recognition: a literature survey," *International Journal of Computer Trends and Technology*, vol. 48, no. 1, pp. 1–4, Jun. 2017, doi: 10.14445/22312803/IJCTT-V48P101.
- [7] M. W. Schurgin, J. Nelson, S. Iida, H. Ohira, J. Y. Chiao, and S. L. Franconeri, "Eye movements during emotion recognition in faces," *Journal of Vision*, vol. 14, no. 13, p. 14, Nov. 2014, doi: 10.1167/14.13.14.
- [8] C. Sawangwong, K. Puangsuwan, N. Boonnam, S. Kajornkasirat, and W. Srisang, "Classification technique for real-time emotion detection using machine learning models," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 4, p. 1478, Dec. 2022, doi: 10.11591/ijai.v11.i4.pp1478-1486.
- [9] F. Bourel, C. C. Chibelushi, and A. A. Low, "Recognition of facial expressions in the presence of occlusion," in *Proceedings of the British Machine Vision Conference 2001*, 2001, pp. 23.1–23.10, doi: 10.5244/C.15.23.
- [10] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006, doi: 10.1109/TPAMI.2006.244.
- [11] X. Huang, G. Zhao, W. Zheng, and M. Pietikainen, "Towards a dynamic expression recognition system under facial occlusion," *Pattern Recognition Letters*, vol. 33, no. 16, pp. 2181–2191, Dec. 2012, doi: 10.1016/j.patrec.2012.07.015.
- [12] S. Kuruvayil and S. Palaniswamy, "Emotion recognition from facial images with simultaneous occlusion, pose and illumination variations using meta-learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 9, pp. 7271–7282, Oct. 2022, doi: 10.1016/j.jksuci.2021.06.012.
- [13] R. Cowie, K. Karpouzis, G. Caridakis, and M. Wallace, *Multimodal user interfaces*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
- [14] H. Alshamsi and V. Kupuska, "Real-time facial expression recognition app development on smart phones," *International Journal of Engineering Research and Applications*, vol. 07, no. 07, pp. 30–38, Jul. 2017, doi: 10.9790/9622-0707033038.
- [15] E. Kremic and A. Subasi, "Performance of random forest and SVM in face recognition," *International Arab Journal of Information Technology*, vol. 13, no. 2, pp. 287–293, 2016.
- [16] H. S. Dadi, G. K. M. Pillutla, and M. L. Makkena, "Face recognition and human tracking using GMM, HOG and SVM in surveillance videos," *Annals of Data Science*, vol. 5, no. 2, pp. 157–179, Jun. 2018, doi: 10.1007/s40745-017-0123-2.
- [17] T. X. Tee and H. K. Khoo, "Facial recognition using enhanced facial features k-nearest neighbor (k-NN) for attendance system," in *Proceedings of the 2020 2nd International Conference on Information Technology and Computer Communications*, Aug. 2020, pp. 14–18, doi: 10.1145/3417473.3417475.




