Facial Expression Recognition Using Attentional Convolutional Network

Ran Huo
Yin Jiang

Abstract-Facial expression recognition has been an active research area over the past few decades, and it is still challenging due to the high intra-class variation. Traditional approaches for this problem rely on hand-crafted features such as SIFT, HOG and LBP, followed by a classifier trained on a database of datasets of images captured in a controlled condition, but fail to perform as good on more challenging datasets with more image variation and partial faces. In this project we implemented deep learning approach based on attentional convolution network, which is able to focus on important parts of the face for facial expression recognition. The model is trained using the FER dataset and achieves significant improvement over traditional approaches. [1]

Introduction

Facial expression is one of the most popular features in emotion recognition due to a number of reasons. They are visible; contain many useful features and are easy to collect a large dataset.
With the use of deep learning, many features can be extracted and learned for a decent facial expression recognition model. However, features extracted as clues of facial expressions come from a few parts of the face, e.g. the mouth and eyes. Other parts, such as ears and hair, play little part in the recognition. Thus, attentional convolution network fits the current need as a model focuses only on important parts of face in facial expression recognition. [1]

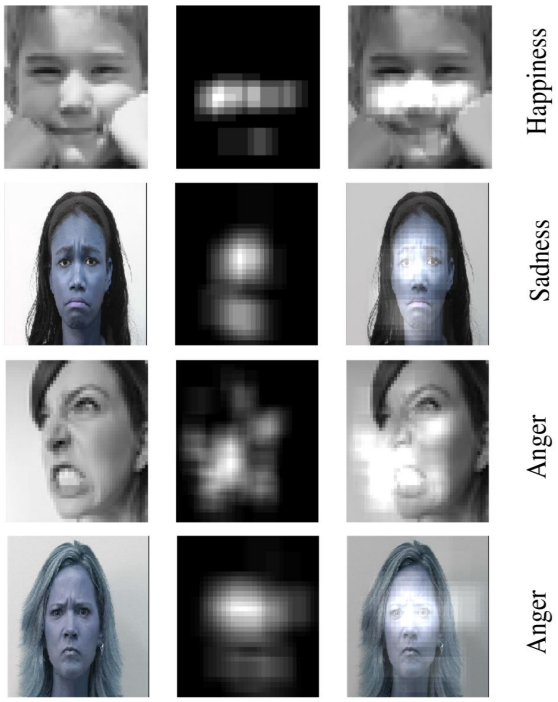
Related Works

Earlier works on emotion recognition, rely on the traditional two-step machine learning approach, where in the first step some features are extracted from the images, and in the second step, a classifier (such as SVM, neural network, or random forest) are used to detect the emotions. Some of the popular hand-crafted features used for facial expression recognition include the histogram of oriented gradients (HOG), local binary patterns (LBP) and Haar features. A classifier would then assign the best emotion to the image. These approaches seemed to work fine on simpler datasets, but with the advent of more challenging datasets, they started to show

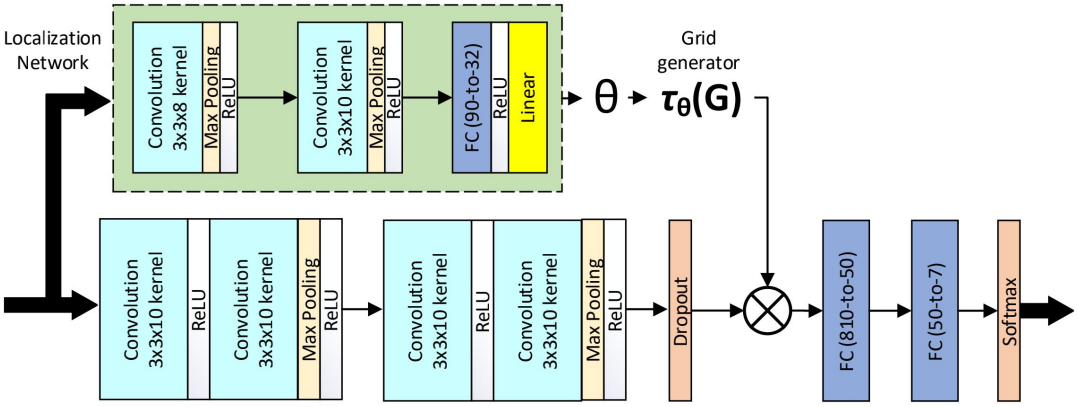
their limitation.
With the great success of deep learning, and more specifically convolutional neural networks, several groups developed deep learning-based models for facial expression recognition, which achieve significant improvements over the traditional works on emotion recognition.

