ECE 176 Final Project

CNN with SE for Facial Expression Recognition

Laura Fleig

UCSD

Winter Quarter 2024

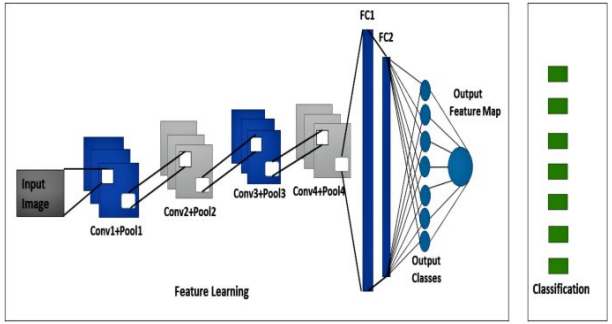


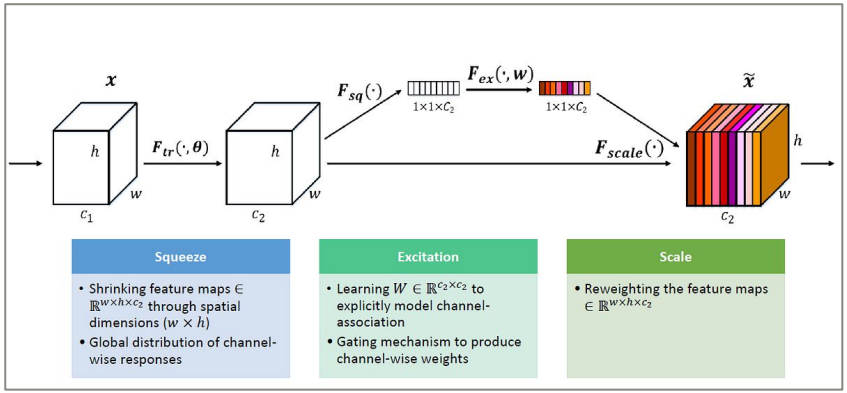
Key Terms

Facial expression recognition (FER)

Convolutional Neural Network (CNN)

Squeeze-and-Excite (SE) blocks

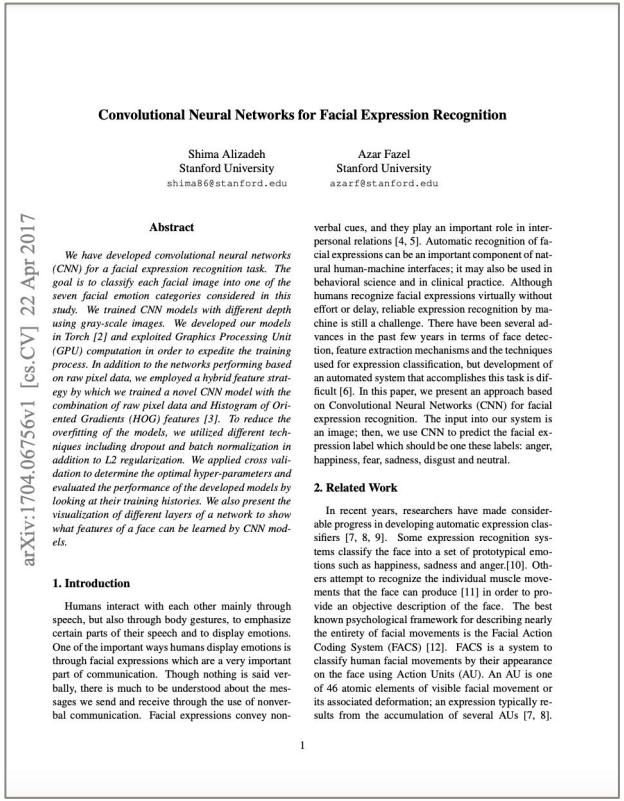




CNN example image source: https://media.springernature.com/m685/springer-static/image/art%3A10.1038%2Fs41598-022-11173-0/MediaObjects/41598_2022_11173_Fig7_HTML.png
SE image source: https://miro.medium.com/v2/resize:fit:1200/1*hljnSRy8uOILJwp581BMtg.png

