AI PROGRAMMING PROJECT
AIP391
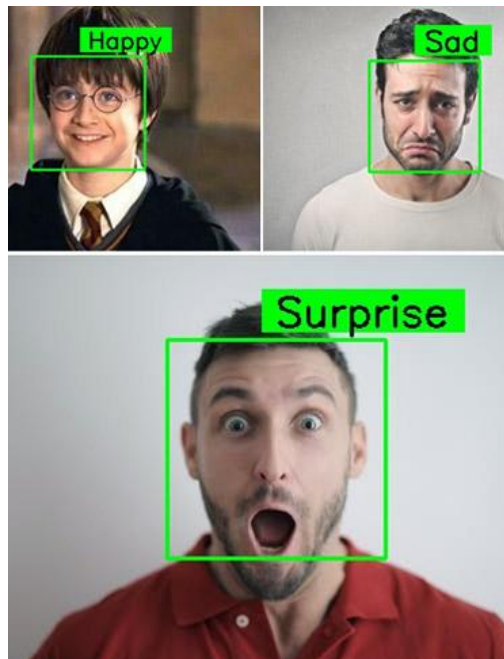
GROUP 05 <3
SE150556 - Nguyễn Thanh Đảm
SE150347 - Nguyễn Phúc Nguyên Anh
SE150402 - Nguyễn Lê Phương Hà

What makes the difference between humans & machines?

Real Time

Facial Expression Recognition

1 Introduction

- **What is Facial Expression Recognition (FER)?**

★ <u>Analyze features</u> extracted from images or video frames.
★ <u>Classify emotion</u>s on human faces.

What are FER's application?

★ Human–computer interaction, Robotics, Computer graphics animation
★ Behavioral psychology: human behavior understanding, mental disorder detection, cognition human emotions, …
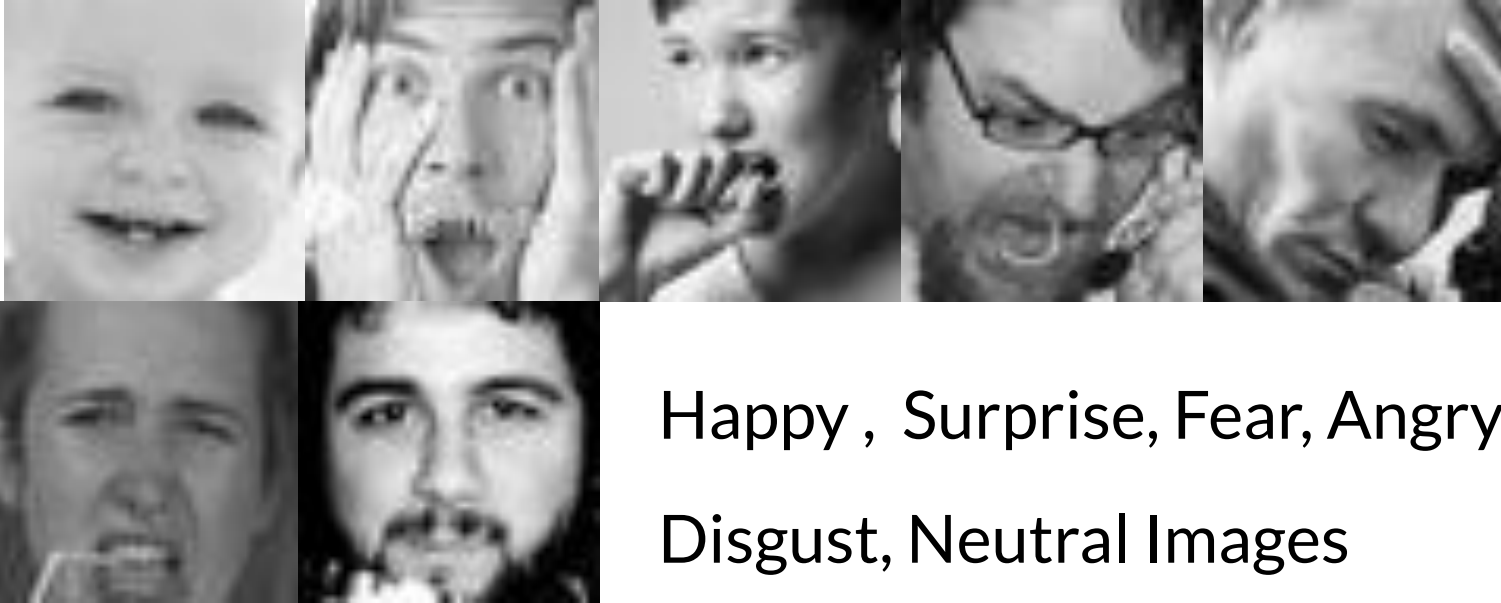★ Assist people with disabilities to communicate

https://youtu.be/OTmPw4iy0hk

2 Data : FER2013

Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
4,953	547	5,121	8,989	6,077	4,002	6,198

