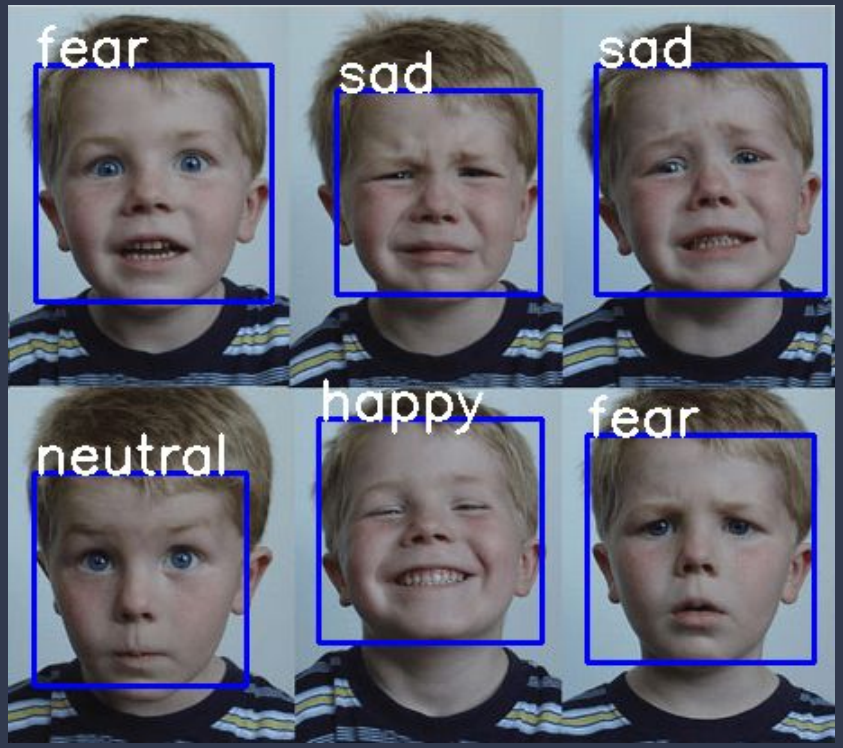
FERAtt: Facial Expression Recognition with Attention Net

Presentors: Haixuan Guo and Robert Johnson



Introduction

Numerous studies have been conducted on automatic facial expression analysis because of its practical importance in sociable robots, medical treatment, driver fatigue surveillance, and many other human-computer interaction systems. [1]

Researchers proposed 7 basic emotions that human can perceive in the same way regardless of culture. These emotions are anger, disgust, fear, happiness, sadness, surprise, and contempt. [1]

Deep Convolutional Neural Networks (CNN) have recently shown excellent performance in a wide variety of image classification tasks, as well as in facial expression recognition. [2]

