I.J. Modern Education and Computer Science, **2022, 6, 53-64**
Published Online on December 8, 2022 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2022.06.05

A Facial Expression Recognition Model using Lightweight Dense-Connectivity Neural Networks for Monitoring Online Learning Activities

Duong Thang Long*
Hanoi Open University, Hanoi, 100000, Viet Nam
E-mail: duongthanglong@hou.edu.vn
ORCID iD: https://orcid.org/0000-0003-0609-9534
*Corresponding Author

Truong Tien Tung
Hanoi Open University, Hanoi, 100000, Viet Nam
E-mail: truongtientung@hou.edu.vn

Tran Tien Dung
Hanoi Open University, Hanoi, 100000, Viet Nam
E-mail: dungtranitd@hou.edu.vn
ORCID iD: https://orcid.org/0000-0003-3921-2396

Received: 17 September, 2022; Revised: 16 October, 2022; Accepted: 20 November, 2022; Published: 08 December, 2022

Abstract: State-of-the-art architectures of convolutional neural networks (CNN) are widely used by authors for facial expression recognition (FER). There are many variants of these models with positive results in studies for FER and successful applications, some well-known models are VGG, ResNet, Xception, EfficientNet, DenseNet. However, these models have considerable complexity for some real-world applications with limitations of computational resources. This paper proposes a lightweight CNN model based on a modern architecture of dense-connectivity with moderate complexity but still ensures quality and efficiency for facial expression recognition. Then, it is designed to be integrated into learning management systems (LMS) for recording and evaluation of online learning activities. The proposed model is to run experiments on some popular datasets for testing and evaluation, the results show that the model is effective and can be used in practice.

Index Terms: Deep learning, Convolution neural network, Dense-Connectivity networks, Facial expression recognition.

1. Introduction

Facial expression recognition (FER) is very interesting in wide research today and it has high applicability in the field of computer vision. The facial expressions of human beings play an important role in any interpersonal communication, it can help others to understand one's emotions or even intentions, making it an indispensable communication element in human interaction. With the development of computer vision technology and its practical application, as referred in [1,2,3,4,5] the results of various studies on facial expression recognition have shown a lot of promising successful applications in the fields of human-machine interaction, animation, medicine and education.

Authors in [6] mentioned that Ekman and Friesen identify six basic human facial expressions which they believe these emotions are presented in all people regardless of nationality, ethnicity or religion. These expressions are happiness (Ha), sadness (Sa), surprise (Su), disgust (Di), anger (An) and fear (Fe). The changing or moving parts of human faces such as raised eyebrows, locked eyebrows, and corners of mouth moving outwards or openings and closings are considered as basic units of change in facial expressions. However, everyone's facial expressions change dramatically, and the expressions shown by different faces are also different from person to person. These factors

greatly impact the operation and efficiency of any FER system including computer vision techniques. The database system of facial expressions was established by authors, describing each expression in detail, laying the foundation for solving the FER problem. Currently, the databases for research on FER are popularly published such as CK+, JAFFE, OuluCASIA, RAF_DB, KDEF and FER2013, they are all described in [2,6,7] and they are used by many authors to experiments run for FER models.

Methods for FER problems can be divided into two categories [6,8]first one uses traditional image processing techniques such as scale-invariant feature transformation (SIFT), histogram (HOG), local binary patterns (LBP) analysis and other one uses machine learning which is geometry-based global features. The extracted features of facial expressions will be used as input for classifiers such as BP, SVM [6,9] obtain final recognition results. However, traditional image processing methods have low efficiency due to the difference, large variation of images for difference of perspectives and the simplicity of recognition model architectures. Recently, the use of deep learning technology with convolutional neural networks (CNN) has been strongly developed and brought high efficiency [6,10] Many features hidden deep in images can be detected and extracted based on training CNN models to create a powerful FER system. Therefore, they are very stable for images with face positions with difference of perspectives and scale changes [11]. There are different CNN models for FER problems that have been proposed in studies based on state-of-the-art architectures such as VGG, SENet, Xception [12], GoogleNet, ResNet [6] or EfficientNet [13-15]

In particular, DenseNet architecture [16] provides lightweight models with a rather small set of training parameters and gives better recognition, it is also used by authors as a backbone network for CNN models [12,17-18]. This architecture has a dense connectivity scheme which is designed to improve the information transferring from early layers of neurons to backward layers, it forwards and concatenates feature maps to subsequence layers and uses growth rate to establish how much each layer contributes to the global state. Specifically, this kind of models are divided into dense blocks (DB), there are several layers of densely connected neurons in a DB, that is, an earlier layer of neurons directly connects to all afterward layers. At the end of each dense block, there is a layer of neurons that plays the role of relaying information (transition layers) to the next block (Fig.1).

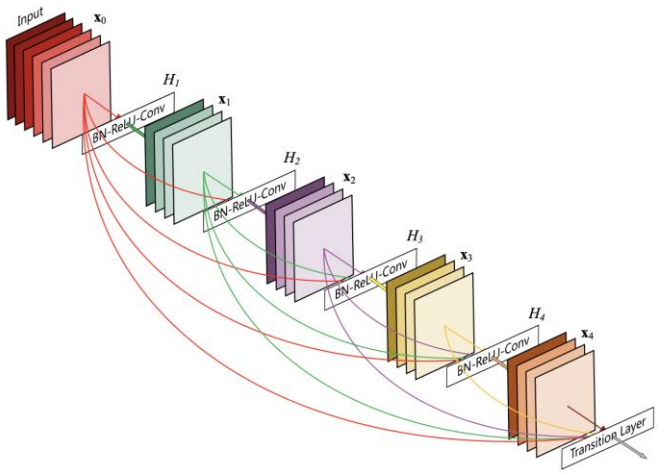

Fig. 1. Block of densely connected neurons layers illustration [16]

Although most of proposed CNN models are highly efficient, they have very complex structures, high computational costs, and require large computational systems in both training models and inference in using models. Therefore, some studies propose to use CNN models in lightweight form such as [1,11,18-21]to be suitable for limitation of computational resource systems in applied reality but still ensure the efficiency of models, high accuracy of recognition and especially these models can be published to plug-in on web-based online environment.

