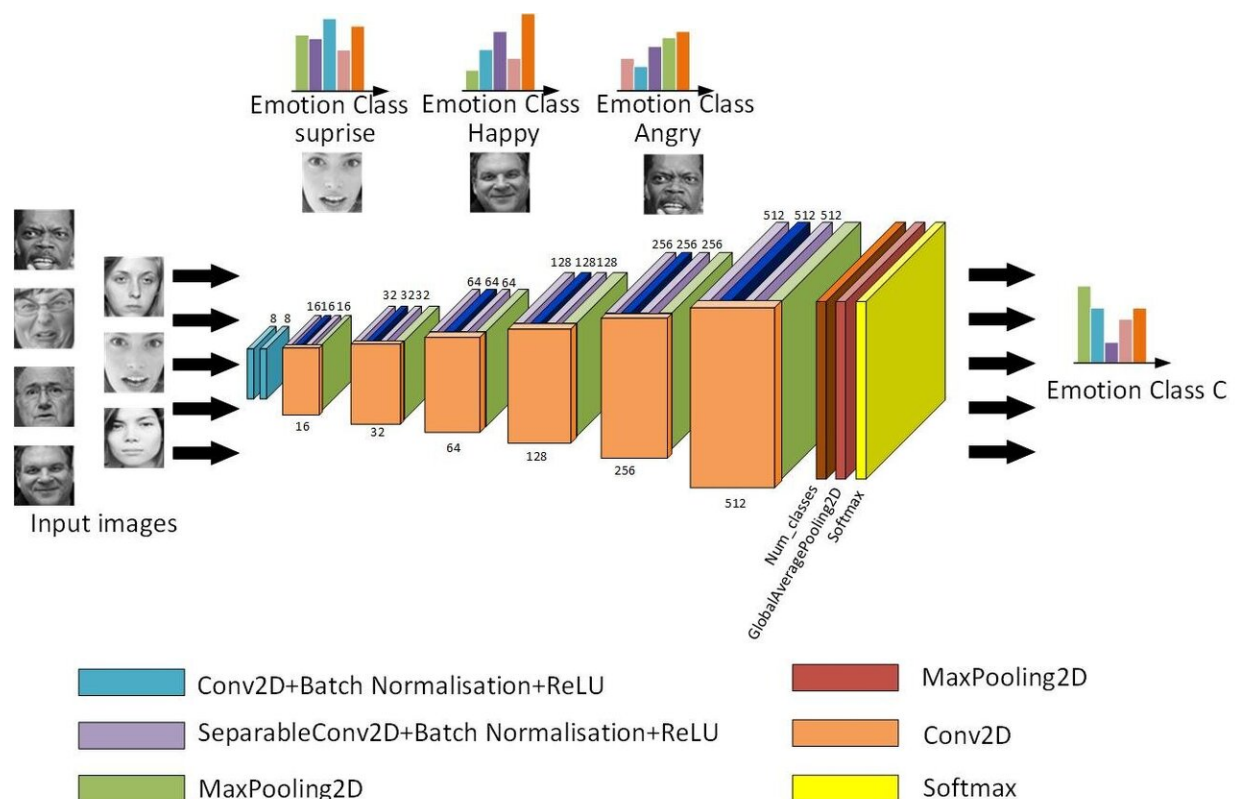# Three convolutional neural network models for facial expression recognition in the wild

May 24 2019, by Ingrid Fadelli

The basic structure of the Light-CNN. Credit: Jie & Yongsheng.

Two researchers at Shanghai University of Electric Power have recently developed and evaluated new neural network models for facial expression recognition (FER) in the wild. Their study, published in Elsevier's Neurocomputing journal, presents three models of

convolutional neural networks (CNNs): a Light-CNN, a dual-branch CNN and a pre-trained CNN.

"Due to the lack of information on non-frontal faces, FER in the wild is a difficult point in computer vision," Qian Yongsheng, one of the researchers who carried out the study, told TechXplore. "Existing natural facial expression recognition methods based on deep convolutional neural networks (CNNs) present several problems, including over-fitting, high computational complexity, single feature and limited samples."

Although many researchers have developed CNN approaches for FER, so far, very few of them have tried to determine what type of network is best suited for this particular task. Aware of this gap in the literature, Yongsheng and his colleague Shao Jie developed three different CNN for FER and carried out a series of evaluations to identify their strengths and weaknesses.