★ Used in ICML competitions and several research papers.
★ 7 facial expressions.
★ 35,887 images are normalized to 48x48 pixels in grayscale.
★ Be not a balanced dataset.

Data: https://www.kaggle.com/msambare/fer2013

Images Of Dataset

Happy , Surprise, Fear, Angry, Sad,

Disgust, Neutral Images

Anatomy and Hypotheses

Universality hypothesis:
- Facial expressions are signals of specific emotions
- Recognized by people regardless of culture,..

Discrete emotion theory:
Assume that there exists 6 basic emotions (angry, disgust,fear, happy, sad, surprise) and a neutral state.

@AlbertoEsFeliz

El secuestro de la amígdala: ante una amenaza (real o percibida), se produce un atajo neuronal donde la amígdala decide por supervivencia antes que nuestro neocórtex, dando una respuesta impulsiva, no lógica.

AMENAZA VISUAL → ALERTA → MIEDO

ANSIEDAD → BLOQUEO (NOS IMPIDE PODER HABLAR EN PÚBLICO)

Vía directa de tálamo a amígdala

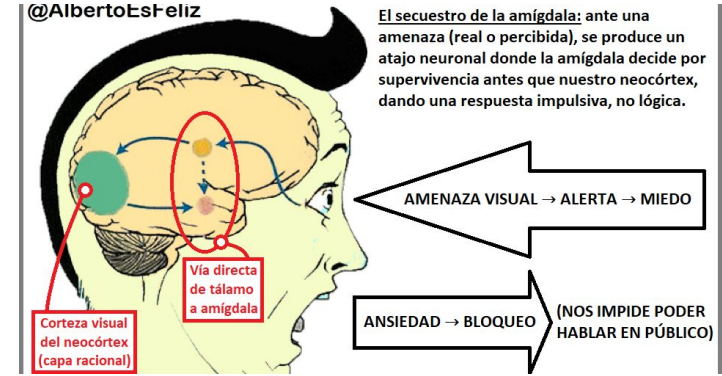Corteza visual del neocórtex (capa racional)

Figure 6.3: Facial expressions are governed by Amygdala and other cortices

Facial Action Coding System (FACS)

The Facial Action Coding System, or FACS (Ekman, Friesen, & Hager, 2002), is a systematic approach to describe what a face looks like when facial muscle movements have occurred.

Facial Actions

1 Inner Brow Raise	9 Nose wrinkle
2 Outer Brow Raise	12 Lip corner pull
4 Brow lower	15 Lip corner depress
5 Eye widen	17 Chin raise
6 Cheek raise	20 Lip stretch
7 Lid tighten	24 Lip press

Facial Action Coding System (FACS)

