Explainable Model Selection of a Convolutional Neural Network for Driver's Facial Emotion Identification

Amany A. Kandeel, Hazem M. Abbas, and Hossam S. Hassanein

Agenda

- Introduction
- Objectives
- Driver's emotion detection approaches
- The proposed model
- Experimental results
- Datasets
- Conclusion

Introduction

Road conditions

Emotion

Weather

Fatigue

Vehicle

Drinking

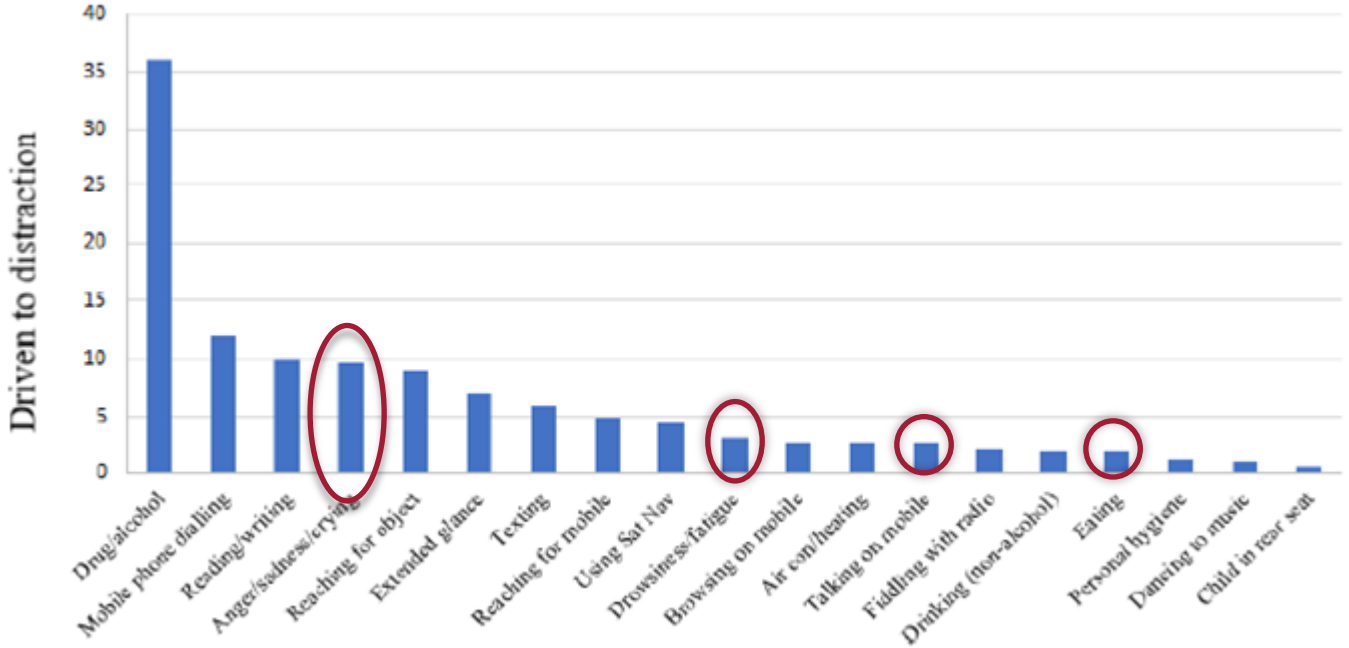
Using mobile

Introduction (cont.)

Facial expression influence on driver distraction

Virginia Tech research [1]

- driving with sadness , anger (10)
- drowsiness or fatigue (3)
- talking on mobile or eating (2)



Objectives

- Implement a Facial Expression Recognition (FER) model using a Convolutional Neural Network (CNN)
- Using explainability methods to find the best architecture

Driver emotion detection

High accuracy

Low complexity

Facial Expression Recognition (FER)

What is FER?