"Our first model is a shallow light-CNN that introduces a depthwise separable module with the residual network module, reducing network parameters by changing the convolution method," Yongsheng said. "The second is a dual-branch CNN, which combines global features and local texture features, trying to obtain richer features and compensate for the lack of rotation invariance of convolution. The third pre-trained CNN uses weights trained in the same distributed large database to retrain on its own small database, reducing training time and improving recognition rate."
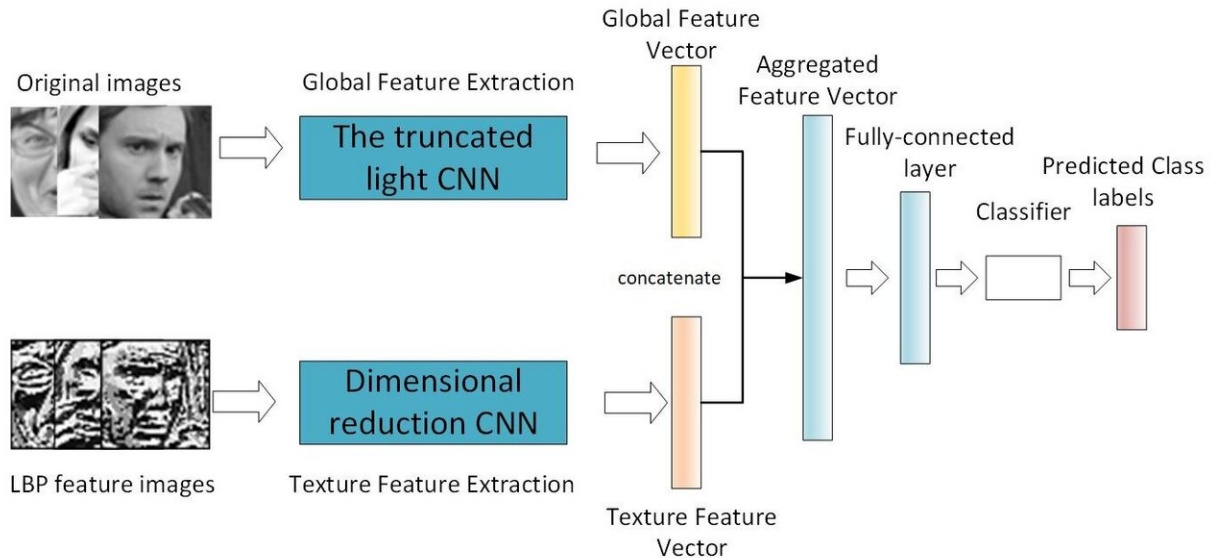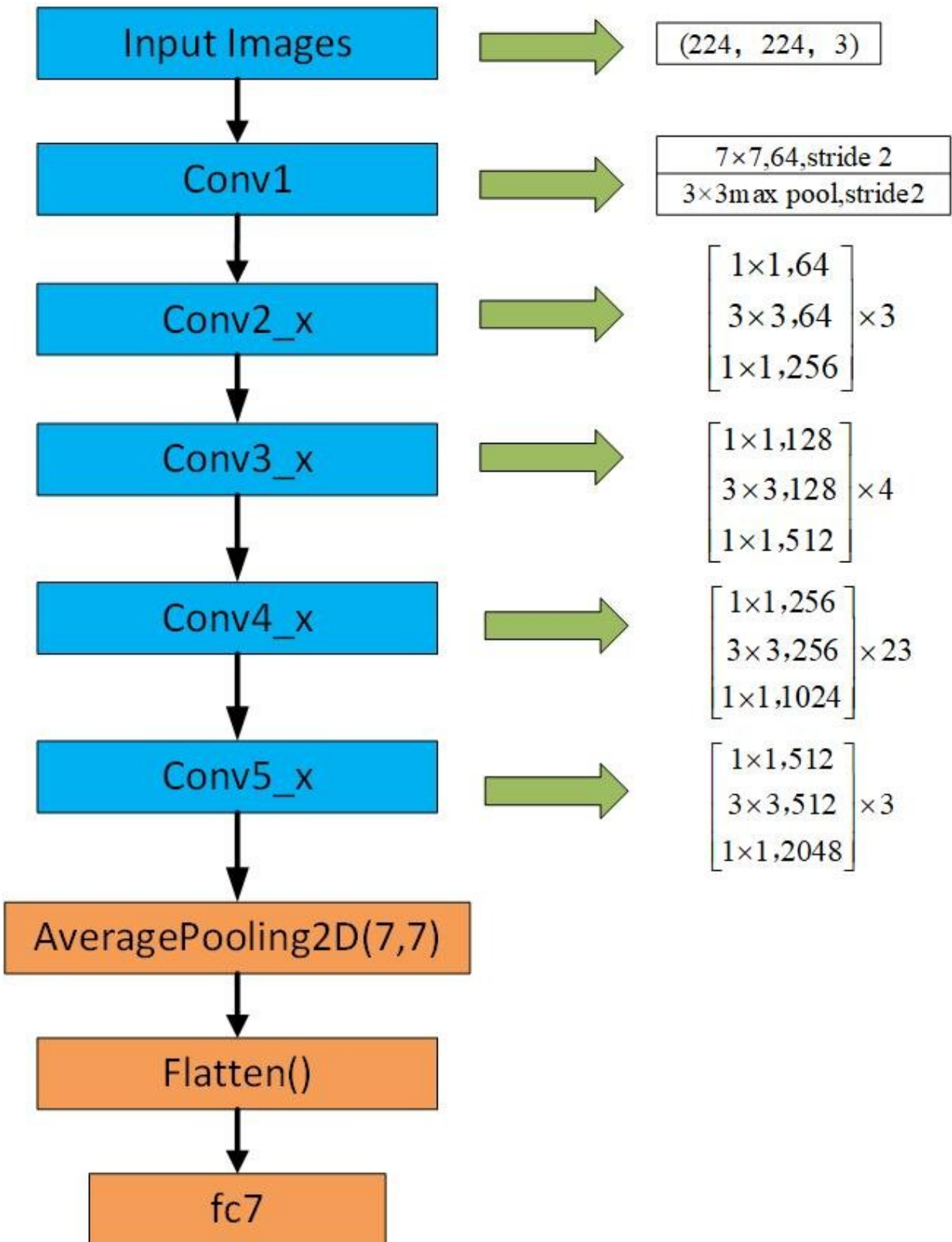
Framework of the dual-branch CNN. Credit: Jie & Yongsheng.