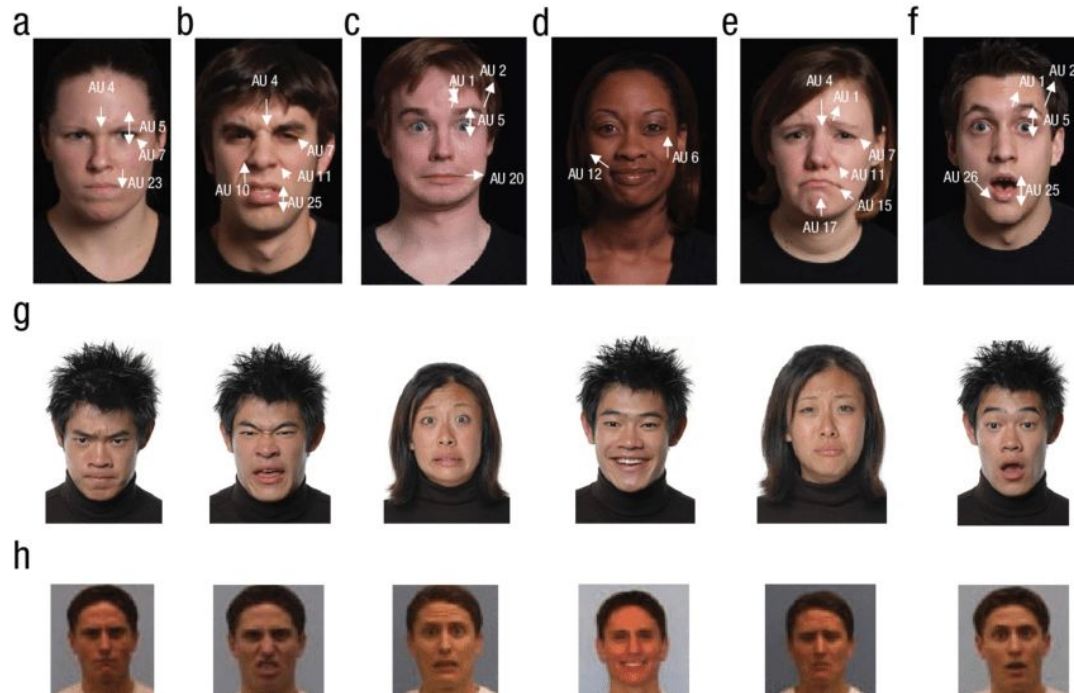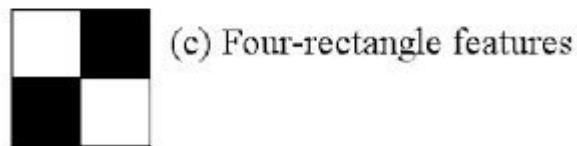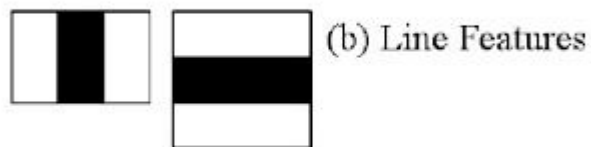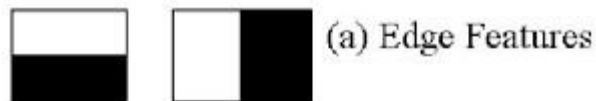| Basic expressions	Involved Action Units
Surprise	AU 1, 2, 5,15,16, 20, 26
Fear	AU 1, 2, 4, 5,15,20, 26
Disgust	AU 2, 4, 9, 15, 17
Anger	AU 2, 4, 7, 9,10, 20, 26
Happiness	AU 1, 6,12,14
Sadness	AU 1, 4,15, 23

Facial Expressions of Emotion

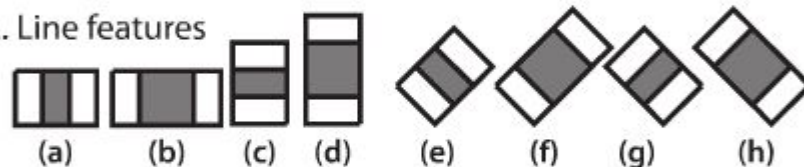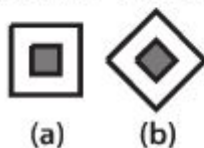Figure 1.1: Facial action ensembles for common-view facial expressions.