FER challenges

- Close in the feature space
- Different representations of emotion
- Brightness, background, occlusion
- The scarcity of facial expression datasets

Ambiguity between angry vs sad expression

Surprise emotion representation for the same person

Three different subjects with the happy expression

Driver Emotion Detection Approaches

- **Biomedical means**) ECG sensors(
- **A self-reporting questionnaire**
- **Analyzing data from sensors** (steering wheel)
- **Manual feature engineering**
  - Feature extraction
  - Classification (SVM)
- **Deep learning**
  - learn directly from the input images
  - Computer vision

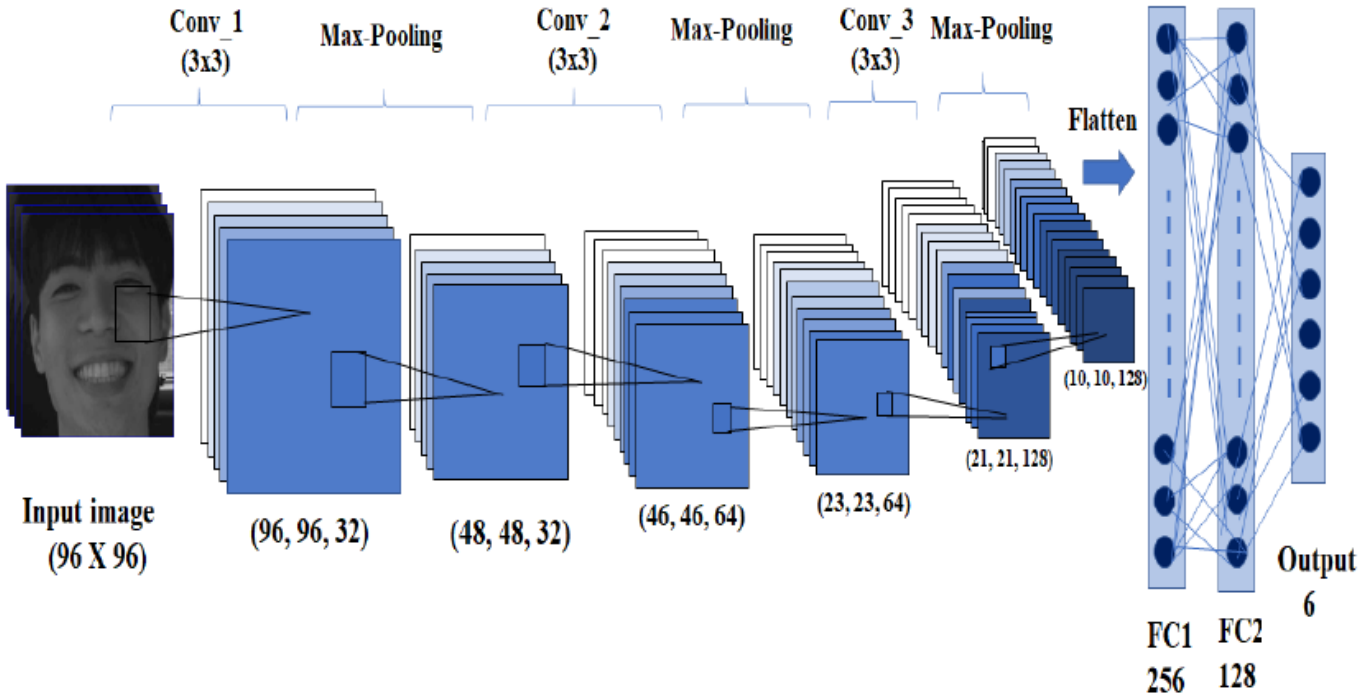
The Proposed FER Model

Face detection and cropping

- Haar Cascades classifier

CNN model

- Feature extraction
  - 3 Convolutional(32, 64, 128)
  - max Pooling
- Classification layer
  - fully connected (256, 128)
  - final output layer (6)