In this study, we propose a lightweight CNN model based on the densely connected block architecture for FER problems. The model has a small number of training parameters for high speed in both training and inferencing. It is lighter in weight and suitable in many practical applications with limitations of computational resources systems. Furthermore, to increase the quality of models in machine learning, we also apply transformations and image processing to augment training data. The next section of this paper is Part 2 which is details of the proposed CNN model and designing an integrated application system of this model with a learning management system (LMS) for recognition and supporting assessment of online learning activities. In Part 3, the model is experimentally running and evaluating on different popular datasets which are CK+, OuluCASIA and JAFFE. These datasets have some differences in illumination and face pose, one is diverse illumination in color whilst two others are grayscale. Results are analyzed and compared with other methods to assess the effectiveness. Finally, Section 4 is the conclusion.

2. Proposed Recognition System

2.1. Design dense-connectivity model

In this section, we design a CNN model based on densely connected blocks from DenseNet type architectures and integrate it into LMS systems to record and support assessment of online learning activities, it is called lightweight densely connected architecture based for facial expressions recognition (LDFER). This model is divided into three main phases as in Fig.2 including: (1) taking pictures of learners from learning devices which are connected, then pre-processing to detect face areas on images and enhance the quality of images if necessary; (2) perform features extraction from learners' images that presents information of facial expressions; and (3) classifying features to recognize labels of facial expression in order to record and support for assessment of learning activities.

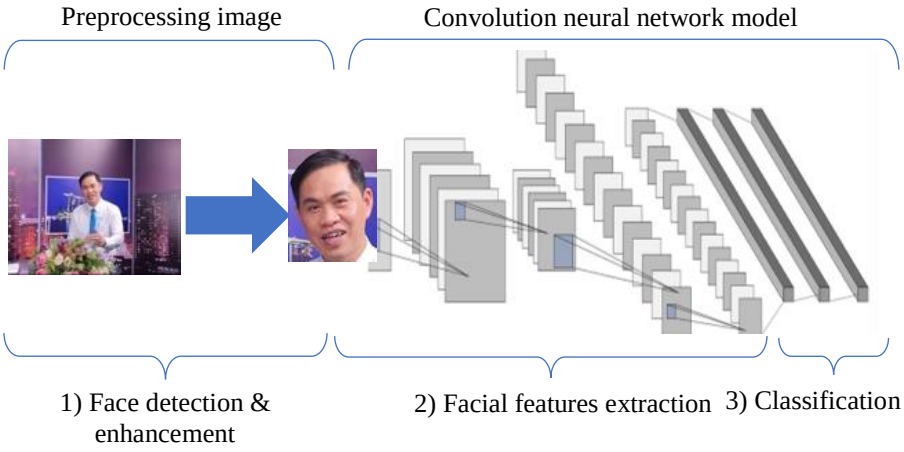

Fig. 2. The overall diagram of LDFER model

The core of LDFER model is a CNN architecture that performs two main functions: extraction of facial expression features by densely connected blocks of neurons (called dense-connectivity blocks, DB) and classification to recognize labels of facial expressions from extracted features. The structure of each DB in this model is designed by a number of neuron layers, each of which has a channel-wise concatenate connection to all its backward layers. In other words, output of an earlier neuron layer is contributed to the input channel of backward layers in the block. In order to avoid gradient explosion when training models, each neuron layer in DB is applied a batch normalization mechanism, it helps to stabilize the distribution of whole training data on the normal distribution across all neuron layers. There are number of feature signal processing (FSP) of DB, each FSP has a set of processing operations including batch normalization (B), activation of neuron by linear operation such as "relu" (R) and convolutional operations (C), we denotes B+R+C. In which, symbol C1 or C3 represents the window-size of kernel function in convolutions 1x1 or 3x3 respectively. At the end of a DB block, there is a neuron layer to aggregate information in the form of an element-wise addition, which plays the role of converting and transferring features for the next DB, this layer is denoted by TS (transition layer), it includes B+R+C1 and average pooling of 2x2 window-size (Pa). Fig.3 illustrates a DB(k) which has parameter k for the number of FSPs in the block.

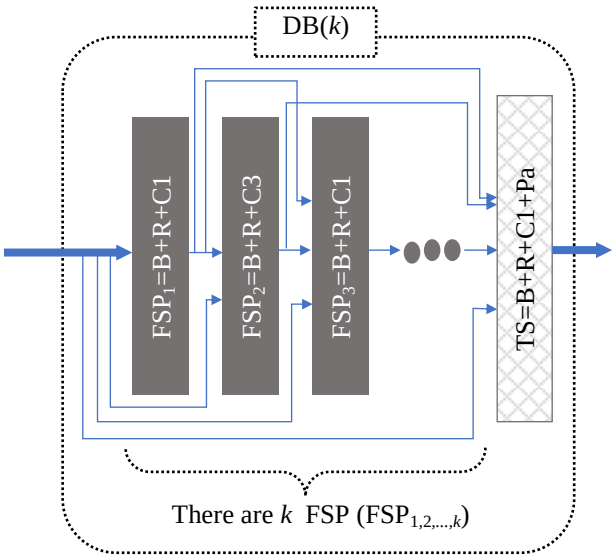

Fig. 3. Connection diagram of dense-connectivity block - DB(k)