Contempt

Sadness

Fear

Surprise

Happy

Anger

Disgust

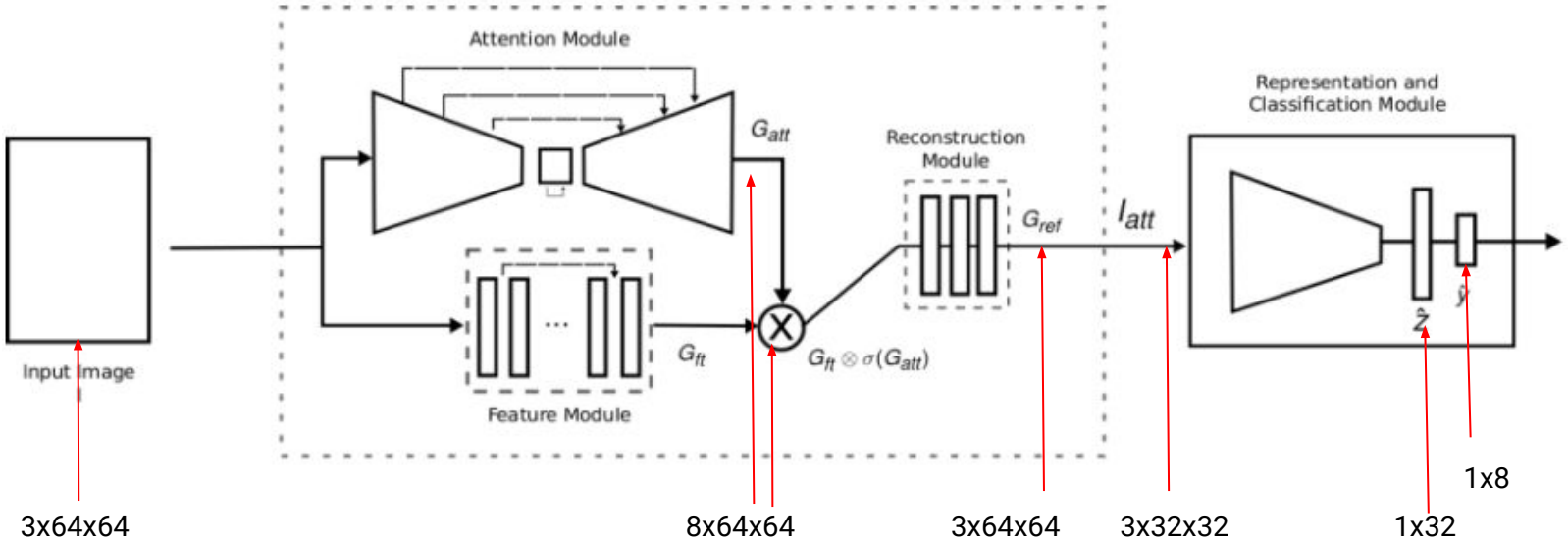
Current Approach

- Conventional FER Approaches[3]

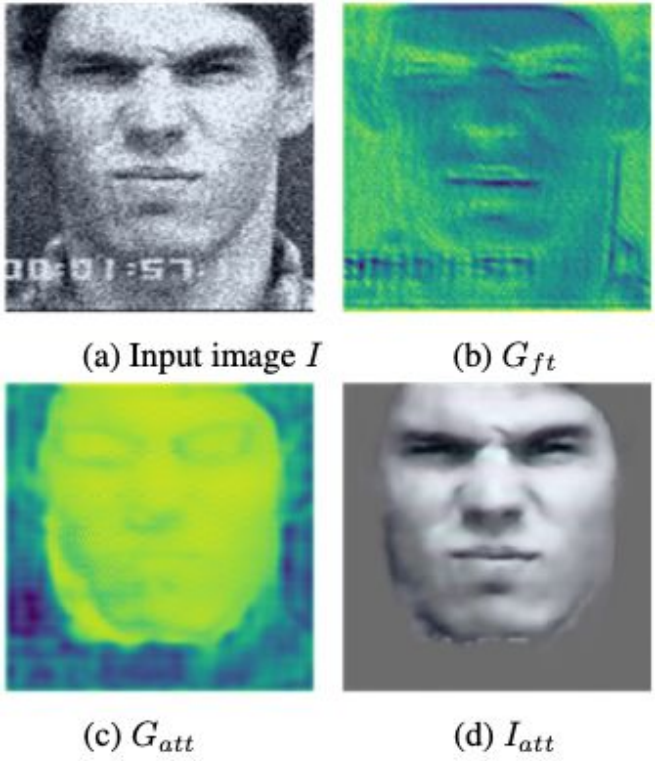
  Image preprocessing, feature extraction, and expression classification

  - Gabor Feature Extraction
  - Local Binary Pattern (LBP)
- Deep Learning-Based FER Approaches[3]
  - Convolutional neural network (CNN)
  - Deep autoencoder (DAE)

Proposed Model



What does the attention module do?



