Rita samir (25)

Sohayla mohamed (31)

Expression Recognition

March 05, 2020

The proposed problem

Humans use different forms of communications such as speech, hand gestures and emotions. Being able to understand one's emotions and the encoded feelings is an important factor for an appropriate and correct understanding. Facial expression is one of the most important components in daily communications of human beings. It is generated by movements of facial muscles. While different people have different kinds of facial expressions caused by their own expressive styles or personalities, many studies show that there are several types of basic expressions shared by different peoples with different cultural and ethnic backgrounds.

Current state of the art performances

- For CK+ dataset

Paper	Accuracy
Frame attention networks for facial expression recognition in videos	99.7%
Facial Motion Prior Networks for Facial Expression Recognition	98%

- For MMI dataset

Paper	Accuracy
DeXpression: Deep Convolutional Neural Network for Expression Recognition	98.63%

[Facial Motion Prior Networks for Facial Expression Recognition](#)	82.74%

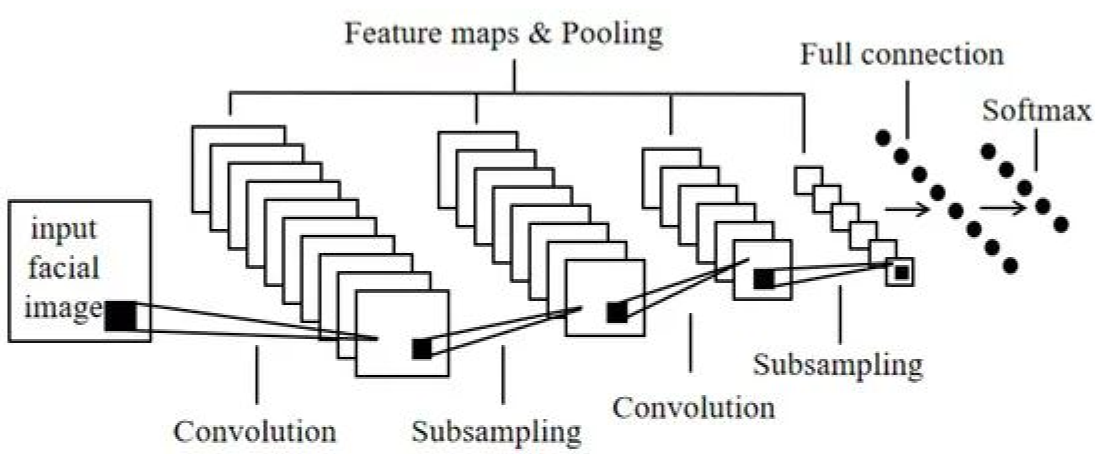
- For FER+ dataset

Paper	Accuracy
Efficient Facial Feature Learning with Wide Ensemble-based Convolutional Neural Networks	87.15%
Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution	85%

Available models and solutions

Deep learning-based approaches highly reduce the reliance on image preprocessing and feature extraction and are more robust to the environments with different elements, e.g., illumination and occlusion, which means that they can greatly outperform the conventional approaches. In addition, it has potential capability to handle high volume data.

- **Convolutional Neural Network (CNN)**



- **Deep Belief Network (DBN)**
    - The DBN is based on Restricted Boltzmann Machine (RBM) and its feature extraction of input signal is unsupervised and abstract.The FER method based on

DBN can learn the abstract information of facial images automatically. Combined with other components, DBN has proved to be an effective FER approach.

- . **Long Short-Term Memory (LSTM)**
    - An RNN composed of LSTM units is commonly referred to as an LSTM network, which is well suited for the temporal features extraction of consecutive frames.
- **Generative Adversarial Network(GAN)**
    - GAN is an unsupervised learning model composed of a generative network and a discriminative network.
    - The generator frontalises a frontal face image based on the input non-frontal image while retaining the identity and expression information, and the discriminator is trained to distinguish and recognise. This face frontalization system is demonstrated as valid for FER with visible head pose variations.
    - GANs are employed to train the generator to generate six basic expressions from a face image while CNN is fine-tuned for each single identity sub-space expression classification.

Datasets

Having sufficient labeled training data that include as many variations of the populations and environments as possible is important for the design of a deep expression recognition system. In this section, we discuss the publicly available databases that contain basic expressions :

- The Extended CohnKanade (CK+) http://www.consortium.ri.cmu.edu/ckagree/
- The MMI https://mmifacedb.eu/
- The Japanese Female Facial Expression (JAFFE) http://www.kasrl.org/jaffe.html
- EmotioNet http://cbcsl.ece.ohio-state.edu/dbform_emotionet.html
- ExpW http://mmlab.ie.cuhk.edu.hk/projects/socialrelation/index.html
- FER 2013, https://datarepository.wolframcloud.com/resources/FER-2013