Each DB has direct connections from an earlier layer of neurons to all subsequent layers in the block, this improves information transferring between neuron layers. Thus, the h^{th} neuron layer which corresponding to a feature signal processing (FSP_h) receives feature maps of all previous layers of neurons in the block $(x_1, x_2, \dots, x_{h-1})$ as input, and the output feature maps of this layer (x_h) is formalized as follows:

$$x_h = FSP_h([x_1, x_2, \dots, x_{h-1}]) \quad (1)$$

where, $[x_1, x_2, \dots, x_{h-1}]$ represents a continuous concatenation of feature maps generated from previous layers (from 1 to h-1). Thus, the output of each DB(k) block is formalized as a composite function from many feature signal processing and a transition layer of neurons (TS) as follows:

$$DB(k) = TS(FSP_k, FSP_{k-1}, \dots, FSP_1) \quad (2)$$

In this architecture, the number of *FSPs* and their size (number of neurons) in dense-connectivity blocks affects the quality of models in feature extraction from input images, and they are also the factors that make complexity of models. Authors often adjust these factors to design a complete connection diagram of CNN models in order to balance the quality of recognition and the computational conditions of real applied environments. The LDFER model in this study (Figure 4) uses 4 dense-connectivity blocks with sizes of 2, 4, 8 and 16 respectively, they are denoted by $DB(k)$ where $k = 2, 4, 8, 16$.

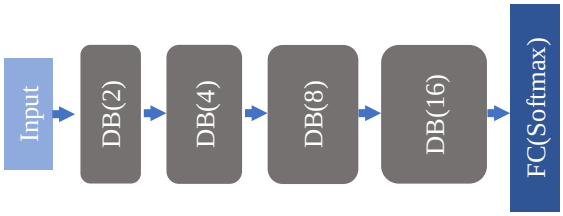

Fig. 4. Diagram of dense-connectivity blocks in LDFER

Finally, the LDFER model has a total of 34 layers of convolutional neurons which are divided into 4 dense-connectivity blocks. These blocks act as feature extraction of the model. This model has about 2.4 million parameters, which is lower than almost all other CNN models for FER problems. Although the LDFER model has quite a lot of convolution layers, we use a small number of kernel functions with small size for convolutional operations, so it results in a low number of parameters or less complexity of the model. Table 1 shows the model in details, where, "⊕" denotes concatenation of neuron layers with a certain repeated number, "→" is forward connection of two layers, "Pm" is a max-pooling operation of specified window size.

Table 1. Details of LDFER model

Type	#Layers (size - strides, number of filters)	#Parameters	#Output shape
Pre-processing	C(7x7-2, 64) → Pm(3x3-2)	3,200	24x24x64
Dense block	2 ⊕ [C(1x1-1, 128) → C(3x3-1, 32)]	95,936	12x12x128
Transition	C(1x1-1, 64) → Pa(2x2-2)	8,256	6x6x64
Dense block	4 ⊕ [C(1x1-1, 128) → C(3x3-1, 32)]	209,024	6x6x192
Transition	C(1x1-1, 64) → Pa(2x2-2)	18,528	3x3x96
Dense block	8 ⊕ [C(1x1-1, 128) → C(3x3-1, 32)]	519,552	3x3x352
Transition	C(1x1-1, 64) → Pa(2x2-2)	62,128	2x2x176
Dense block	16 ⊕ [C(1x1-1, 128) → C(3x3-1, 32)]	1,478,464	2x2x688
Pooling	Average global pooling	-	688
Classifier	Fully connected layer	4,134/4,823/5,512	6/7/8

2.2. Design integration of LDFER and LMS

In this session, we design an integrated system between LDFER model and an available LMS to automatically capture photos and recognize facial expressions of learners for recording and evaluating online learners' activities during whole learning processes. This integrated system is done by application programming interface (API) connections between an LMS and LDFER model, we make an embedded links from LDFER system into LMS. Our proposed model is published as a module that runs on client devices which are web-client or mobile apps. Learners log in LMS through their learning account (identifier and password) to authenticate for learning, the system will ask to open camera or webcam to record videos or take pictures of learning activities with facial expressions. These images are sent to LDFER system for face collection or facial expression recognition. At first using, learners must be collected face

images for training the model. Then, this trained model is used to recognize facial expressions from images of online learning activities. The recognition process is repeated at certain intervals of time to record whole learning processes. Synthesize results of recognized facial expression are stored in histories to contribute for assessment of learning quality, evaluation of learning content and teaching activities of lecturers. From the results, there are notices for learners, lecturers, administrators to know how to adjust their operations in order to get higher and higher quality of education. In order to secure this connection, LMS sends a message of userID and security code to LDFER system at starting a session. The connection diagram and operation process of this integrated system are presented in Fig.5.

