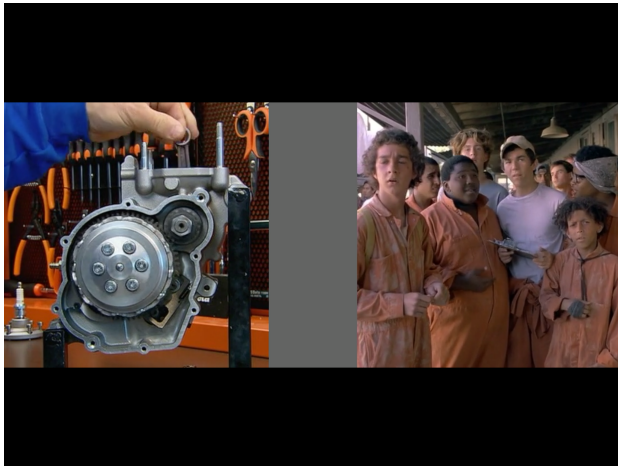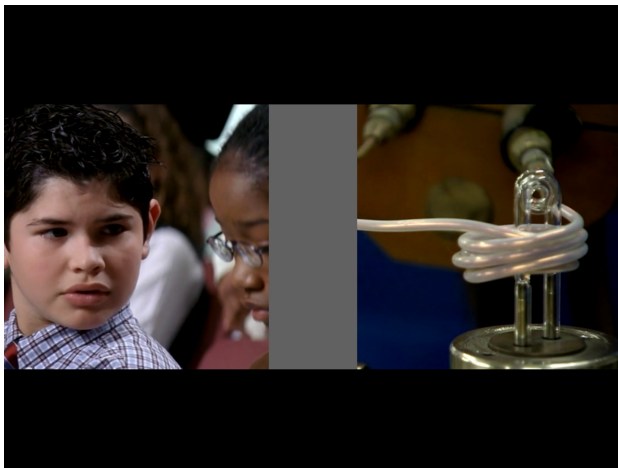# Supplement

(a)

(b)

Fig. 1: **Screenshots from the iPad Experiment (Stimulus):** In our experiment setup, we show two videos simultaneously on the iPad, where one is dynamic naturalistic video, and another one is social communication video. The dynamic naturalistic video contains machines, fractals, toys, cars, nature scenes, and so on; the social communication video contains people interacting with each other. These two types of videos swap left and right periodically as shown in (a) and (b).
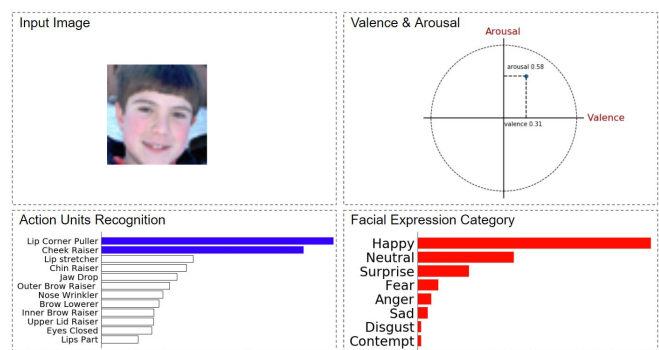

Fig. 2: **CNN Input and Outputs:** The Convolutional Neural Network (CNN) in our system takes an image as input, detects the face in the image, and then analyzes the expression, action units, arousal, and valence in the image. There are eight categories for expression, which are neutral, happy, sad, disgust, anger, surprise, fear, contempt. Our system also recognizes 12 action units, which are: AU 1 (inner brow raiser), AU 2 (outer brow raiser) , AU 4 (brow lowerer), AU 5 upper lid raiser, AU 6 (cheek raiser), AU 9 (nose wrinkler), AU 12 (lip corner puller), AU 17 (chin raiser), AU 20 (lip stretcher), AU 25 (lips part), AU 26 (jaw drop), and AU 43 (eyes closed).