Datasets

JAFFE dataset

- 213 samples

CK+ dataset

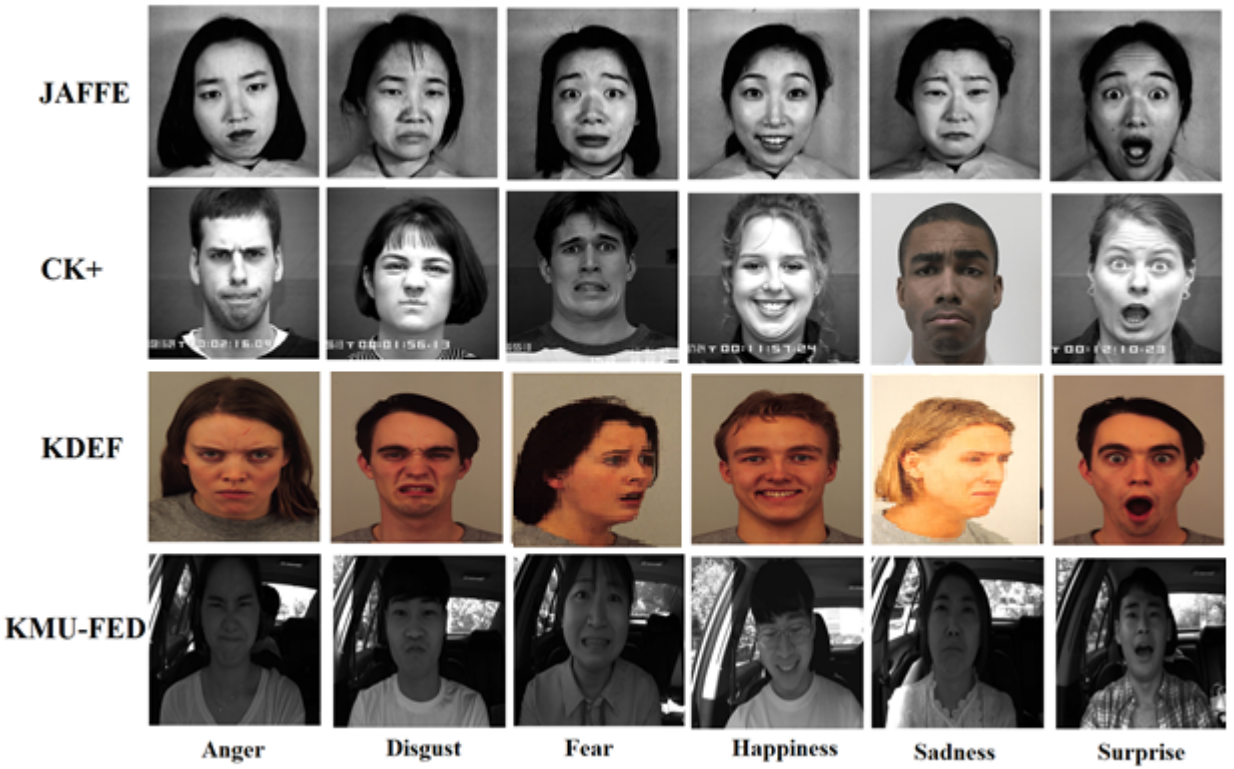
- 327 labeled sequences
- variations in illumination & background

KDEF dataset

- 4900 pictures
- frontal and non-frontal faces, five different angles
- variations in illumination