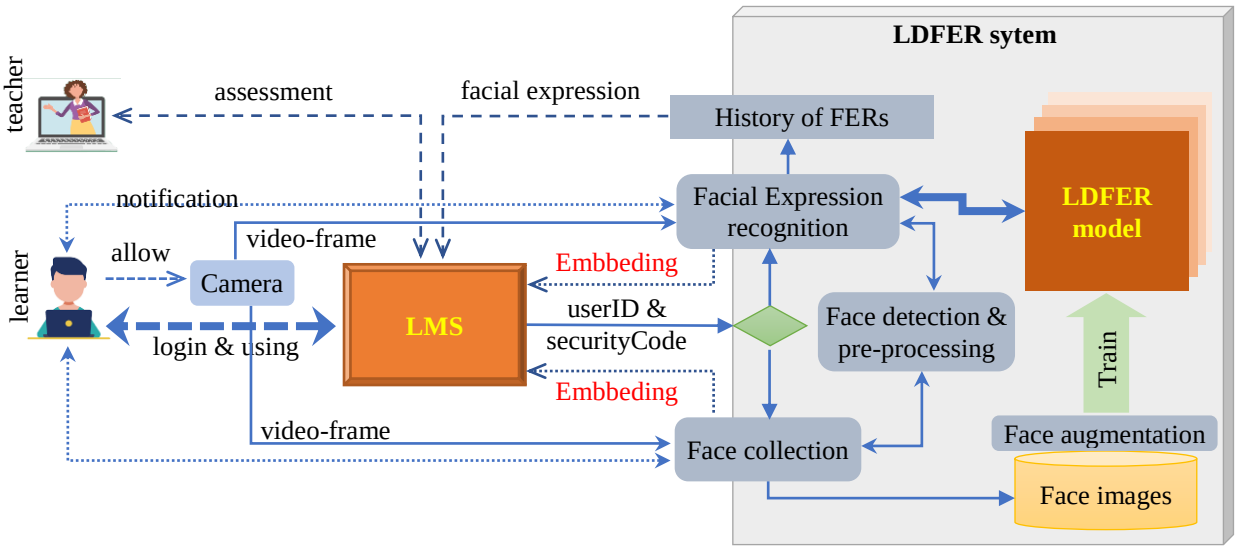

Fig. 5. Diagram of connections between LDFER model and LMS

This integrated system is designed as a quite independent connection. It requires no major modifications to an existing LMS for connecting to LDFER model. LMS can be executed independently as it is, without needing connection to the model. When LMS is connected to LDFER model, it will receive learner's facial expression recognition results during the whole learning process and use these results to synthesize and evaluate online learning activities, then we can notify learners or teachers when needed for improving the quality of education. With this design, we can easily integrate LDFER model into any existing LMS.

2.3. Image pre-processing and augmentation

In real applications, input images are usually taken from user devices, they consist of a background with any object inside. This study uses a well-known CNN-based model called MTCNN as shown in [2] to detect faces on images, then they are cut out background. In case of having no face, we can record a time of no present learners for learning, this does nothing with LDFER model.

In order to avoid overfitting in training models and make the model more stable with high accuracy in recognition, we augment training images as shown in [2,22]by using some 2D image processing techniques. These operations can be noise addition, rotation, flipping, cropping and shifting, brightening or darken color of images. With an input image a, results obtained after image augmentation operations are as follows:

$$\{\Im^{\alpha}(f^{D}(a), p^{\alpha})\} \tag{3}$$

where, f^{D} is a detector for face detection on images, such as MTCNN, p^{α} are parameters for image augmenting operations with a processing $\alpha = \{noise,\ rotation,\ zoom,\ shift,\ contrast, \ldots\}$, $\Im^{\alpha}$ represents the transformation of images for α-augmenting operation. For example, Fig.6 shows 15 augmented images with random parameters from an image in OuluCASIA dataset. The original image is in the first row and the last three rows are augmented images. The augmented images are more diverse, so the training model will give it more stability to feature extraction with any changing style, illumination, position, perspectives and so on of captured images.

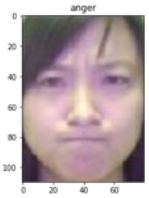

Fig. 6. Augmented images of an original one in OuluCASIA

The parameters of images augmenting operations are selected in a certain range to ensure that important information of facial expressions on images is preserved for feature extraction. For example, the fourth image in the last row of Figure 6 has a large degree of rotation and movement, this can be lost information of facial expressions, therefore it is very difficult for features extraction and recognition. An image can be simultaneously applied to some augmenting operations, in this study we randomly select values of parameters in a suitable range for augmentation.

3. Experiments

3.1. Datasets and parameters

This study uses three datasets for experiments running the LDFER model, they are CK+ (Extended Cohn-Kanade), OuluCASIA and JAFFE. We also mix these datasets into a new one for another running, it is called COJ. The dataset of CK+ has 981 images collected from 118 different people, it has seven basic expressions including anger, disgust, fear, happiness, sadness, surprise and contempt. Images in this dataset have grayscale (a channel of color). The first line of Fig.7 shows some images of CK+ with their title corresponding to facial expression labels of images.

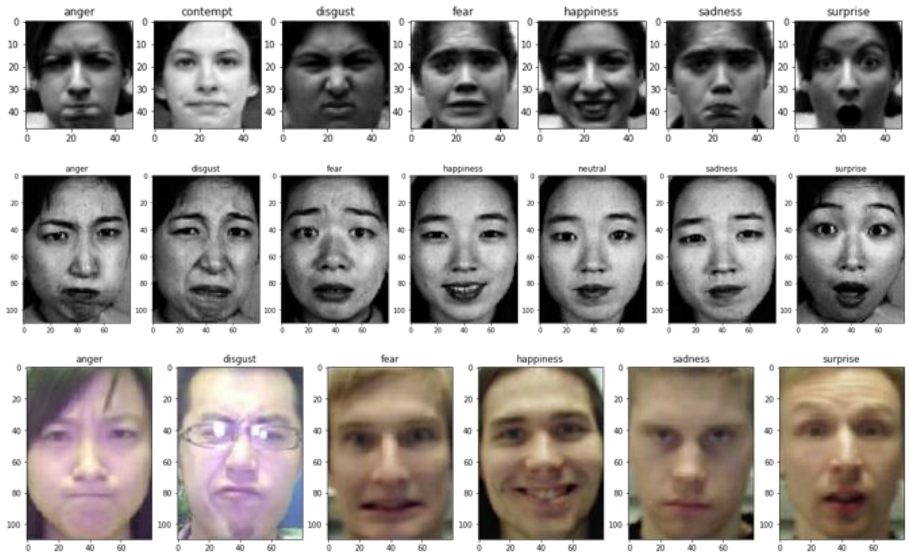

Fig. 7. Example images from three datasets

The JAFFE dataset contains 213 images from 10 different women in Japan. Each person has images with six basic facial expressions including anger, disgust, fear, happiness, sadness and surprise. It also has a neutral expression in some images. This dataset is challenging to train models because it contains very few images in each category. The second line in Fig.7 shows images of seven facial expressions of JAFFE. This dataset is also grayscale. There are 1440

images in the OuluCASIA dataset. It has six facial expressions as CK+ without contempt. This dataset was collected from 80 different people under various illumination conditions and head pose. The last line of Fig.7 shows images in every facial expression of OuluCASIA. Table 2 describes in detail the distribution of images in facial expressions for three datasets. We also create a mixed dataset by taking all images from CK+, JAFFE and OuluCASIA (called CJO dataset). This new one has 2634 images in total with eight labels of facial expressions. We have to convert images into the same size and channels of color, so, color images in OuluCASIA are converted into grayscale. There are two facial expressions of contempt and neutral which have few images compared to others because they are only in one original dataset. So, the CJO dataset has imbalanced classes of images (e.g., class of "surprise" has more than 10 times "neutral").