Haar-Like feature

(a) Edge Features

(b) Line Features

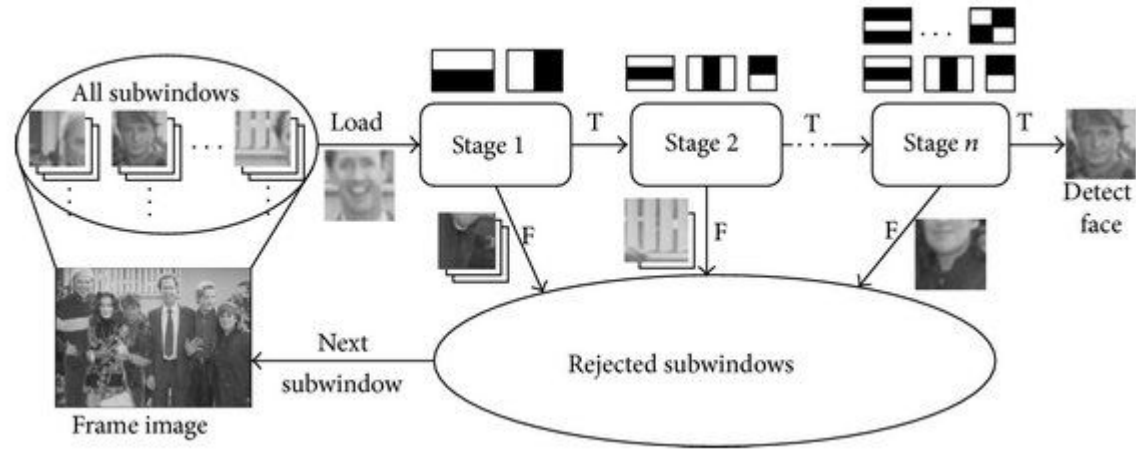(c) Four-rectangle features

1. Edge features
(a) (b) (c) (d)
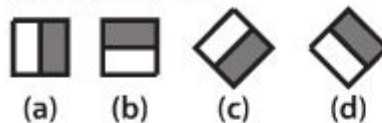
2. Line features
(a) (b) (c) (d) (e) (f) (g) (h)

3. Center-surround features
(a) (b)

$$f(x) = Tổng_{vùng\ đen}(các\ mức\ xám\ của\ pixel) - Tổng_{vùng\ trắng}(các\ mức\ xám\ của\ pixel)$$

Haar Cascade Face Detection algorithm

Convolution Neural Network (CNN)

Input

Kernel

Pooling   Pooling   Pooling

Convolution + ReLU   Convolution + ReLU   Convolution + ReLU

Flatten Layer

Fully Connected Layer

Output

0.2 — Horse
0.7 — Zebra
0.1 — Dog

SoftMax Activation Function

Feature Maps

Feature Extraction   Classification   Probabilistic Distribution
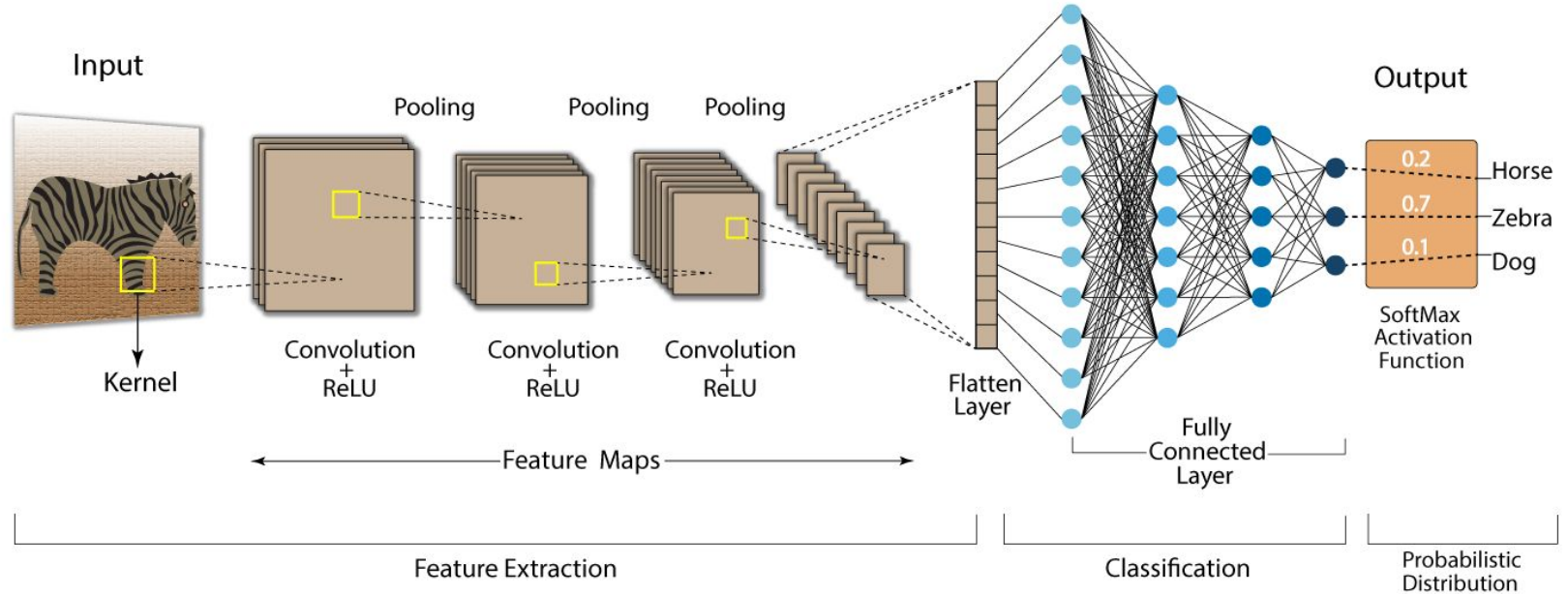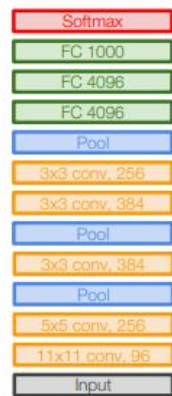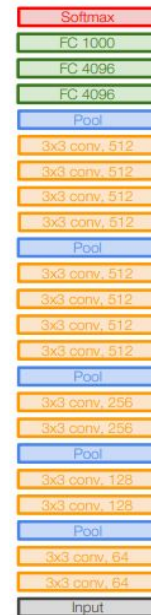
AlexNet
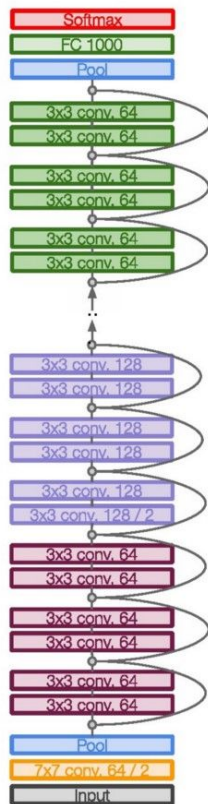
Resnet

