Facial Expression Recognition

Luong Toan Bach 21521845
Nguyen Quoc Truong 21521604

Table of Contents

01. Introduction

Problem Description

FER stands for “Facial Expression Recognition”.
Or “Facial Emotion Recognition”

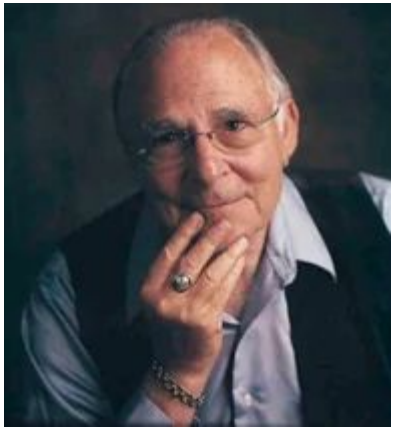
FER is a computer vision task aimed at identifying and categorizing emotional expressions(emotions) depicted on a human face.

Emotion

What is emotion?

There are many emotions but according to Paul Ekman.

Paul Ekman's theory



Paul Ekman

Paul Ekman theorized that some basic human emotions (happiness/enjoyment, sadness, anger, fear, surprise, disgust and contempt) are innate and shared by everyone, and that they are accompanied across cultures by universal facial expressions.

Motivation

Why do we need to FER?

- Healthcare
- Education
- Crime detection
- Business

Visualize

Input

Output

Surprise

- Digital image of face (eyes, nose, mouth)
- Labeled training set (labeled face image)

- String represent emotion

Challenges

Things that make this problem hard

- Illumination
- Low resolution ?
- Duplicate emotion
- Occlusion

02. Methodology

Feature Extraction

Turn image to vector

- Many Feature Extraction methods:
  - HOG
  - SIFT
  - ...

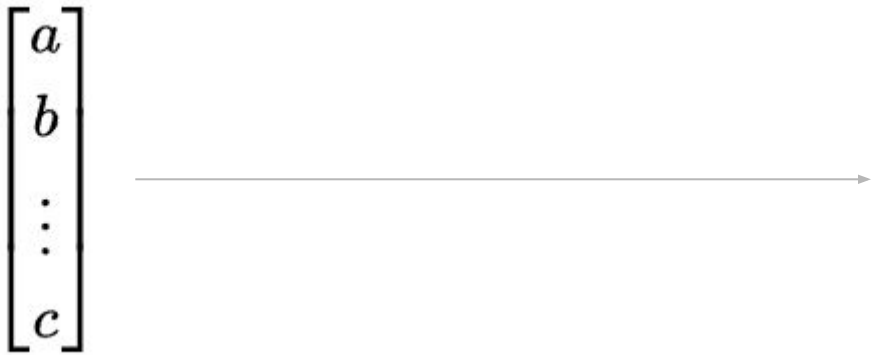
$$\begin{bmatrix} a \\ b \\ \vdots \\ c \end{bmatrix}$$