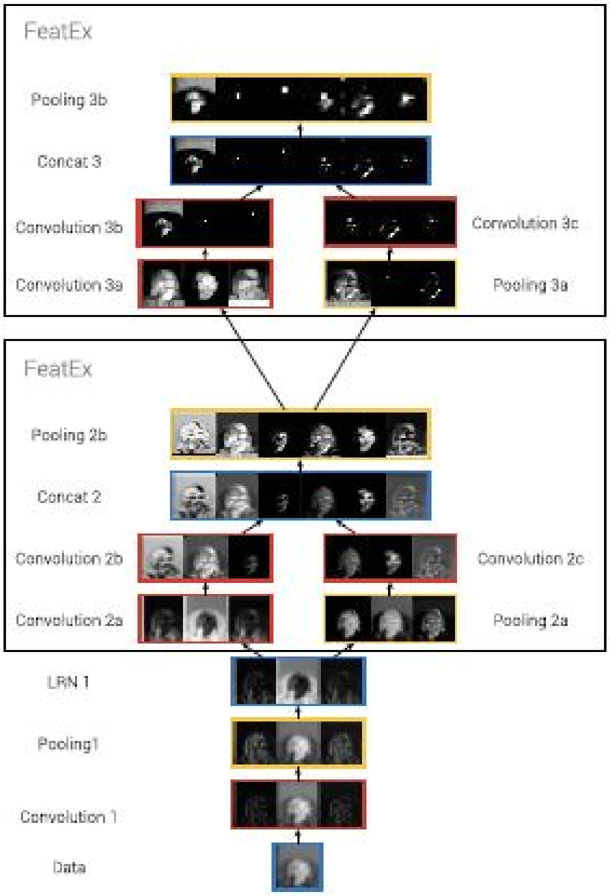
- FER+, https://github.com/microsoft/FERPlus?fbclid=IwAR19fs2UgZGpyIrwlN19qsKn5j_wlAHpkM1xCWs7UVgC1iFSaOnDRxQYHoQ
- others

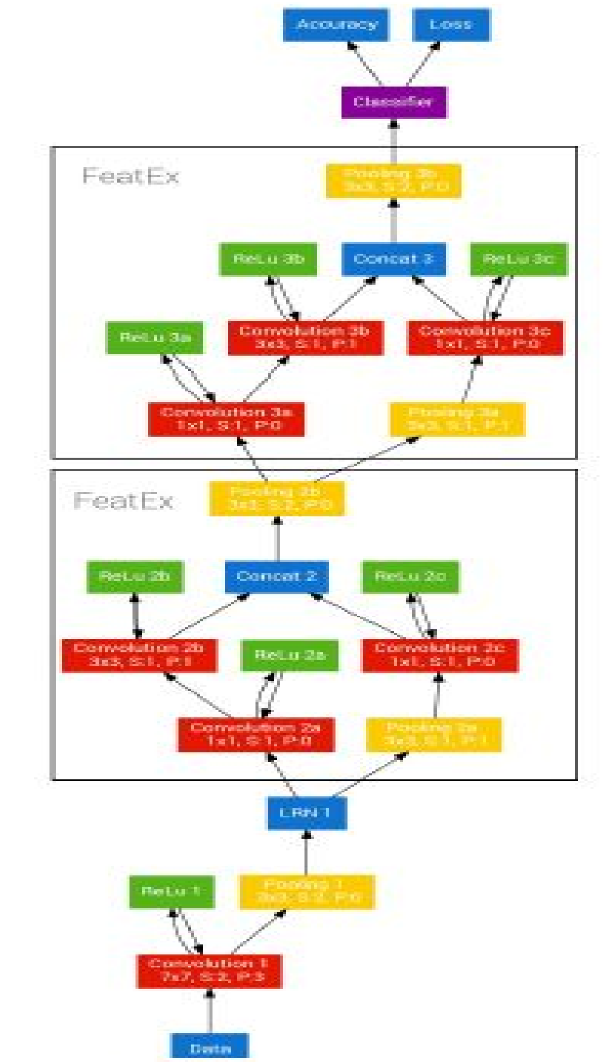
Database	Samples	Subject	Condit.	Elicit.	Expression distribution
CK+ [33]	593 image sequences	123	Lab	P & S	6 basic expressions plus contempt and neutral
MMI [34], [35]	740 images and 2,900 videos	25	Lab	P	6 basic expressions plus neutral
JAFFE [36]	213 images	10	Lab	P	6 basic expressions plus neutral
TFD [37]	112,234 images	N/A	Lab	P	6 basic expressions plus neutral
EmotioNet [43]	1,000,000 images	N/A	Web	P & S	23 basic expressions or compound expressions
ExpW [47]	91,793 images	N/A	Web	P & S	6 basic expressions plus neutral

Model to be used

1. Description

The proposed deep Convolutional Neural Network architecture consists of four parts. The first part automatically preprocesses the data. This begins with Convolution 1, which applies 64 different filters. The next layer is Pooling 1, which down-samples the images and then they are normalized by LRN 1. The next steps are the two FeatEx (Parallel Feature Extraction Block) blocks. The features extracted by these blocks are forwarded to a fully connected layer, which uses them to classify the input into the different emotions.