(a) Input image I (b) G_{ft}

(c) G_{att} (d) I_{att}

Objective Function

$$\min_{\Theta} \{\mathcal{L}_{att}(I_{att}, I \otimes I_{mask}) + \mathcal{L}_{rep}(\hat{z}, y) + \mathcal{L}_{cls}(\hat{y}, y)\}$$

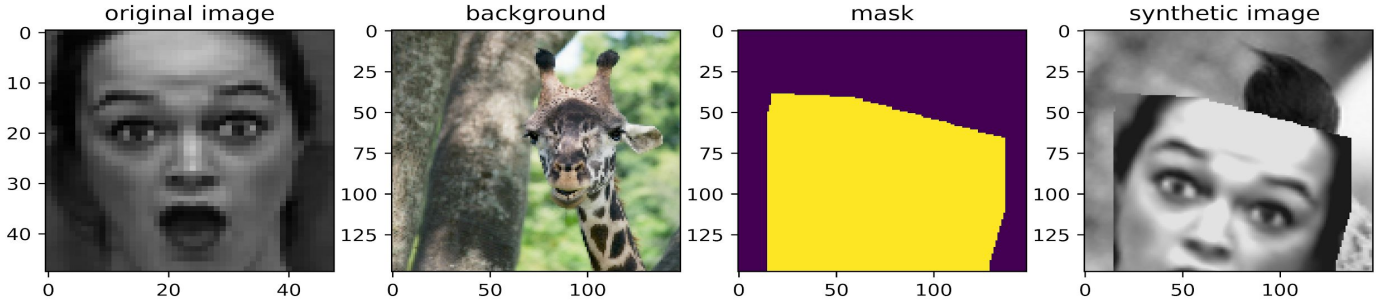
$$\mathcal{L}_{rep} = \mathbb{E}\left\{||P(w_j|f_{\Theta}(x_k)) - P(w_j|x_k)||_2^2\right\}$$

Generator - Synthetic Data

- What to do about the lack of data?
  - The CK dataset includes faces on blank backdrops
  - Pictures of people in the wild don’t usually have labels we can use
- The COCO dataset
  - “Common Objects in Context”
  - Mixed with CK to achieve a synthetic “People in Context” dataset