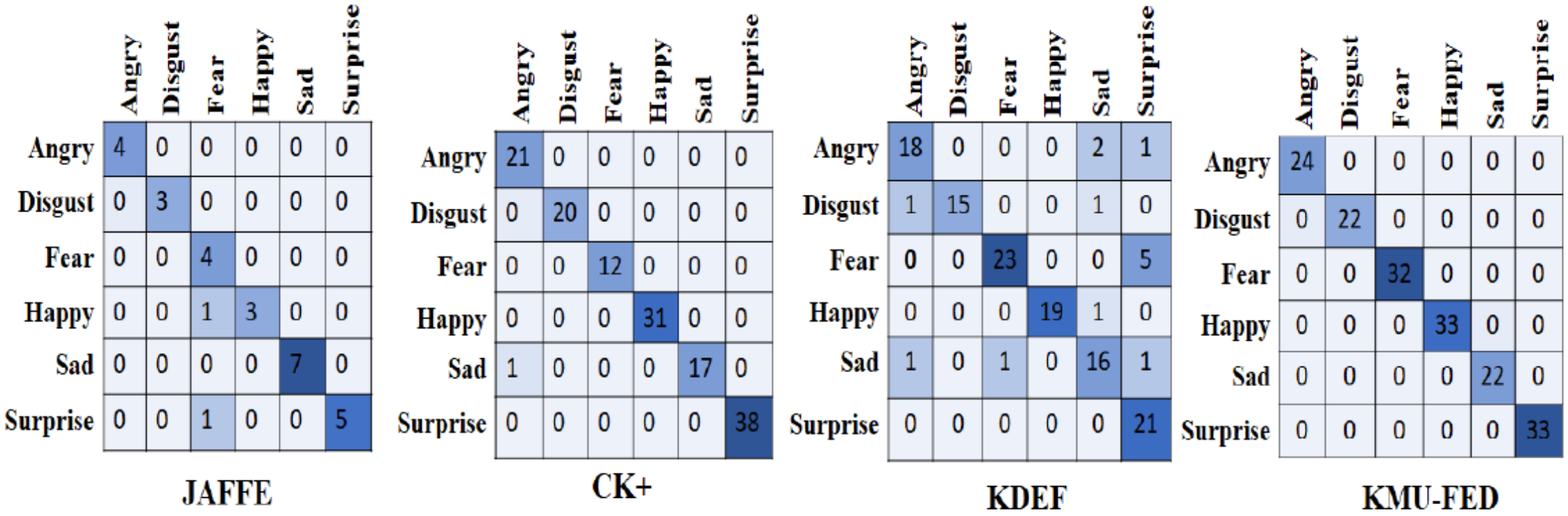
KMU-FED dataset

- real driving environment
- 55 image sequences, 1106 frames
- changes in illumination, occlusions



Datasets

Experiments and Results(cont.)



JAFFE

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	4	0	0	0	0	0
Disgust	0	3	0	0	0	0
Fear	0	0	4	0	0	0
Happy	0	0	1	3	0	0
Sad	0	0	0	0	7	0
Surprise	0	0	1	0	0	5

CK+

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	21	0	0	0	0	0
Disgust	0	20	0	0	0	0
Fear	0	0	12	0	0	0
Happy	0	0	0	31	0	0
Sad	1	0	0	0	17	0
Surprise	0	0	0	0	0	38

