DEVELOPING A DEEP LEARNING-BASED AFFECT RECOGNITION SYSTEM FOR YOUNG CHILDREN

Amir Hossein Farzaneh, Yanghee Kim, Mengxi Zhou, and Xiaojun Qi

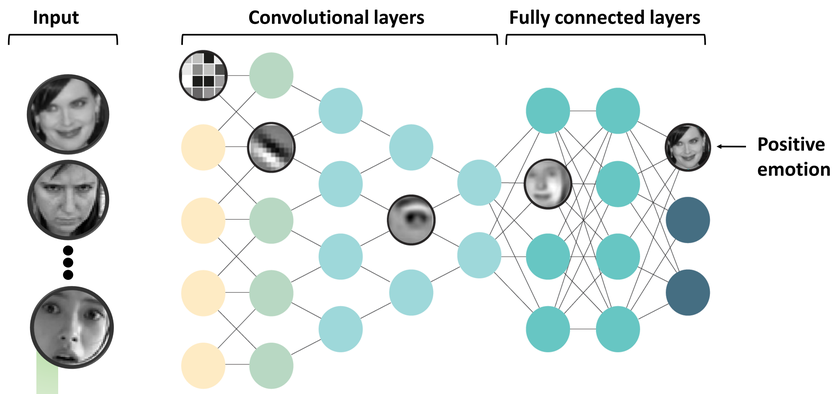
Poster PDF

AFFECT RECOGNITION

Children's interactions with digital devices are rich in emotions and involve a lot of non-verbal responses. Designing and implementing a system that recognizes children's learning-related affective state using a non-verbal channel such as facial expression, is needed to expedite the analysis of children's behaviors in a digital learning environment. In this study, we have leveraged deep learning techniques to predict the emotions of young children.

DEEP LEARNING

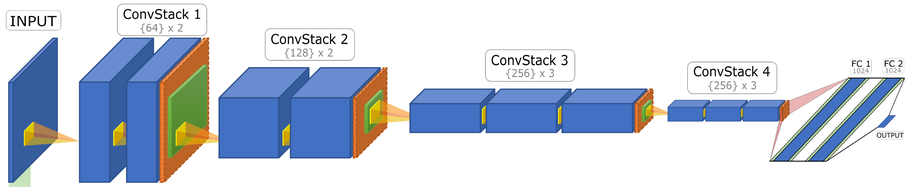
Deep learning algorithms take many forms. Our research group uses a Deep Convolutional Neural Network (DCNN) to automatically extract sophisticated facial features from images and integrated them in fully connected layers that follow. Over multiple iterations, the AI network discovers patterns in the data that can distinguish emotions.



DCNN as a trainable AI

- To train our AI, we use an enhanced Facial Expression Recognition (FER+) dataset with almost 36,000 facial images captured under diverse illuminations, head poses, occlusions, and have a broad spectrum of demographics including people of different ages and races.
- For testing, we use three video sequences which include children interacting with a teaching assistant. In these trial tests, we consider two affective categories: *positive* and *neutral*.

NETWORK TOPOLOGY AND TRAINING



VGG-12 architecture | blue: activation, yellow: convolution, orange: max-pooling, green: dropout

Our VGG-12 network achieves close to state-of-the-art performance while offering a simple architecture. The *output* layer produces the probability of each candidate's emotion being *positive* or *neutral*. During training. FER+ images are augmented with random affine transformations to provide better generalization.

TESTING

To test our system, we first track each child's face in three videos using a CNN-based Multi-Domain Convolutional Neural Network (MDNet) tracker and pass the pre-processed face to the trained CNN. We save the predicted emotion for each child in all frames for evaluation.

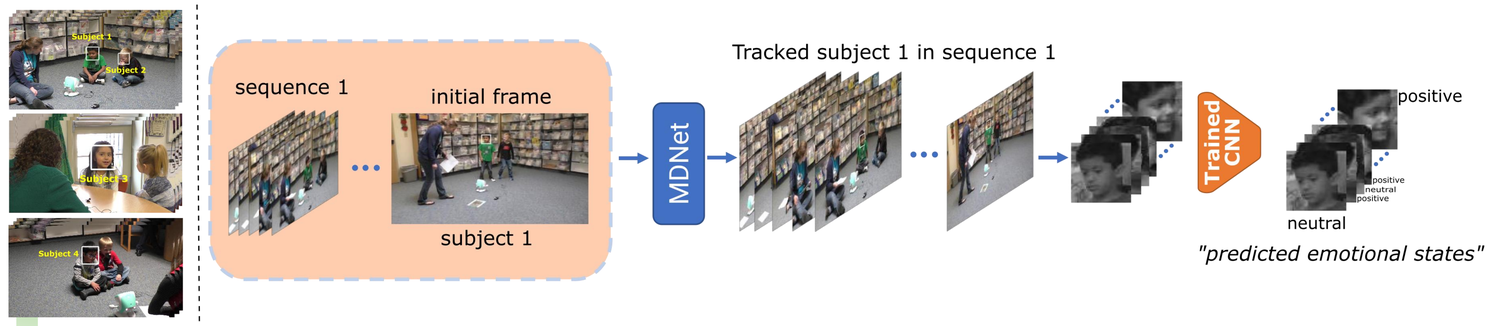


Illustration of the proposed affect recognition system

EVALUATION

We calculate the accuracy of the proposed system as the ratio of *the number of correct predictions* and *the total number of predictions.*

	Trials					
Target Subject	1	2	3	4	5	Average Accuracy
Subject 1|Sequence 1	89.01 %	96.70 %	87.91 %	96.70 %	94.51 %	92.97 %
Subject 2|Sequence 1	98.04 %	96.08 %	96.08 %	100.00 %	98.04 %	97.65 %
Subject 3|Sequence 2	96.96 %	98.48 %	98.18 %	98.78 %	98.18 %	**98.12 %**
Subject 4|Sequence 3	82.69 %	84.62 %	82.69 %	82.69 %	84.62 %	83.46 %
Average Accuracy	91.67 %	93.97 %	91.21 %	**94.54 %**	93.84 %	

Affective state testing accuracy for 4 children in 3 test video sequences

farzaneh@aggiemail.usu.edu z1841378@students.niu.edu

The 20th International Conference on Artificial Intelligence in Education (AIED-2019)