The researchers carried out extensive evaluations of their CNN models on three datasets that are commonly used for FER: the public CK+, multi-view BU-3DEF and FER2013 datasets. Although the three CNN models presented differences in performance, they all achieved promising results, outperforming several state-of-the-art approaches for FER.

"At present, the three CNN models are used separately," Yongsheng explained. "The shallow network is more suitable for embedded hardware. The pre-trained CNN can achieve better results, but requires pre-trained weights. The dual-branch network is not very effective. Of course, one could also try to use the three models together."

In their evaluations, the researchers observed that by combining the residual network module and the depthwise separable module, as they did for their first CNN model, network parameters could be reduced.

This could ultimately resolve some of the shortcomings of computing hardware. In addition, they found that the pre-trained CNN model transferred a large database to its own database and could hence be trained with limited samples.

Input Images	(224, 224, 3)
Conv1	7×7,64,stride 2 3×3 max pool,stride2
Conv2_x	$\begin{bmatrix} 1\times1,64 \\ 3\times3,64 \\ 1\times1,256 \end{bmatrix} \times 3$
Conv3_x	$\begin{bmatrix} 1\times1,128 \\ 3\times3,128 \\ 1\times1,512 \end{bmatrix} \times 4$
Conv4_x	$\begin{bmatrix} 1\times1,256 \\ 3\times3,256 \\ 1\times1,1024 \end{bmatrix} \times 23$
Conv5_x	$\begin{bmatrix} 1\times1,512 \\ 3\times3,512 \\ 1\times1,2048 \end{bmatrix} \times 3$
AveragePooling2D(7,7)	
Flatten()	
fc7	

The framework of the pretrained CNN. Credit: Jie & Yongsheng.

The three CNNs for FER proposed by Yongsheng and Jie could have numerous applications, for instance, aiding the development of robots that can identify the facial expressions of humans they are interacting with. The researchers are now planning to make additional adjustments to their models, in order to further enhance their performance.

"In our future work, we will try to add different traditional manual features to join the dual-branch CNN and change the fusion mode," Yongsheng said. "We will also use cross-database training network parameters to get better generalization capabilities and adopt a more effective deep transfer learning approach."