KDEF

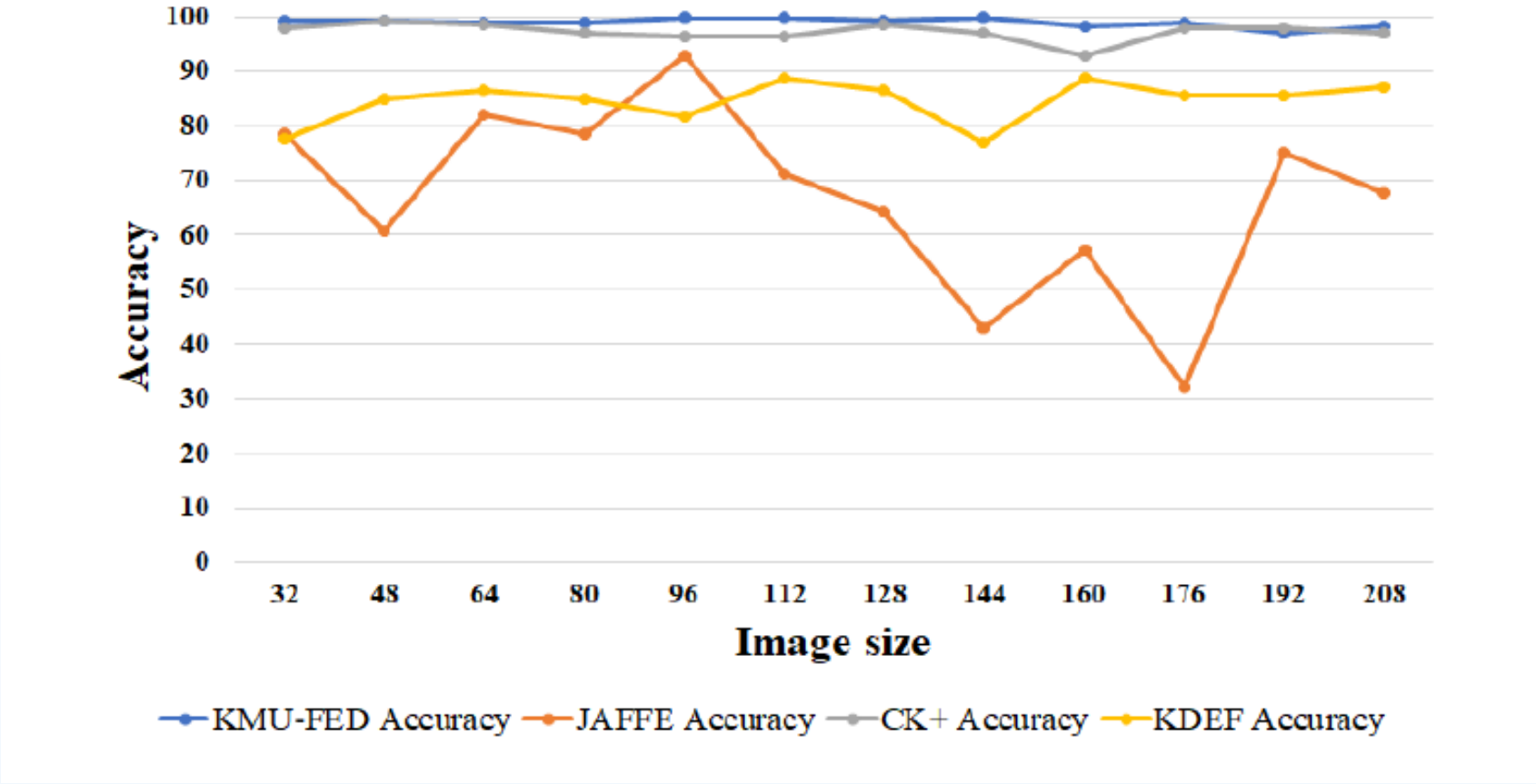
	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	18	0	0	0	2	1
Disgust	1	15	0	0	1	0
Fear	0	0	23	0	0	5
Happy	0	0	0	19	1	0
Sad	1	0	1	0	16	1
Surprise	0	0	0	0	0	21

KMU-FED

	Angry	Disgust	Fear	Happy	Sad	Surprise
Angry	24	0	0	0	0	0
Disgust	0	22	0	0	0	0
Fear	0	0	32	0	0	0
Happy	0	0	0	33	0	0
Sad	0	0	0	0	22	0
Surprise	0	0	0	0	0	33