Inception

VGG

TESTED DEEP ARCHITECTURES AND THEIR TEN-CROP TEST ACCURACY ON FER2013

Name	Architecture	Depth	Parameters	Accuracy
VGG	CCPCCPCCPCCPFF	10	1.8 m	72.7%
Inception	CIPIIPIIPIIPF	16	1.2 m	71.6%
ResNet	3R4R6R3RPF	33	5.3 m	72.4%

3. Method

Convolution + ReLU + BatchNorm
Max pooling
Fully connected + ReLU
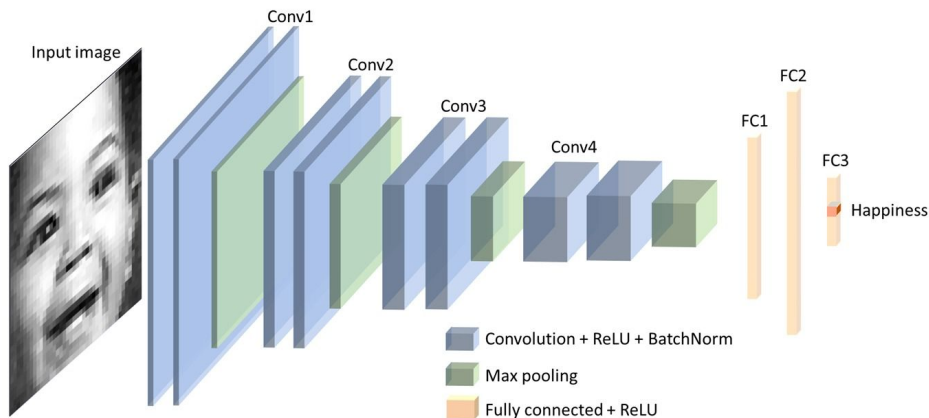
|				ConvNet Configuration			
	A	A-LRN	B	C	D	E
	11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)						
	conv3-64	conv3-64 LRN	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool						
	conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool						
	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool						
	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool						
	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool						
FC-4096						
FC-4096						
FC-1000						
soft-max						

Evaluation metrics

Accuracy:

A simple measurement for classification tasks, supply just one numerica

information. $$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

Confusion Matrix
• A table that describes the performance of a classification model.
• Much better with imbalanced data.
• Can describe where the model is confused.