Turn “Vector” to Vector

Classification Model

Turn vec into prediction

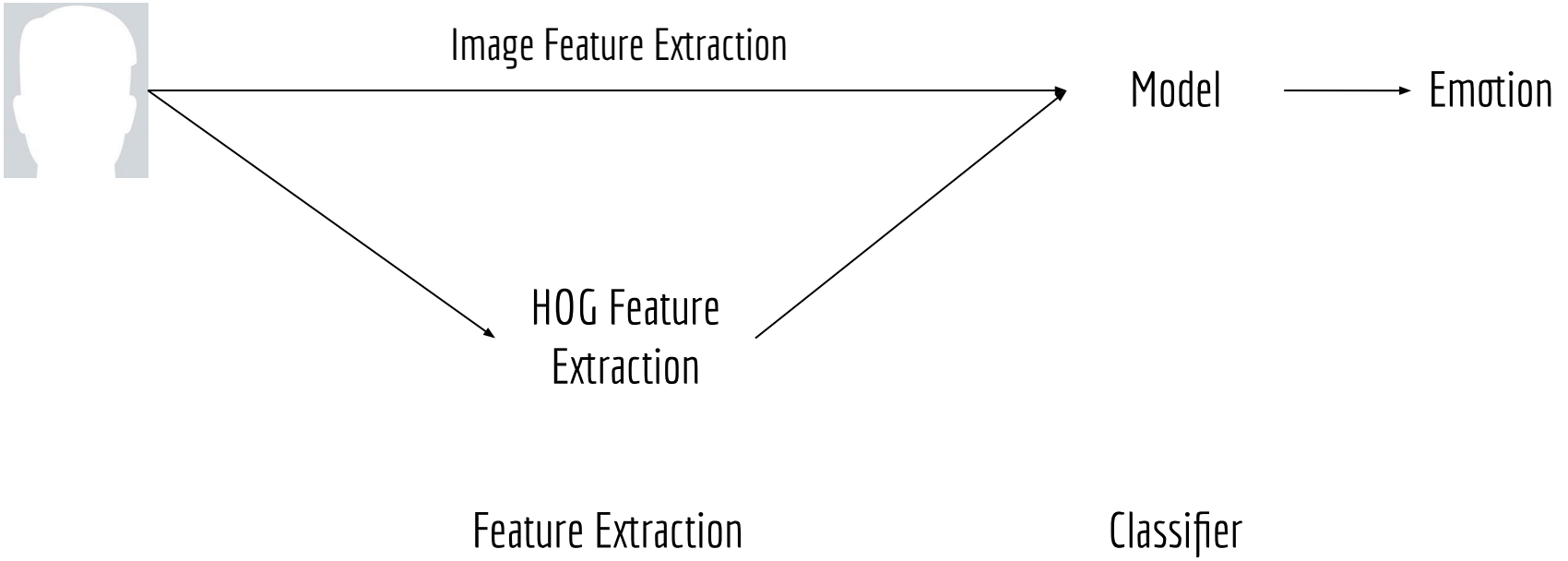
- Many Classification models:
  - Logistic regression
  - Naive Bayes
  - Decision Tree
  - Random Forest
  - ...



Turn vec into prediction

Pipeline

Input (Image)