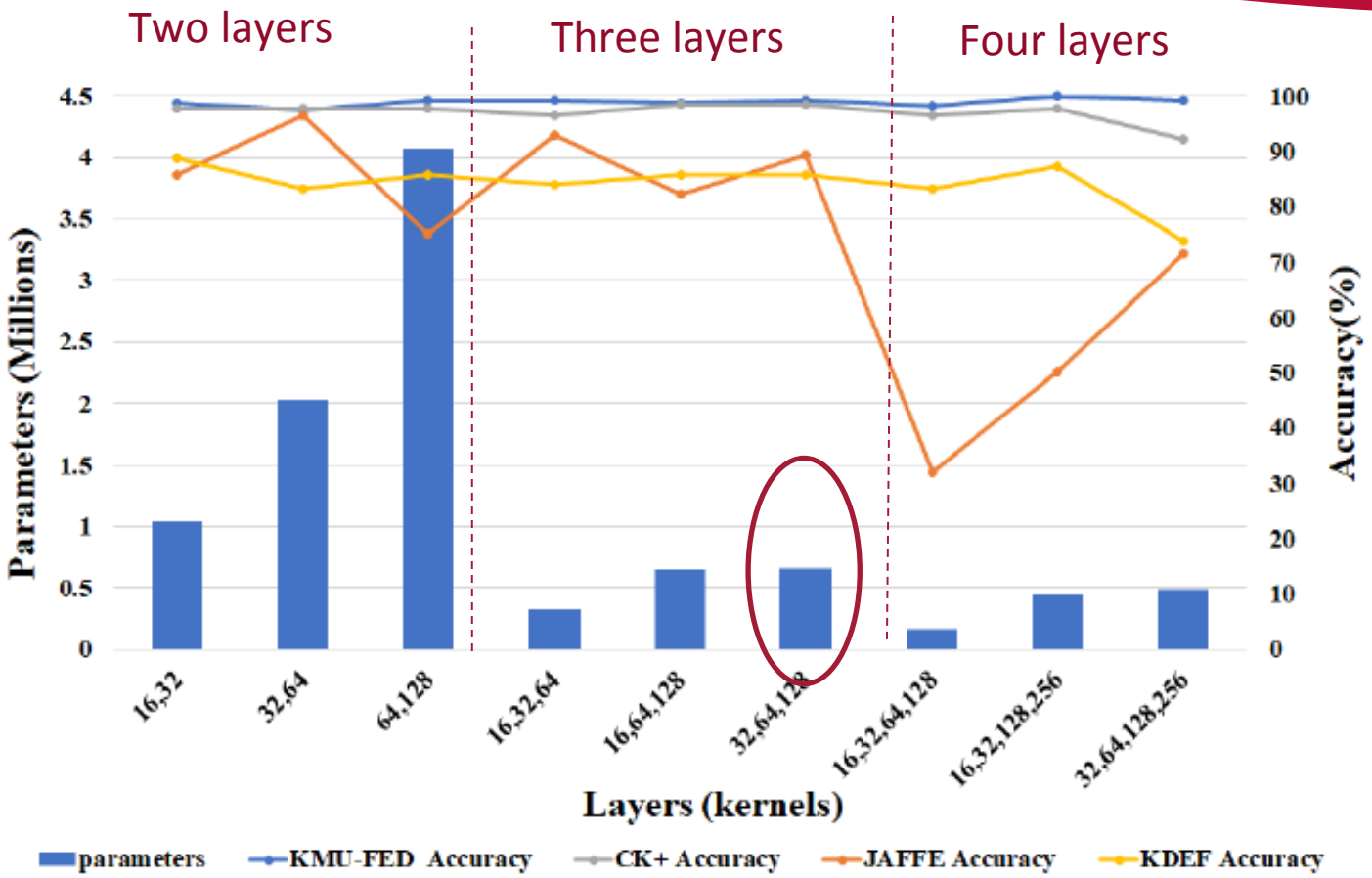
The confusion matrix of each dataset

Experiments and Results(cont.)



Dataset accuracy versus different image sizes

Experiments and Results(cont.)



Dataset accuracy versus different CNN architectures

Experiments and Results(cont.)