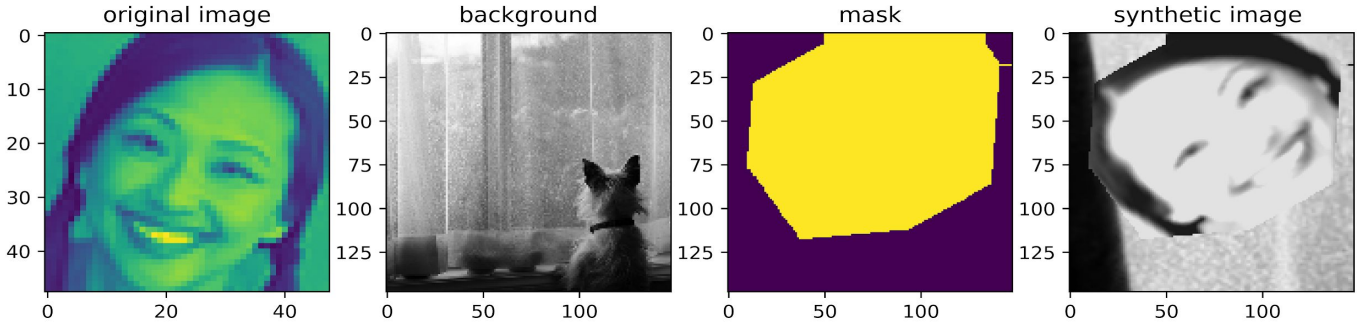
Synthetic images

CK+



Surprise

FERPlus



Happiness

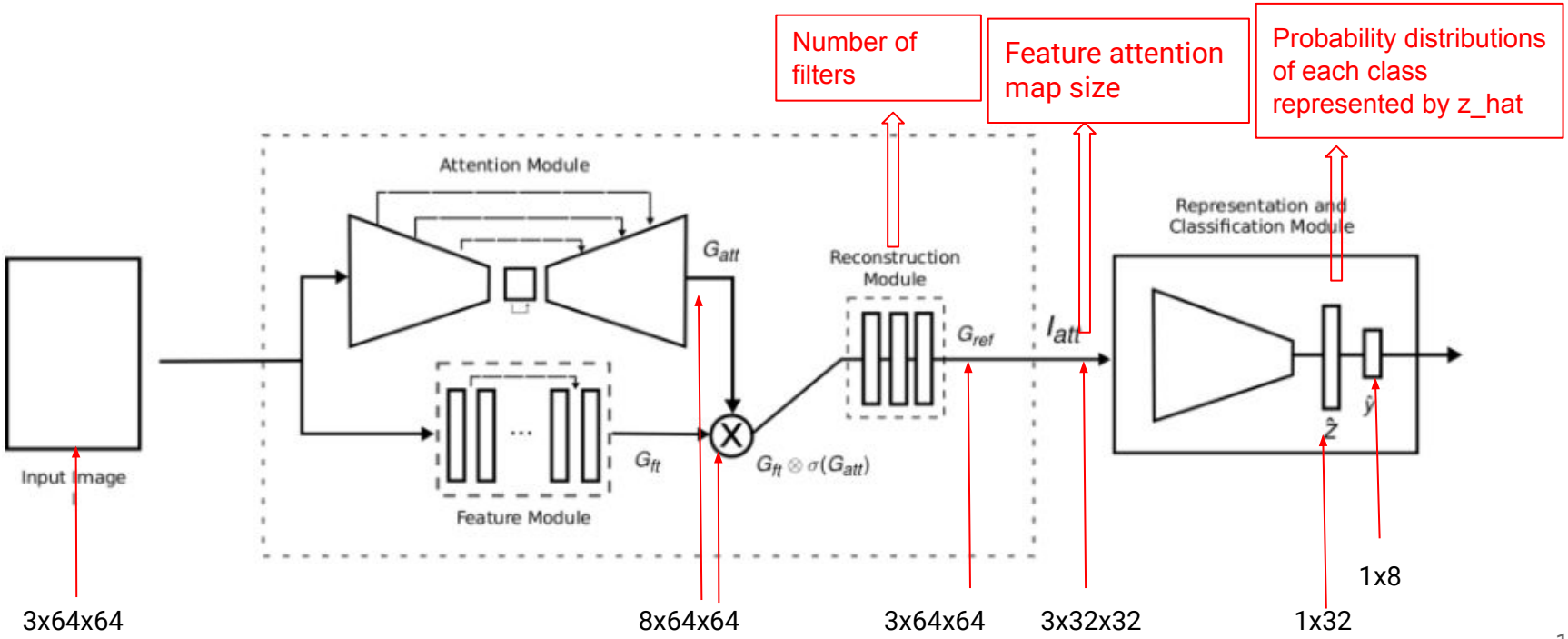
Shortcomings

- Small data, less than 1000
- Feature attention map size of 32 and number of filters in reconstruction module are borrowed from other literatures without experimenting with other values.
- The assumption that the probabilities for each class of images can be estimated by the proposed neural network, following a Gaussian-like beta distribution is not persuasive .

Improvement

- Larger dataset: FERPlus(total: 28236)
- Feature map size to be [4, 8, 16, 32(default)]
- Number of filters in reconstruction module to be [8, 16, 32(default), 64,128]
- Different distributions other than the Gaussian-like beta distribution in representation loss (Beta distribution)

Model – our changes



Dataset

Table1: Number of emotions in CK+ and FERPlus

Emotion	CK+	FERPlus
Angry	45	2098
Contempt	18	120
Disgust	58	116
Fear	25	536
Happy	69	7284
Sadness	28	3022
Surprise	82	3136
Neutral*	33	8733

Table2: Train and test size of CK+ and FERPlus

Dataset	CK+		FERPlus	
Type	real	synthetic	real	synthetic
Train size	323	1000	25045	30000
Test Size	35	2000	3191	5000

Experimental Results

- **PreActResNet18**: includes classification
- **FERAtt+Cls**: includes attention and classification
- **FERAtt+Rep+Cls**: includes attention, classification, and representation, and Gaussian Manifold Loss for training

CK+

- PreActResNet18 outperforms the other two models on CK+ real data
- Synthetic dataset performances better than real dataset
- FERAtt+Cls is better than FERAtt+Rep+Cls
- There is a significant difference between our results and the results in the paper