Original Reference

S. Alizadeh and A. Fahel, “Convolutional Neural Networks for Facial Expression Recognition”, 2017.



Convolutional Neural Networks for Facial Expression Recognition

Shima Alizadeh
Stanford University
shima86@stanford.edu

Azar Fazel
Stanford University
azarf@stanford.edu

Abstract

We have developed convolutional neural networks (CNN) for a facial expression recognition task. The goal is to classify each facial image into one of the seven facial emotion categories considered in this study. We trained CNN models with different depth using gray-scale images. We developed our models in Torch [2] and exploited Graphics Processing Unit (GPU) computation in order to expedite the training process. In addition to the networks performing based on raw pixel data, we employed a hybrid feature strategy by which we trained a novel CNN model with the combination of raw pixel data and Histogram of Oriented Gradients (HOG) features [3]. To reduce the overfitting of the models, we utilized different techniques including dropout and batch normalization in addition to L2 regularization. We applied cross validation to determine the optimal hyper-parameters and evaluated the performance of the developed models by looking at their training histories. We also present the visualization of different layers of a network to show what features of a face can be learned by CNN models.

1. Introduction

Humans interact with each other mainly through speech, but also through body gestures, to emphasize certain parts of their speech and to display emotions. One of the important ways humans display emotions is through facial expressions which are a very important part of communication. Though nothing is said verbally, there is much to be understood about the messages we send and receive through the use of nonverbal communication. Facial expressions convey nonverbal cues, and they play an important role in interpersonal relations [4, 5]. Automatic recognition of facial expressions can be an important component of natural human-machine interfaces; it may also be used in behavioral science and in clinical practice. Although humans recognize facial expressions virtually without effort or delay, reliable expression recognition by machine is still a challenge. There have been several advances in the past few years in terms of face detection, feature extraction mechanisms and the techniques used for expression classification, but development of an automated system that accomplishes this task is difficult [6]. In this paper, we present an approach based on Convolutional Neural Networks (CNN) for facial expression recognition. The input into our system is an image; then, we use CNN to predict the facial expression label which should be one these labels: anger, happiness, fear, sadness, disgust and neutral.

2. Related Work

In recent years, researchers have made considerable progress in developing automatic expression classifiers [7, 8, 9]. Some expression recognition systems classify the face into a set of prototypical emotions such as happiness, sadness and anger.[10]. Others attempt to recognize the individual muscle movements that the face can produce [11] in order to provide an objective description of the face. The best known psychological framework for describing nearly the entirety of facial movements is the Facial Action Coding System (FACS) [12]. FACS is a system to classify human facial movements by their appearance on the face using Action Units (AU). An AU is one of 46 atomic elements of visible facial movement or its associated deformation; an expression typically results from the accumulation of several AUs [7, 8].

Experimental Overview

1. Data augmentation/preprocessing
2. Shallow network
3. Deeper network
4. SE network based on deeper network
5. Improved SE network
6. Train on FER-2013
7. Train final model on RAF-DB

Methods: Shallow Network

Layer	Output Shape	Param #
Conv2D	(48, 48, 32)	320
BatchNormalization	(48, 48, 32)	128
ReLU	(48, 48, 32)	0
Dropout	(48, 48, 32)	0
Conv2D	(48, 48, 64)	18496
BatchNormalization	(48, 48, 64)	256
ReLU	(48, 48, 64)	0
MaxPooling2D	(24, 24, 64)	0
Dropout	(24, 24, 64)	0
Flatten	(36864)	0
Dense	(512)	18874880
ReLU	(512)	0
Dropout	(512)	0
Dense	(7)	3591

Table 1: Shallow Network Architecture

Methods: Deeper Network

Layer	Output Shape	Param #
Conv2D	(48, 48, 64)	640
BatchNormalization	(48, 48, 64)	256
ReLU	(48, 48, 64)	0
MaxPooling2D	(24, 24, 64)	0
Dropout	(24, 24, 64)	0
Conv2D	(24, 24, 128)	204928
BatchNormalization	(24, 24, 128)	512
ReLU	(24, 24, 128)	0
MaxPooling2D	(12, 12, 128)	0
Dropout	(12, 12, 128)	0
Conv2D	(12, 12, 512)	590336
BatchNormalization	(12, 12, 512)	2048
ReLU	(12, 12, 512)	0
MaxPooling2D	(6, 6, 512)	0
Dropout	(6, 6, 512)	0
Conv2D	(6, 6, 512)	2359808
BatchNormalization	(6, 6, 512)	2048
ReLU	(6, 6, 512)	0
MaxPooling2D	(3, 3, 512)	0
Dropout	(3, 3, 512)	0
Flatten	(4608)	0
Dense	(256)	1179904
BatchNormalization	(256)	1024
ReLU	(256)	0
Dropout	(256)	0
Dense	(512)	131584
BatchNormalization	(512)	2048
ReLU	(512)	0
Dropout	(512)	0
Dense	(7)	3591