Explainable Model

Deep learning
- High performance
- Black box

Explainable Artificial Intelligence (XAI)
- Appropriate explanation
- improves the transparency and the trust

Saliency Map
Gradient-Weighted Class Activation Map (Grad-CAM)

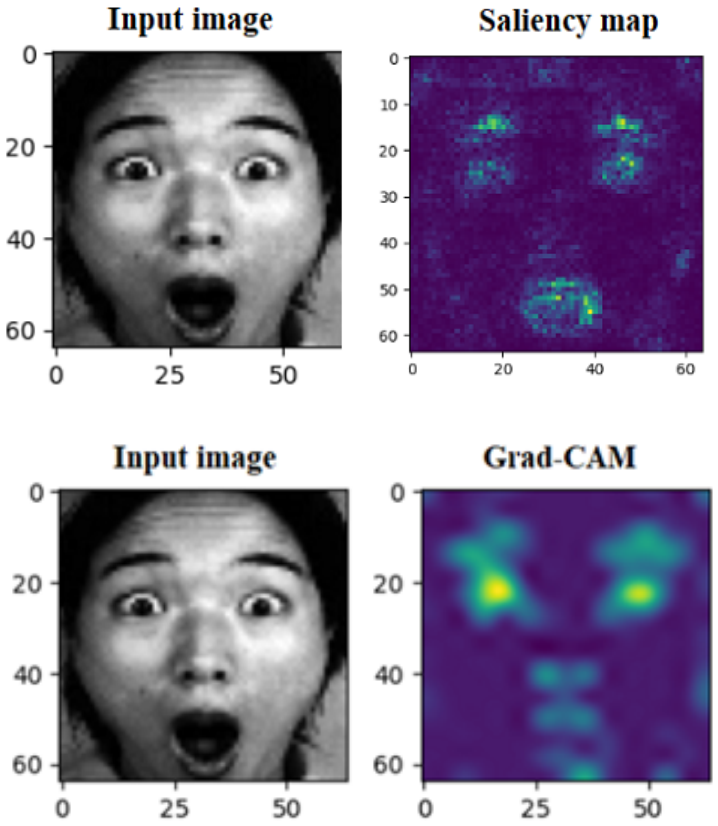
Experiments and Results(cont.)

Saliency Map [2]

- Local interpretable
- Areas importance
- Difference between a particular position and its surrounding

Grad-CAM [3]

- The gradient information
- Related regions to each class
- Class Activation Map (CAM) approach
- Guided Backpropagation



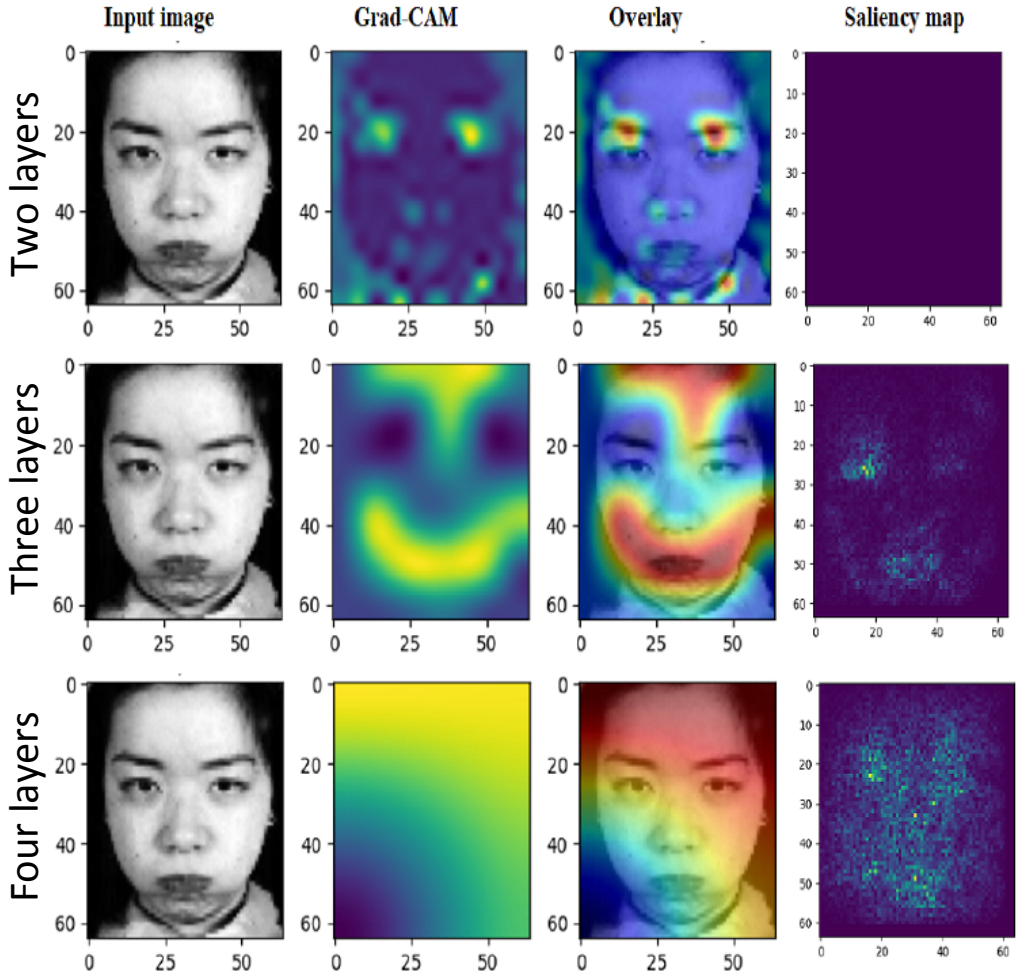
Experiments and Results(cont.)