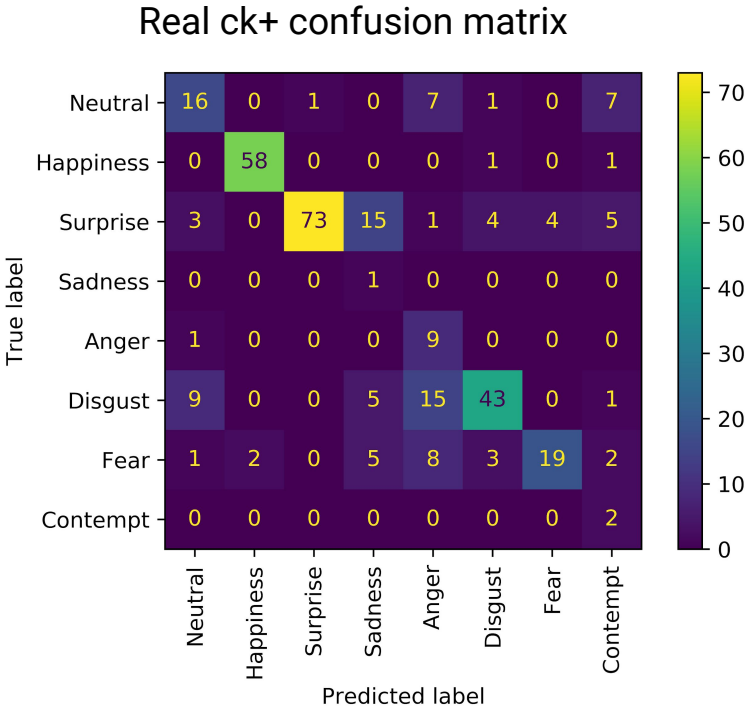
Dataset	CK+(real)				CK+(synthetic)			
Method	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
PreActResNet18	62.90%	51.80%	53.50%	52.60%	54.20%	37.90%	34.30%	36.00%
FERAtt+Cls	25.00%	3.60%	14.30%	5.70%	54.00%	52.20%	39.90%	45.20%
FERAtt+Rep+Cls	11.10%	1.60%	14.30%	2.90%	40.20%	40.60%	30.90%	35.10%

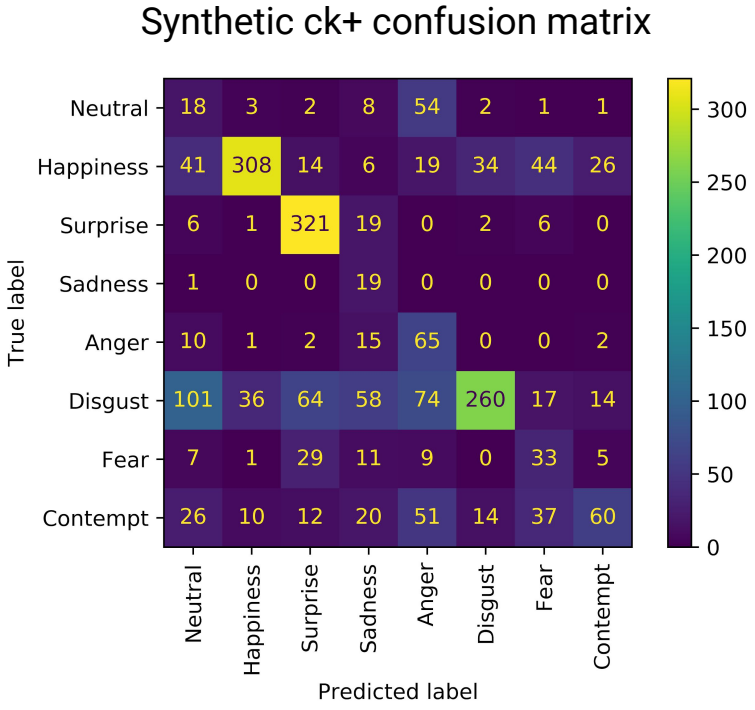
FERPlus

- PreActResNet18 outperforms the other two models
- Synthetic dataset performs worse than real dataset
- FERAtt+Cls is better than FERAtt+Rep+Cls
- The overall performance of model on FERPlus is better than on CK+