2. Why this model

Our choice criteria was a trade off between simple architecture and great performance compared with other state of the art models. In our opinion this model is a great choice since its architecture is based on CNN models and achieves great performance on some known datasets; even leads them in accuracy on the MMI dataset.

3. Evaluation metric

Accuracy using the confusion matrix :

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}$$

TP: True Positive, TN: True Negative, FN: False Negative, FP: False Positive

4. Dataset

Train our models using CK+ dataset, to try to reproduce its results before making any updates.

The Extended CohnKanade (CK+) database is the most extensively used laboratory-controlled database for evaluating FER systems. CK+ contains 593 video sequences from 123 subjects. The sequences vary in duration from 10 to 60 frames and show a shift from a neutral facial expression to the peak expression. Among these videos, 327 sequences from 118 subjects are labeled with seven basic expression labels (anger, contempt, disgust, fear, happiness, sadness, and surprise) based on the Facial Action Coding System (FACS). Because CK+ does not provide specified training, validation and test sets, the algorithms evaluated on this database are not uniform. For static-based methods, the most common data selection method is to extract the last one to three frames with peak formation and the first frame (neutral face) of each sequence. Then, the subjects are divided into n groups for person-independent n-fold cross-validation experiments, where commonly selected values of n are 5, 8 and 10.

5. plots/graphs

Visualize our model’s accuracy using accuracy vers epochs plots..

The proposed updates to the literature model

- FER2013 dataset is one of the most challenging datasets in facial expression recognition task, as the following table shows :

	[68] Cubic SVM+HoG	57.17
FER2013 [114]	[84] CNN	72.10
	[87] CNN(DeeperCNN)	61.10

State of the art models have struggled to obtain a good accuracy on the dataset due to human-based labeling which can be very subjective and ambiguous expressions in images, which prompted researchers at microsoft Barsoum et la.[1] to update the dataset to what's now known as FER+.

- Even with the update, many models stray away from using FER+ as their dataset which gives us room to try and check the state of the art models on it. The highest accuracy recorded for FER+ is obtained by a paper published in 2020 by Siqueira et la.[3].
- The proposed literature model based its experiments on CK+ and MMI datasets and holds the highest accuracy on MMI dataset compared to other models, we want to run the model on FER+ dataset and do hyperparameters tuning to check the highest accuracy we can get.

Graduation projects

- **Breast Mass Classification and Segmentation in Digital Mammogram**

  Breast mass is one of the most distinctive signs for diagnosis of breast cancer, and its marginal information reflects the growth pattern and biological characteristics. Generally speaking, benign masses are regular in shape, and masses with irregular margins are often malignant. So, our problem considers two tasks, predicting if breast mass is benign or malignant and segmenting it. We train two models separately for each task. Mainly, Our network is encoder-decoder architecture: the encoder is a densely-connected CNN and the decoder is a CNN integrated with AGs. Our dataset is CBIS-DDSM dataset, includes approximately 2,500 cases and every case contains two views of each breast, as well as some associated patient information (age, breast density rating, rating for abnormalities and keyword description of abnormalities) and image information (scanner, spatial resolution and so on). Images containing suspicious areas have associated pixel-level ‘‘ground truth’’ information about the locations and types of suspicious regions.

- **Text to Image Generation using GANs:**
  Text Visualization is a fundamental problem towards automatically generating images according to natural language descriptions. It is a challenging problem in computer vision that opens new avenues for a lot of innovative applications. It also drives research progress in multimodal learning and inference across vision and language, which is one of the most active research areas in recent years. Text to image synthesis has many exciting and practical applications such as photo editing or computer-aided content creation.

  Recently, Automatic synthesis of realistic images from text has become popular with deep convolutional and recurrent neural network architectures to aid in learning discriminative text feature representations. In another domain, Deep Convolutional Generative Adversarial Networks are able to synthesize images such as interiors of bedrooms from a random noise vector sampled from a normal distribution.

  In our work, we develop a novel deep architecture and GAN formulation to effectively bridge these advances in text and image modeling, translating visual concepts from characters to pixels. The featured algorithms learn a text feature representation that captures the important visual details and then use these features to synthesize a compelling image that a human might not predict correctly.

Citations

[1] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer and Zhengyou Zhang Microsoft Research, Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution

[2]Peter Burkert, Felix Trier, Muhammad Zeshan Afzal , Andreas Dengel and Marcus Liwicki German Research Center for Artificial Intelligence, DeXpression: Deep Convolutional Neural Network for Expression Recognition

[3] Kai Wang Shenzhen, Xiaojiang Peng Shenzhen Institutes of Advanced Technology Jianfei Yang Nanyang Shijian Lu Nanyang Technological University, Singapore, Suppressing Uncertainties for Large-Scale Facial Expression Recognition

Online resources/papers

1. Deep Facial Expression Recognition: A Survey
2. Article of Facial Expression Recognition: A Survey
3. Facial Action Coding System (FACS) - A Visual Guidebook
4. Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution
5. MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Face Images
6. Suppressing Uncertainties for Large-Scale Facial Expression Recognition
7. Facial expression databases