Project proposal

Problem:

Facial expression recognition using attentional convolutional network

Given a face image, it is clear that not all parts of the face are important in detecting a specific emotion, and in many cases, we only need to attend to the specific regions to get a sense of the underlying emotion. To tackle this, we can add an attention mechanism through spatial transformer network into traditional covolutional network.

Solution:

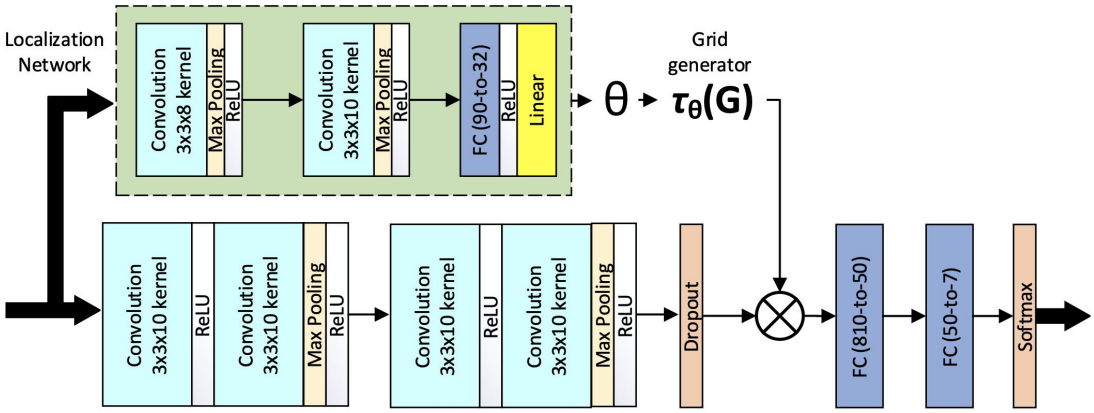

Fig. 3: The proposed model architecture

Part 1

The feature extraction part consists of four convolutional layers, each two followed by max-pooling layer and ReLU activation function. They are then followed by a dropout layer and two fully-connected layers.

Part 2

The spatial transformer consists of two convolutional layers (each foolowed by max-pooling and ReLU), and two fully-connected layers. Use affine transformation to warp the input to the output.

Datasets:

We are going to use the JAFFE and CK+ to train and test our model.

JAFFE contains 213 images of the 7 facial expressions posed by 10 Japanese female models.

CK+ contains a total of 593 sequences across 123 subjects. We will use the last frame of these sequences for our experiment.
If time permit, we will also test the model using FERG dataset which contains 55,767 annotated face images.

Team member and responsibilities:

Yin Jiang A20250197
Image pre-processing and model implementation part 1.
Ran Huo A20437439
Model implementation part 2 and model output.