03. Implementation

Dataset

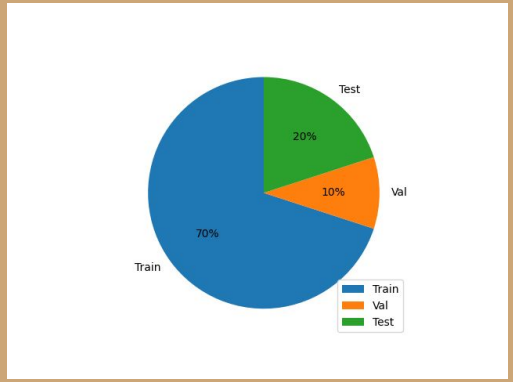
FER-2013

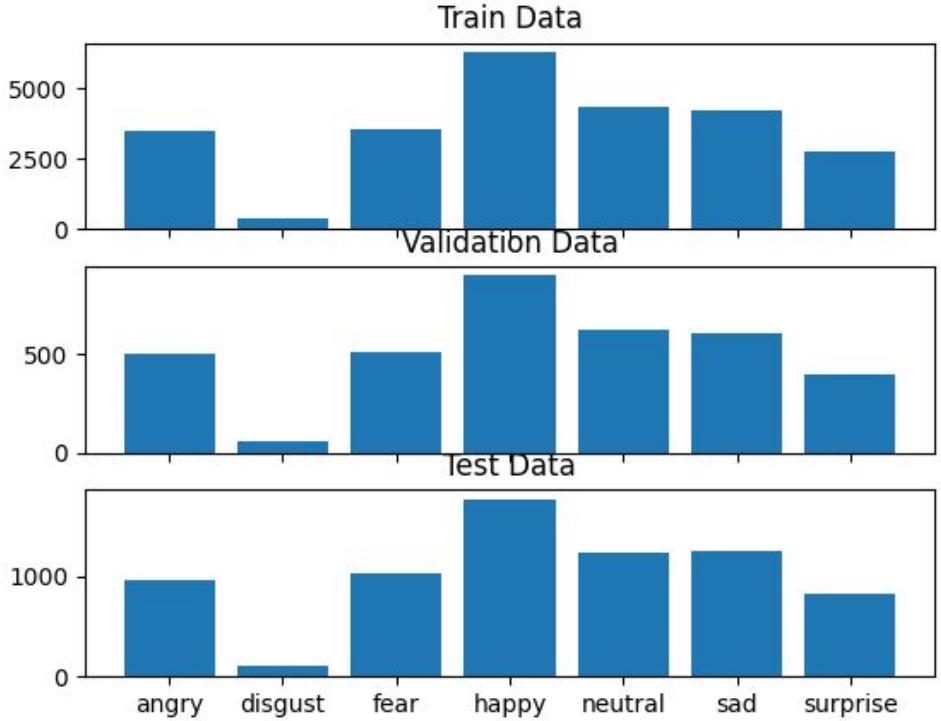
- The data consists of 48x48 pixel grayscale images of faces.
- The dataset consists 7 files corresponding to 7 emotions (sadness, happiness, fear, anger, surprise, disgust, neutral).
- Training set: 28,709 images (80%)
- Test set: 7,178 images (20%)
- Link

Dataset in practice

When we use it

- Split training set of dataset into 2 sets:
  - Training data 25,121 (70%)
  - Validation data 3,588 (10%)
- Test set 7,178 (20%)
- Normalize image feature





Distribution of emotion in each dataset

Metric

Accuracy is equal to f1-score(micro)

- The proportion of true predict to all predicts.

$$Accuracy = \frac{Correct\ prediction}{Total\ cases} * 100\%$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100\%$$

Results

We have try somes

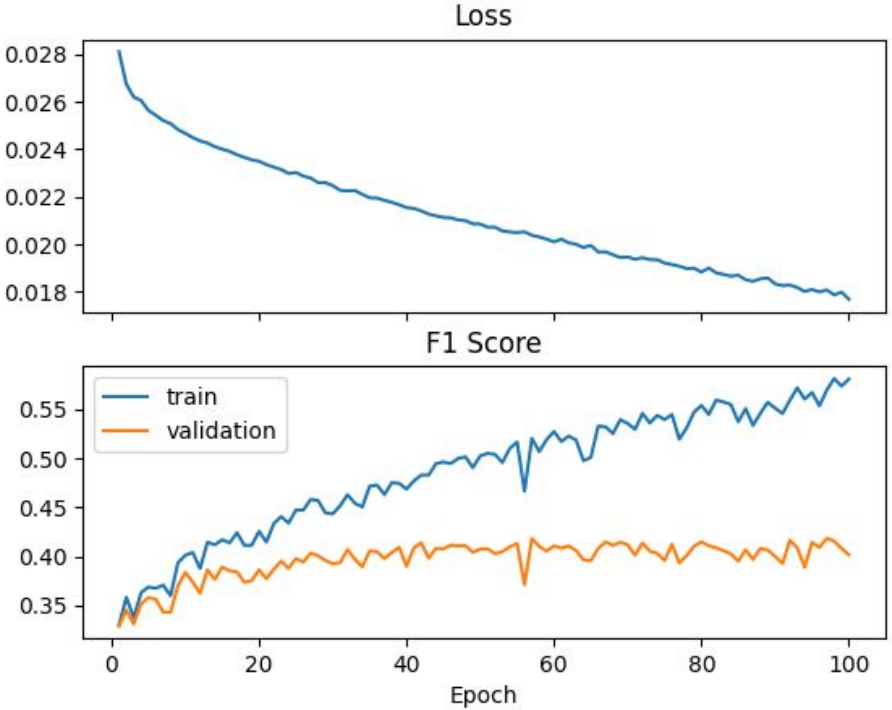
We have try some combination with different settings:

- Feature Extraction:
  - Vanilla image
  - HOG
  - Keypoint mask (harris)
- Classification Model:
  - KNN
  - SVM
  - MLP

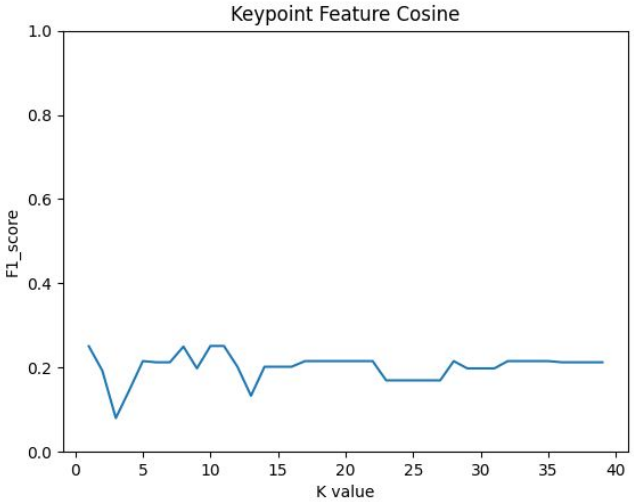
Settings

Different thing in same model

- KNN:
  - Feature
  - Number of K
    - Plot 0 < K < 41
  - Distance metric
- SVM
  - Feature
- MLP
  - Feature
  - Number node, layer



Turning Epoch



Turning K value

Turning hyperparameter

Model Type	Model	Feature	Setting	Layer	Accuracy
Lazy	KNN	Image	Euclidean		38.89%
			Cosine		38.86%
		HOG	Euclidean		49.59%
			Cosine		49.59%
		Keypoint	Euclidean		14.26%
			Cosine		17.29%
Linear	SVM	Image	Linear SVC		34.47%
		HOG			43.60%
		Keypoint			17.18%
Neural Network	MLP	Image	Airhmetic Mean	(2304, 1155, 7)	39.24%
		HOG		(900, 453, 7)	49.99%
		Keypoint		(2304, 1155, 7)	24.71%
		Image	Geometric Mean	(2304, 126, 7)	38.98%
		HOG		(900, 79, 7)	44.47%
		Keypoint		(2304, 126, 7)	24.71%
		Image	Jeff Heaton's	(2304, 1543, 7)	39.96%
		HOG		(900, 607, 7)	51.03%
		Keypoint		(2304, 1543, 7)	24.71%

- Total models: 18
  - KNN: 6
  - SVM: 3
  - MLP: 9

Results

Another approach

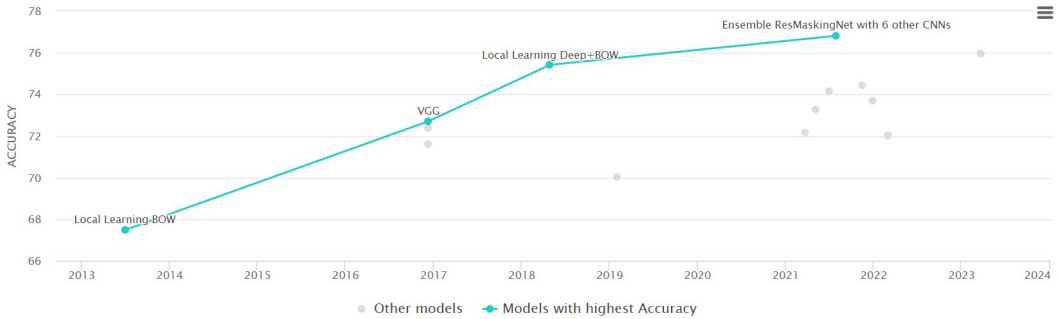
Better than us for sure

- **Sota** for this problem and dataset FER-2013 is Ensemble ResMaskingNet with 6 other CNNs