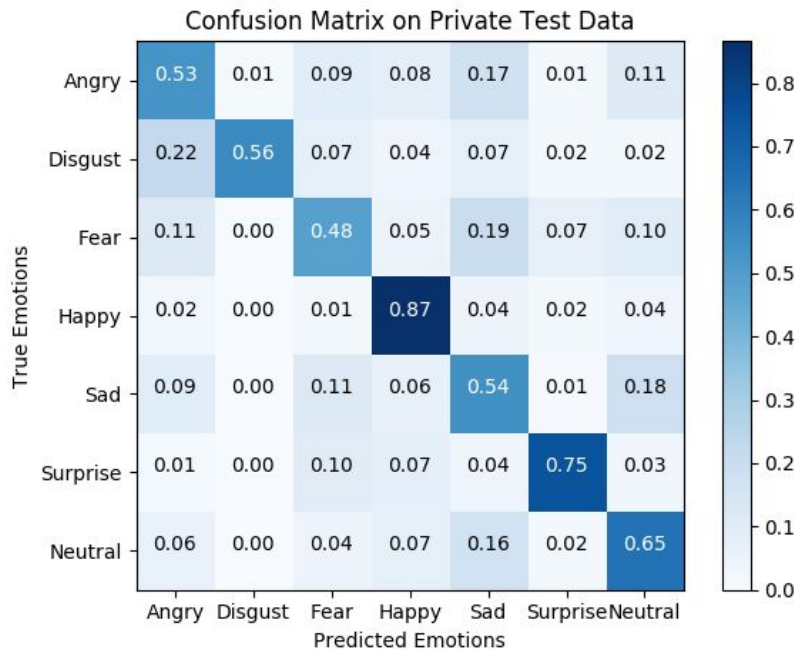# 4 Evaluation
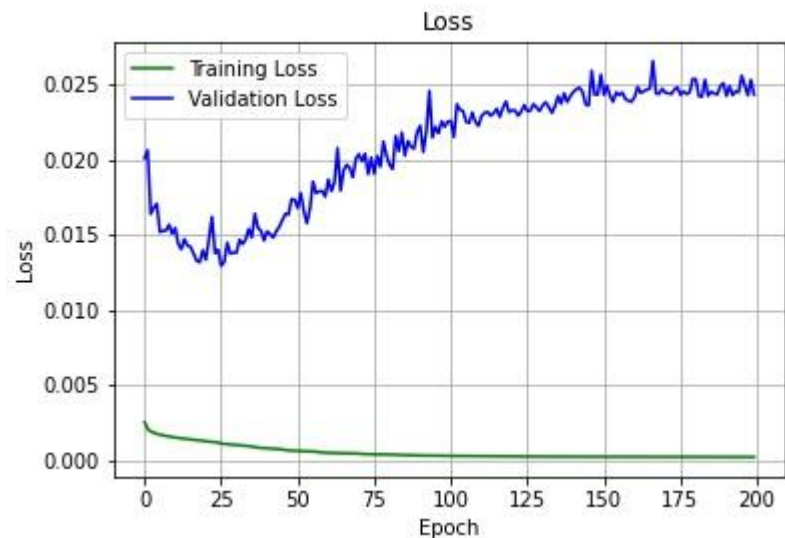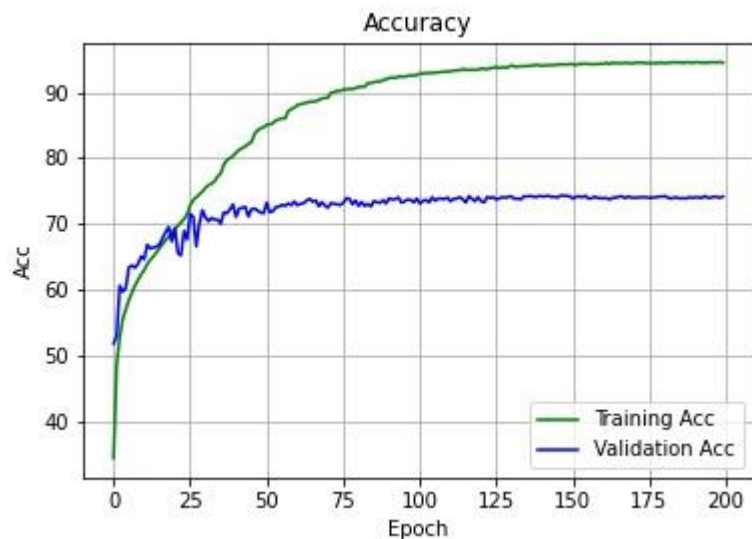
4 Evaluation on Fer2013

Figure 4.3: CM of method on FER2013

Angry→Fear   Angry→Disgust

Fear → Sad   Fear → Sad

Figure 4.4: Failure examples mainly due to wrongly labelling or emotional intensity

5 Conclusion & Demo

★ Limited resources:
   ○ The data is small and unevenly distributed.
   ○ Equipment (computer) is limited.
★ Low accuracy under natural conditions: head/face pose, light, …
★ Failure to recognize complex emotions.

THANK YOU