Table 2. Distribution of images in facial expressions

Facial expressions	CK+	JAFFE	OuluCASIA	CJO
anger	135	30	240	405
contempt	54	-	-	54
disgust	177	29	240	446
fear	75	32	240	347
happiness	207	31	240	478
neutral	-	30	-	30
sadness	84	31	240	355
surprise	249	30	240	519
Total	**981**	**213**	**1440**	**2634**

For experimental running, we use a popular scenario of *k*-fold cross-validation, *k*=5 in our case. Thus, a dataset is randomly divided into 5 folds of images in the same size. In each run of training model, we use a fold for testing (D^{te}) and another fold for evaluation of selecting model (D^{va}), the remaining folds is used to train model (D^{tr}). This scenario is repeated 5 times with every fold for testing, the final result is calculated by mean and deviation of 5 runs. We augment training data by applying $\Im^{\alpha}$ image transformations as in E.q (3). Parameters of image augmenting operations are chosen randomly within a range given in Table 3. Every image of datasets is augmented by 10 times with random parameters, so, the training data is enlarged 10 times. This brings data diversity, it can avoid overfitting and get high accuracy of recognition.

Table 3. Parameters of experimental running

No	Parameters	Value
1	Maximum rotation relative to original image (radian, negative is left rotation)	±0.1π
2	Maximum movement relative to size of original image (percentage, negative is left shifting)	±10%
3	Maximum contrast (ratio)	0.1
4	Maximum noise coefficient according to Gausian	0.1
5	Maximum scaling relative to size of original image (negative is downscaling)	±10%
6	Initial learning rate (using Adam method [23])	10^{-3}
7	Batch size	128
8	Epoches	150

In this study, we use Tensorflow as the deep learning framework. A computer system with TPU and 32GB RAM is used to train our model and conduct experiments.

3.2. Results and discussion

Loss and accuracy of the training LDFER model are averaged over 5 runs under a 5-fold cross-validation scenario as shown in Fig.8. Each subfigure is shown loss and accuracy of both training data (solid line) and validation data (dotted line). There are many changes in the training process of JAFFE, because few images with facial expressions are closer in classes. For training data, they all have about the first 50 epoches to get the high accuracy and low loss, corresponding.

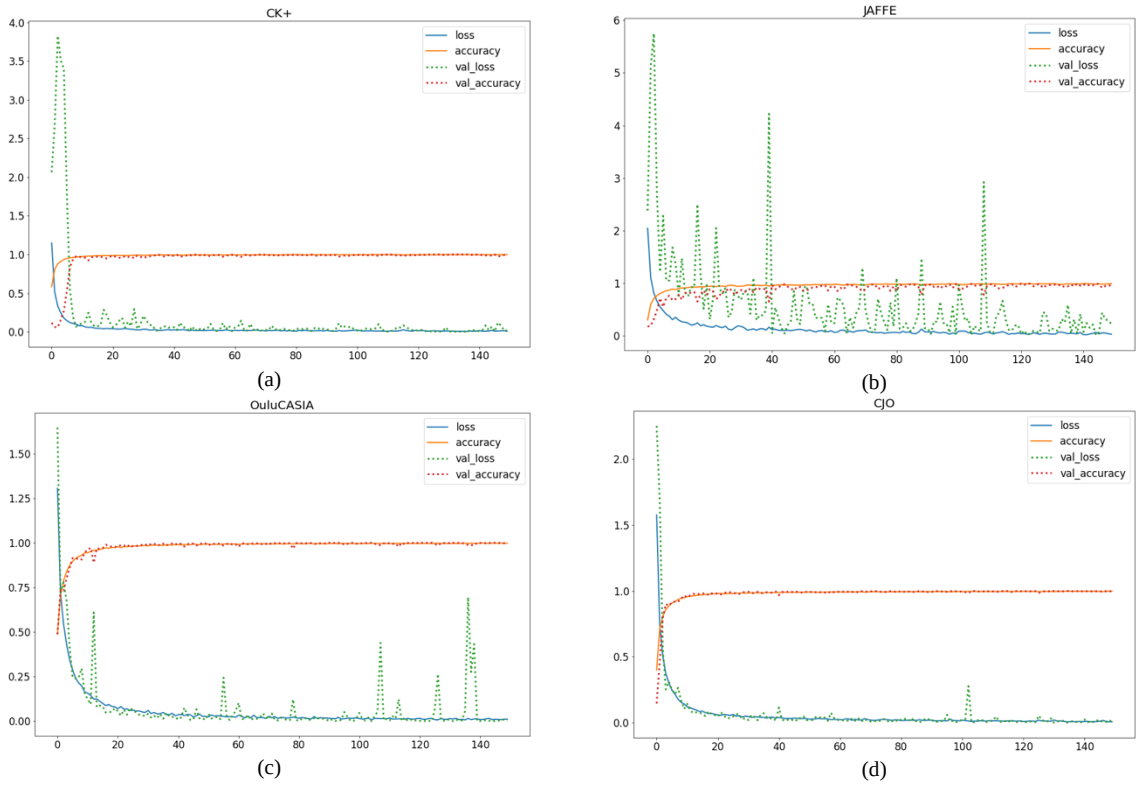

Fig. 8. Loss/accuracy of training process

In LDFER model, operations of dense-connectivity blocks (DB) plays the role of features extraction for facial expression recognition. Here, we show a visual representation of the DB operations by using gradient-based localization. This method shows the concentration or interest of convolutional neurons on local areas of an image when extracting features, it is also known as heatmap of activated neurons layer on the image. Fig.9 shows heatmap of the final convolutional layer of LDFER model on some images with every facial expressions in CJO dataset. Images all have heatmaps mostly focusing on areas that are important to represent facial expressions such as the mouth and eye. These areas are color highlighted on images. This intuitively shows that LDFER model focuses on important image regions to extract descriptive features for facial expressions and vice versa, when these image areas are not taken into account, it is difficult to identify which a proper facial expression.