Facial Expression Recognition (FER) on FER2013

Leaderboard Dataset

View Accuracy by Date for All models



Filter: ResNet VGG Self-Attention untagged

Edit Leaderboard

Rank	Model	Accuracy	Extra Training Data	Paper	Code	Result	Year	Tags
1	**Ensemble ResMaskingNet with 6 other CNNs**	76.82	✓	Challenges in Representation Learning: A report on three machine learning contests			2021	
2	**Segmentation VGG-19**	75.97	✓	A novel facial emotion recognition model using segmentation VGG-19 architecture			2023	

FER2013 Benchmark (Facial Expression Recognition (FER)) | Papers With Code

04. Demo & Conclude

Demo

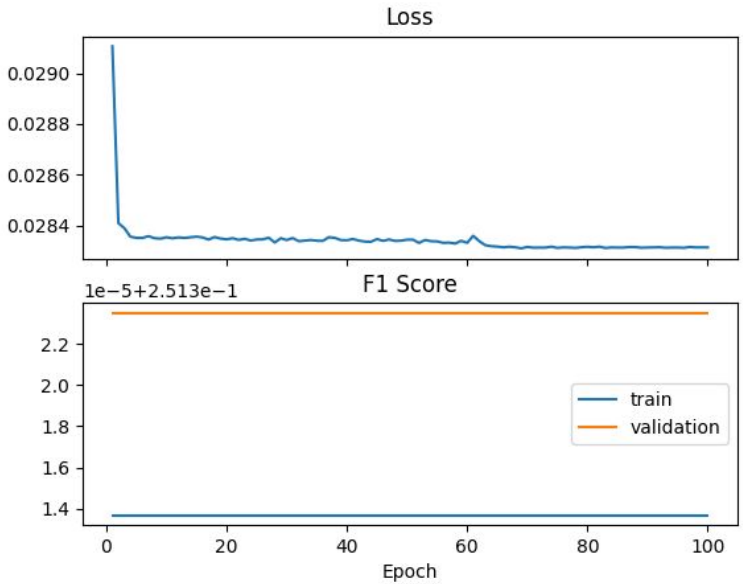
Link to demo: this

- Demo online through Google Colab

Conclude

Some words about our experiments

- Keypoint doesn’t work well.
- Feature Extraction is the heart of the algorithm.
- CNN doesn’t need FE because it consist Convolution Layer to do FE.



Model Type	Model	Feature	Setting	Layer	Accuracy
Lazy	KNN	Image	Euclidean		38.89%
			Cosine		38.86%
		HOG	Euclidean		49.59%
			Cosine		49.59%
		Keypoint	Euclidean		14.26%
			Cosine		17.29%
Linear	SVM	Image	Linear SVC		34.47%
		HOG			43.60%
		Keypoint			17.18%
Neural Network	MLP	Image	Airhmetic Mean	(2304, 1155, 7)	39.24%
		HOG		(900, 453, 7)	49.99%
		Keypoint		(2304, 1155, 7)	24.71%
		Image	Geometric Mean	(2304, 126, 7)	38.98%
		HOG		(900, 79, 7)	44.47%
		Keypoint		(2304, 126, 7)	24.71%
		Image	Jeff Heaton's	(2304, 1543, 7)	39.96%
		HOG		(900, 607, 7)	51.03%
		Keypoint		(2304, 1543, 7)	24.71%

Evidence

Future Work

- Using conner feature instead of mask.
- Turning batch size
- Understanding why keypoint feature doesn’t work well

Do you have any questions?

Please contact us

Thanks for your listening