Table 2: Deeper Model Architecture

Methods: SE Network

Layer	Output Shape	Param #
InputLayer	(48, 48, 1)	0
Conv2D	(48, 48, 64)	640
ReLU	(48, 48, 64)	0
BatchNormalization	(48, 48, 64)	256
GlobalAveragePooling2D	(64)	0
Dense	(4)	256
Dense	(64)	256
Reshape	(1, 1, 64)	0
Multiply	(48, 48, 64)	0
Conv2D	(48, 48, 128)	73856
ReLU	(48, 48, 128)	0
BatchNormalization	(48, 48, 128)	512
GlobalAveragePooling2D	(128)	0
Dense	(8)	1024
Dense	(128)	1024
Reshape	(1, 1, 128)	0
Multiply	(48, 48, 128)	0
MaxPooling2D	(24, 24, 128)	0
Dropout	(24, 24, 128)	0
Conv2D	(24, 24, 512)	590336
ReLU	(24, 24, 512)	0
BatchNormalization	(24, 24, 512)	2048
GlobalAveragePooling2D	(512)	0
Dense	(32)	16384
Dense	(512)	16384
Reshape	(1, 1, 512)	0
Multiply	(24, 24, 512)	0
MaxPooling2D	(12, 12, 512)	0
Conv2D	(12, 12, 512)	2359808
ReLU	(12, 12, 512)	0
BatchNormalization	(12, 12, 512)	2048
GlobalAveragePooling2D	(512)	0
Dense	(32)	16384
Dense	(512)	16384
Reshape	(1, 1, 512)	0
Multiply	(12, 12, 512)	0
MaxPooling2D	(6, 6, 512)	0
Dropout	(6, 6, 512)	0
Flatten	(18432)	0
Dense	(256)	4718848
BatchNormalization	(256)	1024
Dropout	(256)	0
Dense	(512)	131584
BatchNormalization	(512)	2048
Dropout	(512)	0
Dense	(7)	3591

Table 3: SE Model Architecture

Methods: Improved SE Network

Layer	Output Shape	Param #
InputLayer	(48, 48, 1)	0
Conv2D	(48, 48, 32)	320
ReLU	(48, 48, 32)	0
BatchNormalization	(48, 48, 32)	128
GlobalAveragePooling2D	(32)	0
Dense	(2)	64
Dense	(32)	64
Reshape	(1, 1, 32)	0
Multiply	(48, 48, 32)	0
Conv2D	(48, 48, 64)	18496
ReLU	(48, 48, 64)	0
BatchNormalization	(48, 48, 64)	256
GlobalAveragePooling2D	(64)	0
Dense	(4)	256
Dense	(64)	256
Reshape	(1, 1, 64)	0
Multiply	(48, 48, 64)	0
MaxPooling2D	(24, 24, 64)	0
Dropout	(24, 24, 64)	0
Conv2D	(24, 24, 128)	73856
ReLU	(24, 24, 128)	0
BatchNormalization	(24, 24, 128)	512
GlobalAveragePooling2D	(128)	0
Dense	(8)	1024
Dense	(128)	1024
Reshape	(1, 1, 128)	0
Multiply	(24, 24, 128)	0
MaxPooling2D	(12, 12, 128)	0
Conv2D	(12, 12, 128)	147584
ReLU	(12, 12, 128)	0
BatchNormalization	(12, 12, 128)	512
GlobalAveragePooling2D	(128)	0
Dense	(8)	1024
Dense	(128)	1024
Reshape	(1, 1, 128)	0
Multiply	(12, 12, 128)	0
MaxPooling2D	(6, 6, 128)	0
Dropout	(6, 6, 128)	0
Flatten	(4608)	0
Dense	(1024)	4719616
BatchNormalization	(1024)	4096
Dropout	(1024)	0
Dense	(256)	262400
BatchNormalization	(256)	1024
Dropout	(256)	0
Dense	(7)	1799

Table 4: Improved SE Model Architecture

Dataset 1: FER-2013

MANAS SAMBARE · UPDATED 4 YEARS AGO

1023 | New Notebook | Download (63 MB)

FER-2013

Learn facial expressions from an image

Data Card | Code (312) | Discussion (6) | Suggestions (0)

About Dataset

The data consists of 48×48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centred and occupies about the same amount of space in each image.