- [18] F. Zhi-Peng, Z. Yan-Ning, and H. Hai-Yan, "Survey of deep learning in face recognition," in *2014 International Conference on Orange Technologies*, Sep. 2014, pp. 5–8, doi: 10.1109/ICOT.2014.6954663.
- [19] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 2, pp. 1015–1027, 2020, doi: 10.11591/ijeecs.v18.i2.pp1015-1027.
- [20] E. Gubin Mounq, C. Chuan Wooi, M. Mohd Sufian, C. Kim On, and J. Ahmad Dargham, "Ensemble-based face expression recognition approach for image sentiment analysis," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 12, no. 3, p. 2588, Jun. 2022, doi: 10.11591/ijece.v12i3.pp2588-2600.
- [21] J. Donahue *et al.*, "Long-term recurrent convolutional networks for visual recognition and description," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 2625–2634, doi: 10.1109/CVPR.2015.7298878.
- [22] J. Li, T. Qiu, C. Wen, K. Xie, and F.-Q. Wen, "Robust face recognition using the deep C2D-CNN model based on decision-level fusion," *Sensors*, vol. 18, no. 7, p. 2080, Jun. 2018, doi: 10.3390/s18072080.
- [23] M. Yu, H. Zheng, Z. Peng, J. Dong, and H. Du, "Facial expression recognition based on a multi-task global-local network," *Pattern Recognition Letters*, vol. 131, pp. 166–171, Mar. 2020, doi: 10.1016/j.patrec.2020.01.016.
- [24] M. A. Saleh, A. T. Yong, Y. M. Yusoff, and N. Marbukhari, "Facial expression recognition: a new dataset and a review of the literature," *Turkish Online Journal of Qualitative Inquiry*, vol. 12, no. 6, pp. 9804–9811, 2021.
- [25] D. Y. Liliana, "Emotion recognition from facial expression using deep convolutional neural network," *Journal of Physics: Conference Series*, vol. 1193, p. 12004, Apr. 2019, doi: 10.1088/1742-6596/1193/1/012004.
- [26] H. Ding, S. K. Zhou, and R. Chellappa, "FaceNet2ExpNet: regularizing a deep face recognition net for expression recognition," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, May 2017, pp. 118–126, doi: 10.1109/FG.2017.23.
- [27] S. Aneja, N. Aneja, P. E. Abas, and A. G. Naim, "Transfer learning for cancer diagnosis in histopathological images," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 1, p. 129, Mar. 2022, doi: 10.11591/ijai.v11.i1.pp129-136.
- [28] S. Shaees, H. Naeem, M. Arslan, M. R. Naeem, S. H. Ali, and H. Aldabbas, "Facial emotion recognition using transfer learning," in *2020 International Conference on Computing and Information Technology (ICCIT-1441)*, Sep. 2020, pp. 1–5, doi: 10.1109/ICCIT-144147971.2020.9213757.
- [29] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, vol. 2, no. 7, Mar. 2015.
- [30] J. Yim, D. Joo, J. Bae, and J. Kim, "A gift from knowledge distillation: fast optimization, network minimization and transfer learning," in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4133–4141.
- [31] Y. Cheng, B. Jiang, and K. Jia, "A deep structure for facial expression recognition under partial occlusion," in *2014 Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Aug. 2014, pp. 211–214, doi: 10.1109/IIH-MSP.2014.59.
- [32] K. Slimani, M. Kas, Y. El Merabet, Y. Ruichek, and R. Messoussi, "Local feature extraction based facial emotion recognition: a survey," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, p. 4080, Aug. 2020, doi: 10.11591/ijece.v10i4.pp4080-4092.
- [33] A. Mishra, D. Marr, and I. Labs, *Apprentice : using knowledge distillation techniques to improve low -precision net - work a curacy*. 2018.
- [34] C. Bucilua, R. Caruana, and A. Niculescu-Mizil, "Model compression," in *In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2020, pp. 535–541.
- [35] P. L. C. Courville, A. Goodfellow, I. J. M. Mirza, and Y. Bengio, "FER-2013 face database," Universit de Montreal QC Canada, 2013.
- [36] A. Anwar and A. Raychowdhury, "Masked face recognition for secure authentication," *arXiv preprint arXiv:2008.11104*, pp. 1–8, Aug. 2020.

BIOGRAPHIES OF AUTHORS



Boutaina Hdioud    holds a Doctor of Informatics degree from ENSIAS, Mohammed V University Morocco in 2017. She is currently an associate professor at Computer Science Department in High National School for Computer Science and Systems Analysis - ENSIAS, Rabat, Morocco. She is a member of research group IRDA (Information Retrieval and Data Analytics) of the research laboratory ADMIR (Advanced Digital enterprise Modeling and Information Retrieval) of ENSIAS. Her research includes computer vision, Reconnaissance faciale, machine learning, NLP. She can be contacted at boutaina.hdioud@um5.ac.ma.



Mohammed El Haj Tirari    has a Ph.D. degree in Statistics from Free University of Brussels. From 1995 to 2004, he was an associate professor and researcher at ULB, from 2004 to 2010 he was Lecturer/Researcher at the National School for Statistics and Information Analysis (Rennes, France). He also taught at the University Pierre and Marie Curie (Paris, France). Since 2010, he is Full Professor at National Institute of Statistics and Applied Economics-Morocco. He participated as trainer in EU training programs: Medstat II, Medstat III and AMICO. He has several years experience in the field of applied statistics, survey sampling. More specifically, sampling methods, estimation methods, estimation of the precision in complex survey designs, treatment of non-response and treatment of the measurement errors, methods of datamining and statistical modeling. He can be contacted at mtirari@hotmail.fr.