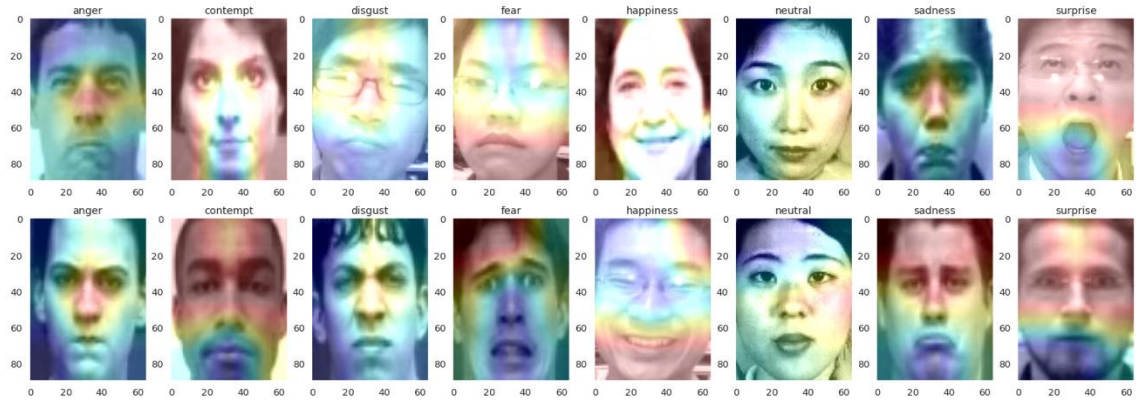

Fig. 9. Heatmaps of LDFER model on images of mixed dataset

To illustrate overall results of the LDFER model, we construct a confusion matrix from 5 runs on a mixed dataset known as CJO in Fig.10. Each row in the matrix is a label of facial expressions in the dataset, each column corresponds to a facial expression label which is predicted by the model. For a run of training model, we apply the model to recognize all entire images of the dataset (i.e., it includes training data D^{tr}, validation data D^{va} and testing data D^{te}) and it obtains total results of recognition. Each image in the dataset is applied 5 times corresponding to 5 runs of the training model, so, the sum of a row of the matrix is exactly equal to the number of images corresponding to the facial expression label in the dataset multiplied by 5. This matrix contains the total number of results on the entire CJO dataset. As in the confusion matrix, there are two labels of "contempt" and "neutral" which have no cases of misrecognition in

both predicting and being predicted by the model. It provides that these facial expressions have distinguished well from others.

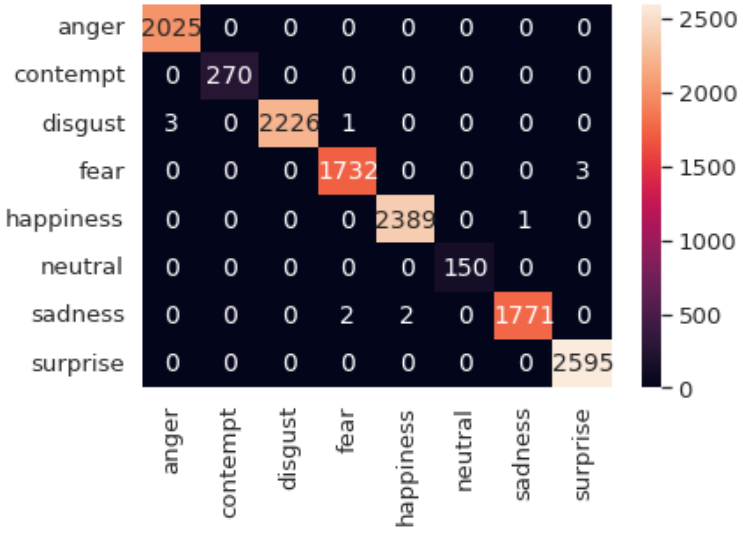

Fig. 10. Confusion matrix of LDFER model on CJO dataset

There are four labels of “contempt”, “disgust”, “neutral” and “surprise” which are correctly recognized for all images. It means that no image of other classes is recognized into these labels. The “disgust” and “sadness” have highest confusion with 4 images, they are recognized into “anger”, “fear” or “happiness”. There are three “fear” images being predicted into “disgust”, one “happiness” image is predicted into “sadness”. There are a total of 12 confused images as shown in Fig.11, accounting for 0.09% of the total 13170 recognized images. Title of each image is a couple of labels with the “>” symbol inside, the left is the target label and the right is the predicted label. Through this confusion matrix analysis, the facial expression labels of “fear”, “sadness” and “disgust” have a higher degree of confusion among them than the others. In fact, some cases are difficult to distinguish intuitively, i.e., the last image of the first row in Fig.11.