The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). The training set consists of 28,709 examples and the public test set consists of 3,589 examples.

Usability
7.50

License
Database: Open Database, Cont...

Expected update frequency
Not specified

Data Augmentation

```
train_data_gen = ImageDataGenerator(
    rescale=1./255,
    validation_split=0.2,
    horizontal_flip=True,
    rotation_range=10,
    width_shift_range=0.08,
    height_shift_range=0.08,
    brightness_range=[0.8, 1.2],
    zoom_range=[0.92, 1.08],
    fill_mode='nearest'
)
```

Results on FER-2013

- Shallow network: 41%
- Deeper network: 56%
- Add SE blocks: 61%
- Improved SE network: 63%
  - 62.3% test accuracy

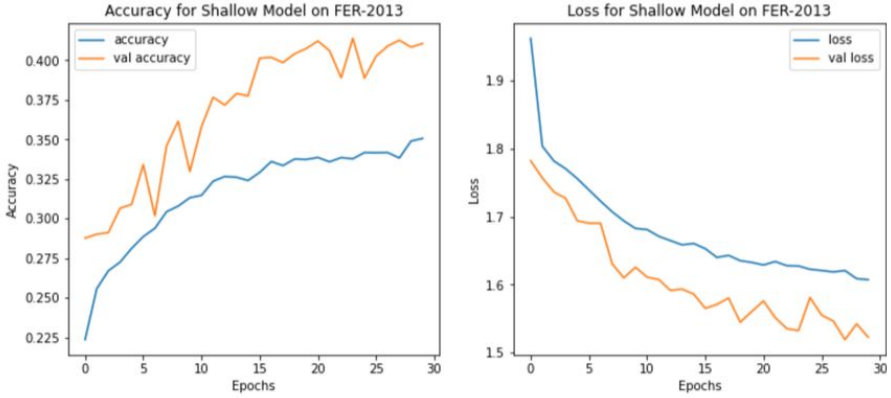


Figure 1: Shallow Model Plots

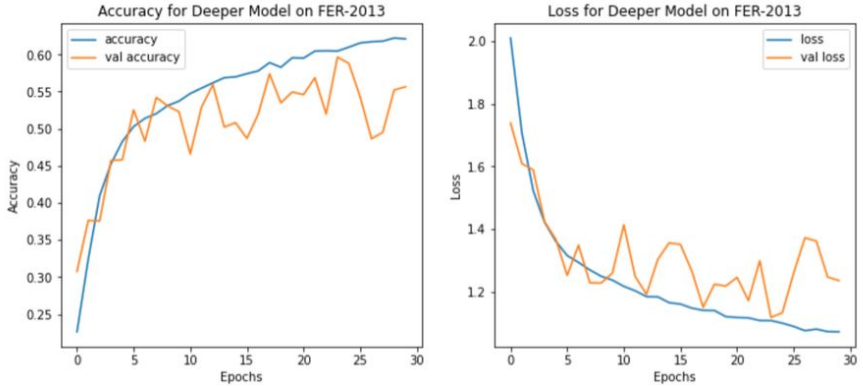


Figure 2: Deeper Model Plots

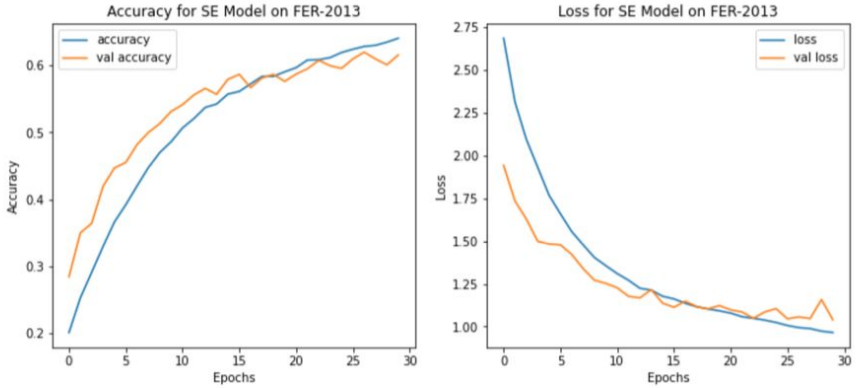


Figure 3: SE Model Plots

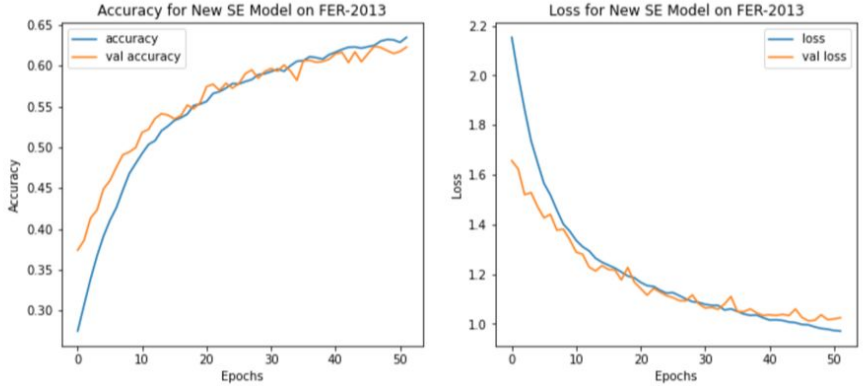


Figure 4: Improved SE Model Plots on FER-2013

Dataset 2: RAF-DB

DEV-SHUVOALOK · UPDATED 6 MONTHS AGO

8 | New Notebook | Download (40 MB)

RAF-DB DATASET

For Recognize emotion from Facial expression

Data Card | Code (18) | Discussion (0) | Suggestions (0)

About Dataset

The Real-world Affective Faces Database (RAF-DB) is a dataset for facial expression. This version Contains 15000k facial images tagged with basic or compound expressions by 40 independent taggers. Images in this database are of great variability in subjects' age, gender and ethnicity, head poses, lighting conditions, occlusions, (e.g. glasses, facial hair or self-occlusion), post-processing operations (e.g. various filters and special effects), etc.