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. Adding spatial transformer network to convulutional network enables us to focus on important face regions (Figure 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. The spatial transformer (the localization network) consists of two convolution layers (each followed by max-pooling and ReLU), and two fully-connected layers. After regressing the transformation parameters, the input is transformed to the sampling grid $T(\theta)$ producing the warped data. The spatial transofrmer module essentially tries to focus on the most relevant part of the image, by estimating a sample over the attended region. [3] An affine transformation is used to warp the input to the output (Figure 2).



The model is trained by optimizing a loss function (cross-entropy) using stochastic gradient descent approach (Adam).

```
criterion= nn.CrossEntropyLoss()
optmizer= optim.Adam(net.parameters(),lr= lr)
Train(epochs, train_loader, val_loader, criterion, optmizer, device)
```

Database

We use FER dataset to train the models.
FER: The facial expression recognition dataset contains 28,709 examples, 48x48 pixel grayscale images of faces. Facial expression categories: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. [2]

Implementation

Since the dataset is saved as a csv file, we need to load the csv file and split train and validation data sets.
The first column contains the image pixel data and the second contains label. The data loader saves each image pixel data into a picture.
A deep_emotion class object is created to construct the attentional convolutional network including convolutional network and spatial transformer network.
By setting up the cross-entropy loss function and Adam optimizer with L2 regularization, the model is trained using the train and validation data sets.
Model accuracy and loss is reported.

Result comparing

The paper provided comparison table between different works and proposed algorithm.

TABLE I: Classification Accuracies on FER 2013 dataset

Method	Accuracy Rate
Bag of Words [52]	67.4%
VGG+SVM [53]	66.31%
GoogleNet [54]	65.2%
Mollahosseini et al [19]	66.4%
The proposed algorithm	70.02%

TABLE II: Classification Accuracy on FERG dataset

Method	Accuracy Rate
DeepExpr [2]	89.02%
Ensemble Multi-feature [49]	97%
Adversarial NN [48]	98.2%
The proposed algorithm	99.3%

As we can see through the table, the proposed algorithm outperforms among other deep learning models.
Because of the availability of databases, we only have one dataset (FER). Our model was trained using the FER dataset and achieved 61.385% accuracy comparing with the 70.02% accuracy from the paper.

```
Training Loss: 0.00808909     Validation Loss 0.01246600    Training Acuuarcy 61.385%     Validation Acuuarcy 44.920%
----------------------Training Finished----------------------------------
```

Program Configuration

The program has been uploaded to the bitbucket. The data folder is empty because the dataset is too large. To download the data, go to https://drive.google.com/drive/folders/1ZlWVNjsifyPPWXj-9RFFJfyBPGnnHxpM?usp=sharing
Download the “train.csv” file to data folder.

Teamwork

Ran Huo
Model implementation 60%
Project proposal
Yin Jiang
Model implementation 40%
Final report and presentation

References

[1] Minaee, S & Amirali, A. 2019. Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network.
[2] Friesen, E., and P. Ekman. "Facial action coding system: a technique for the measurement of facial movement." Palo Alto, 1978.
[3] Jaderberg, Max, Karen Simonyan, and Andrew Zisserman. "Spatial transformer networks." Advances in neural information processing systems, 2015.
[4] Challenges in Representation Learning: Facial Expression Recognition Challenge https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data