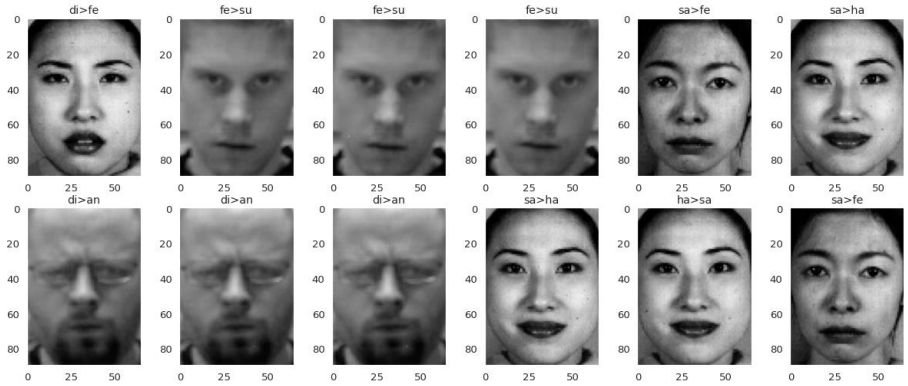

Fig. 11. Images of confused recognition

For trained LDFER model of each run, accuracy of recognition on testing data (D^{te}) is calculated as shown in Table 4. The end line of each dataset is the mean and standard deviation (in bold) of 5 runs. We also run experiments of the Efficient model in [13] with the same scenario and parameters, this model is base version (B0) for having the smallest size. The result is shown in the last column of Table 4. Although the Efficient model has above 4 million parameters which is 70% more than the proposed LDFER model, the recognition accuracy of our proposed model is slightly lower (0.1% in CK+ and 0.08% in CJO) than the Efficient model. For JAFFE and OuluCASIA datasets, the LDFER model is significantly higher (2.18% in JAFFE and 0.14% in OuluCASIA) than the Efficient model. Our proposed model has 100% accuracy at all runs of OuluCASIA, this case is reached in two datasets of CK+ and CJO by Efficient model.

Table 4. Accuracy of recognition on testing data (D^{te})

Dataset/Run	**Proposed LDFER**	**Efficient** [13]
1	100	100
2	100	100
3	100	100
4	99.49	100
5	100	100
CK+	**99.90 (±0.002)**	**100**
1	100	97.62
2	100	97.62
3	97.62	95.24
4	100	97.62
5	97.78	100
JAFFE	**99.08 (±0.009)**	**97.62 (±1.683)**
1	100	100
2	100	100
3	100	100
4	100	99.31
5	100	100
OuluCASIA	**100**	**99.86 (±0.003)**
1	99.81	100
2	100	100
3	99.81	100
4	100	100
5	100	100
Mixed dataset (CJO)	**99.92 (±0.001)**	**100**

Comparison of our results with others is in Table 5. The symbol * denotes how division of datasets for testing in experimental running, where, 5F or 10F is 5-folds or 10-folds respectively in the cross-validation scenario, 0.2T means 20% of data samples are used for testing. It shows that LDFER model has the highest accuracy in two datasets, JAFFE and OuluCASIA, while in CK+ it has the second highest of accuracy. Results in [24] and [25] are taken as the best case of CNN models because these papers are survey studies. Although the proposed model has 2.4 million parameters which 40% lower than the Efficient, it has one more dataset of highest accuracy than the Efficient.

Table 5. Comparison of results

Model (#conv/params) \ **Dataset**		**CK+**	**JAFFE**	**OuluCASIA**
The best case in [24]*10F	(22/6.8)	98.62	98.90	88.92
The best case in [25]*10F	(22/6.8)	99.60	95.80	91.67
Attentional CNN [26]*0.2T	(6/-)	98.00	92.80	-
Dynamic MTL [17]*10F	(20/13)	99.50	-	89.60
Deep MTL [27]*10F	(35/-)	97.85	-	89.23
Dilated ResNet [18]*5F	(55/1.6)	84.27	80.09	-
Octave CNN [11]*5F	(23/2.5)	97.30	-	-
Efficient [13]*5F	(78/4.1)	**100**	97.62	99.86
Proposed LDFER*5F	(34/2.4)	99.90	**99.08**	**100**

4. Conclusion

In this paper, we have proposed a convolutional neural network model for facial expression recognition problems. Structure of the model is based on dense-connectivity architecture. It has a moderate depth, i.e., the number of convolutional layers is not too large and a small number of parameters especially, which is called a lightweight model (LDFER). The recognition result is significant on testing data, it has the lowest accuracy at 99.08% of JAFFE dataset and the highest accuracy at 100% of OuluCASIA dataset. In comparison, LDFER model has the highest accuracy in JAFFE and OuluCASIA datasets, and the second highest accuracy in CK+ dataset. With the same running scenario and parameters, the Efficient model has the highest accuracy in only CK+, but it has a larger number of model parameters than the LDFER model. Thus, our proposed model has significant results in applications. In particular, it is a lightweight model, so it is easy to integrate on real applied systems with limitation of computational resources and suitable for a variety of practical conditions, it still gives good results for applied problems.

We have also designed a system to integrate the LDFER model into an online learning management system (LMS) to support recording and assessment of online learning activities. Accordingly, each learner is recorded in details of the

whole learning process on LMS, then they can be measured and evaluated for the learning process, if there are abnormalities, the system can report to teachers and administrators, of course, this supports to remind and help learners achieve higher and higher learning results. This integrated system follows an open connected mechanism, so it operates quite independently and is designed to ensure safety and security of transforming data between LMS and LDFER model.

In future studies, we will improve the model by using hybrid integration between state-of-the-art architectures to achieve higher quality of feature extraction for recognition and run experiments on more complex datasets for evaluation.

Acknowledgment

The authors wish to thank our colleagues at Hanoi Open University for assistance in research. This work was supported in part by a grant from Hanoi Open University.

References

[1] D.T.Long, "*A Lightweight Face Recognition Model Using Convolutional Neural Network for Monitoring Students in E-Learning*," I.J. Modern Education and Computer Science, vol. 6, pp. 16-28, 2020.

[2] D.T.Long, "*A Facial Expressions Recognition Method Using Residual Network Architecture for Online Learning Evaluation*," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 25, no. 6, pp. 1-10, 2021.

[3] L.Yong, Z.Jiabei, S.Shiguang and C.Xilin, "*Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism*," IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2439-2450, 2019.

[4] L.Shan and W.Deng, "*Deep Facial Expression Recognition: A Survey,"* IEEE Transactions on Affective Computing, Vols. 1949-3045, pp. 1-20, 2020.

[5] E.Derman and A.A.Salah, "*Continuous Real-Time Vehicle Driver Authentication Using Convolutional Neural Network Based Face Recognition*," 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018.

[6] S. Li and W. Deng, "*Deep Facial Expression Recognition: A Survey*," IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 2022.