For More Info Visit: Here

Usability ⓘ
8.82

License
Other (specified in description)

Expected update frequency
Not specified

Results on RAF-DB

- Improved SE network:
  - 77% validation accuracy on initial training on labeled dataset
  - 80% validation accuracy on combined training round
  - 76.5% test accuracy

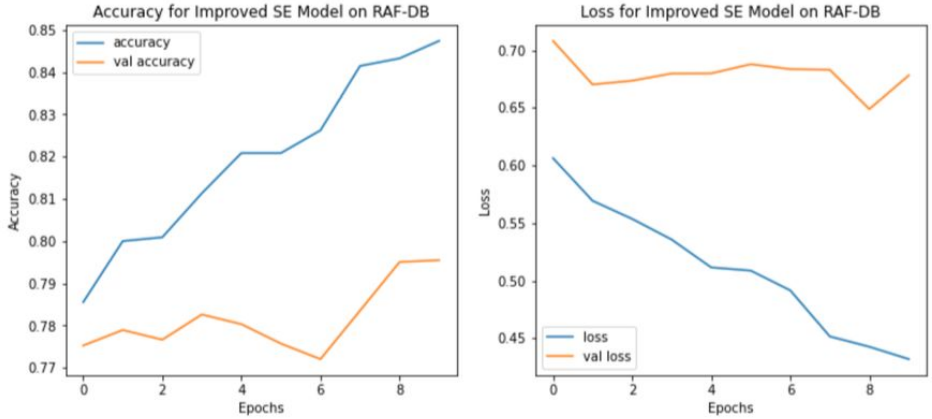


Figure 5: Improved SE Model Plots on RAF-DB (just combined training epochs)

Figure 6: Improved SE Model Prediction Visualization

References

[1] S. Minaee and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," 2019.

[2] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[3] Hujie-Frank, "Hujie-frank/senet: Squeeze-and-excitation networks." [Online]. Available: https://github.com/hujie-frank/SENet

[4] S. Alizadeh and A. Fazel, "Convolutional neural networks for facial expression recognition," 2017.

[5] M. Sambare, "Fer-2013," Jul 2020. [Online]. Available: https://www.kaggle.com/datasets/msambare/fer2013

[6] S. Li, W. Deng, and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 2584–2593.

[7] S. Li and W. Deng, "Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 356–370, 2019.

[8] Z. Min, Q. Ge, and C. Tai, "Why the pseudo label based semi-supervised learning algorithm is effective?" 2023.

[9] J. Yu, Z. Cai, R. Li, G. Zhao, G. Xie, J. Zhu, W. Zhu, Q. Ling, L. Wang, C. Wang *et al.*, "Exploring large-scale unlabeled faces to enhance facial expression recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5803–5810.