Grad-CAM

- Two layers --> more details
- Three layers --> more accurate
- Four layers --> not clear

Saliency map

- Two layers --> no details
- Three layers --> accurate and clear
- Four layers --> ambiguous



Experiments and Results(cont.)

Comparison between our approach and the state-of-the-art FER approaches

Method	JAFFE	CK+	KDEF	KMU-FED
Jeong et al. [4]	--	92.60%	--	94.70%
Zhang et al. [5]	--	93.60%	--	97.30%
Lopes et al. [6]	53.44%	96.76%	--	--
Yang et al. [7]	92.21%	97.02%	--	--
Puthanidam et al. [8]	100%	--	89.58%	--
Pandey et al. [9]	--	--	83.33%	--
Proposed	**92.85%**	**99.28%**	**88.88%**	**100%**

Conclusion and Future Work

- A CNN model to recognize the driver's emotions
- Frontal face datasets
- Two interpretability methods

- Non-frontal samples
- Identifying the driver's distraction in real-time

References

[1] Knapton, S.: Which emotion raises the risk of a car crash by nearly 10 times? https://www.telegraph.co.uk/news/science/science-news/12168472/Whichemotion- raises-the-risk-of-a-car-crash-by-nearly-10-times.html? (Feb 2016)

[2] Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2013)

[3] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618{626 (2017)

[4] Jeong, M., Ko, B.C.: Driver's facial expression recognition in real-time for safe driving. Sensors 18(12), 4270 (2018)

[5] Zhang, J., Mei, X., Liu, H., Yuan, S., Qian, T.: Detecting negative emotional stress based on facial expression in real time. In: 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP). pp. 430{434 (2019)

[6] Lopes, A.T., de Aguiar, E., De Souza, A.F., Oliveira-Santos, T.: Facial expression recognition with convolutional neural networks: coping with few data and the training sample order. Pattern Recognition 61, 610{628 (2017)

[7] Yang, B., Cao, J., Ni, R., Zhang, Y.: Facial expression recognition using weighted mixture deep neural network based on double-channel facial images. IEEE Access 6, 4630{4640 (2017)

[8] Puthanidam, R.V., Moh, T.S.: A hybrid approach for facial expression recognition. In: Proceedings of the 12th International Conference on Ubiquitous Information Management and Communication. pp. 1{8 (2018)

[9] Pandey, R.K., Karmakar, S., Ramakrishnan, A., Saha, N.: Improving facial emotion recognition systems using gradient and laplacian images. arXiv preprint arXiv:1902.05411 (2019)

Thank you