[7] L.Patrick, C.F.Jeffrey, K.Takeo, S.Jason and A.Zara, "*The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression*," IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, Vols. eISSN: 2160-7516, pp. 1-8, 2010.

[8] M.Longbiao, Y.Yan, X.Jing-Hao and W.Hanzi, "*Deep Multi-task Multi-label CNN for Effective Facial Attribute Classification*," IEEE Transactions on Affective Computing, no. DOI 10.1109/TAFFC.2020.2969189, pp. 1-11, 2020.

[9] M.A.Shakik, D.Issam and D.Elio, "*Facial expression recognition using three-stage support vector machines*," Visual Computing for Industry, Biomedicine, and Art, vol. 2, no. 24, pp. https://doi.org/10.1186/s42492-019-0034-5, 2019.

[10] M.Wang and W.Deng, "*Deep Face Recognition: A Survey*," Neurocomputing, vol. 429, pp. 215-244, 2021.

[11] S.-C. Lai, C.-Y. Chen and J.-H. Li, "*Efficient Recognition of Facial Expression with Lightweight Octave Convolutional Neural Network*," Journal of Imaging Science and Technology, pp. 040402.1-9, 2022.

[12] A. Greco, N. Strisciuglio, M. Vento and V. Vigilante, "*Benchmarking deep networks for facial emotion recognition in the wild*," Multimedia Tools and Applications, pp. https://doi.org/10.1007/s11042-022-12790-7, 2022.

[13] M. Tan and Q. V. Le, "*EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*," Proceedings of the 36th International Conference on Machine Learning, pp. 6105-6114, 2019.

[14] M.Z.Alom, T.M.Taha and C.Yakopcic, "*A State-of-the-Art Survey on Deep Learning Theory and Architectures*," Electronics, vol. 8, no. 292, pp. 1-67, 2019.

[15] M.Sandler, A.Howard, M.Zhu, A.Zhmoginov and L.C.Chen, "*MobileNetV2: Inverted Residuals and Linear Bottlenecks*," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4510-4520, 2018.

[16] G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger, "*Densely Connected Convolutional Networks*," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), no. ISSN:1063-6919, pp. 1-9, 2018.

[17] G. Zhao, H. Yang and M. Yu, "*Expression Recognition Method Based on a Lightweight Convolutional Neural Network*," IEEE Access, vol. 18, pp. 38528 - 38537, 2020.

[18] R. R. Devaram and A. Cesta, "*LEMON: A Lightweight Facial Emotion Recognition System for Assistive Robotics Based on Dilated Residual Convolutional Neural Networks*," Sensors, vol. 22, no. 3366, pp. 1-20, 2022.

[19] N. Zhou, R. Liang and W. Shi, "*A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection*," IEEE Access, vol. 9, pp. 5573 - 5584, 2020.

[20] P. N. R. Bodavarapu and P. Srinivas, "*An Optimized Neural Network Model for Facial Expression Recognition over Traditional Deep Neural Networks*," International Journal of Advanced Computer Science and Applications, vol. 12, no. 7, pp. 443-451, 2021.

[21] Y. Nan, J. Ju, Q. Hua, H. Zhang and B. Wang, "*A-MobileNet: An approach of facial expression recognition*," Alexandria Engineering Journal, vol. 61, p. 4435–4444, 2022.

[22] P.Simone, F.Alessandro and A.Luigi, "*Evaluation of Data Augmentation Techniques for Facial Expression Recognition Systems*," Electronics, vol. 9, no. 1892, doi:10.3390/electronics9111892, pp. 1-12, 2020.

[23] D.P.Kingma and J.L.Ba, "*Adam: A Method For Stochastic Optimization*," CoRR, no. https://arxiv.org/abs/1412.6980, 2015.

[24] Y. Huang, F. Chen, S. Lv and X. Wang, "*Facial Expression Recognition: A Survey*," Symmetry, vol. 11, no. 1189 (doi:10.3390/sym11101189), pp. 1-28, 2019.

[25] W.Deng and S. Li, "*Deep Facial Expression Recognition: A Survey*," IEEE Transactions on Affective Computing, vol. 13, pp. 1195-1215, 2022.

[26] S. Minaee, M. Minaei and A. Abdolrashidi, "*Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network*," Sensors, vol. 21, no. 3046 (https://doi.org/10.3390/s21093046), pp. 1-16, 2021.

[27] R. Zhao, T. Liu, J. Xiao, D. P. Lun and K.-M. Lam, "*Deep Multi-task Learning for Facial Expression Recognition and Synthesis Based on Selective Feature Sharing*," 25th International Conference on Pattern Recognition (ICPR), pp. 4412-4419, 2020.

Authors' Profiles

Duong Thang Long is a lecturer of Information Technology major at Hanoi Open University. He received a PhD degree of Information Technology from Vietnam Academy of Science and Technology (VAST) in 2011. His research interests are Machine Learning, Artificial Intelligence, Deep Learning, Computer Vision, Fuzzy Logic and soft computing with real-world applications. He can be contacted at email: duongthanglong@gmail.com or duongthanglong@hou.edu.vn.

Truong Tien Tung is a lecturer of Information Technology major at Hanoi Open University. He received a PhD degree in Information Technology from Russia. His research interest is Database systems, Data structure and Algorithms, Social Learning. He can be contacted at email: truongtientung@hou.edu.vn.

Tran Tien Dung is a lecturer of Information Technology major at Hanoi Open University. He received a Master of Science degree in Information Technology from Le Quy Don university, Vietnam. His research interest is Database systems, Computer network, Infrastructure of IT. He can be contacted at email: dungtranitd@hou.edu.vn.

How to cite this paper: Duong Thang Long, Truong Tien Tung, Tran Tien Dung, "A Facial Expression Recognition Model using Lightweight Dense-Connectivity Neural Networks for Monitoring Online Learning Activities", International Journal of Modern Education and Computer Science(IJMECS), Vol.14, No.6, pp. 53-64, 2022. DOI:10.5815/ijmecs.2022.06.05