Dataset	FERPlus(real)				FERPlus(synthetic)			
Method	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
PreActResNet18	79.20%	57.30%	51.20%	54.10%	61.20%	40.40%	35.00%	37.50%
FERAtt+Cls	79.10%	74.60%	58.20%	65.40%	59.40%	40.70%	31.80%	35.70%
FERAtt+Rep+Cls	77.20%	56.10%	58.80%	57.40%	56.40%	43.20%	25.00%	31.70%

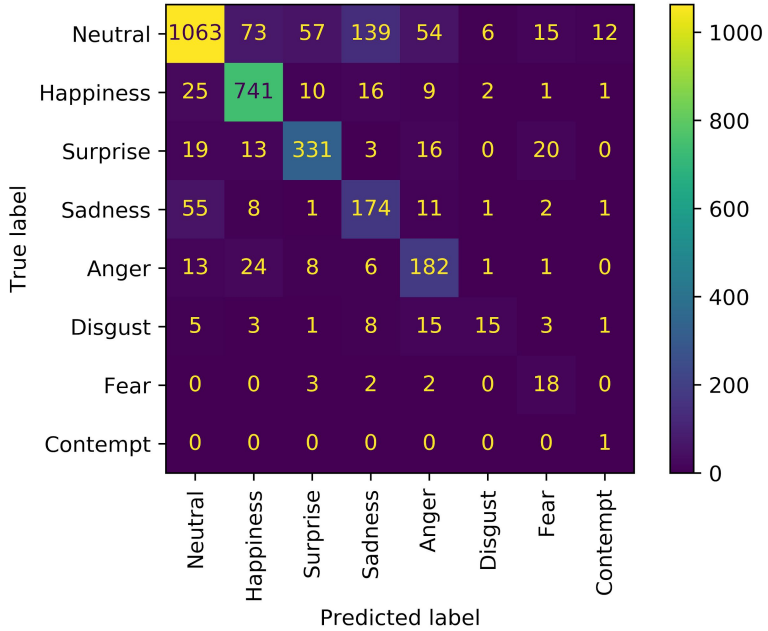
Confusion Matrix on CK+



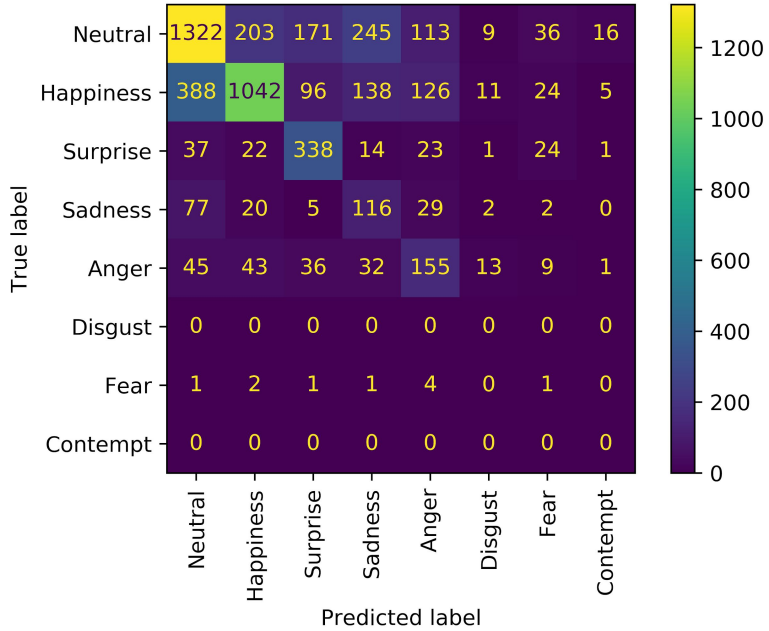


Confusion Matrix on FERPlus

Real FERPlus confusion matrix

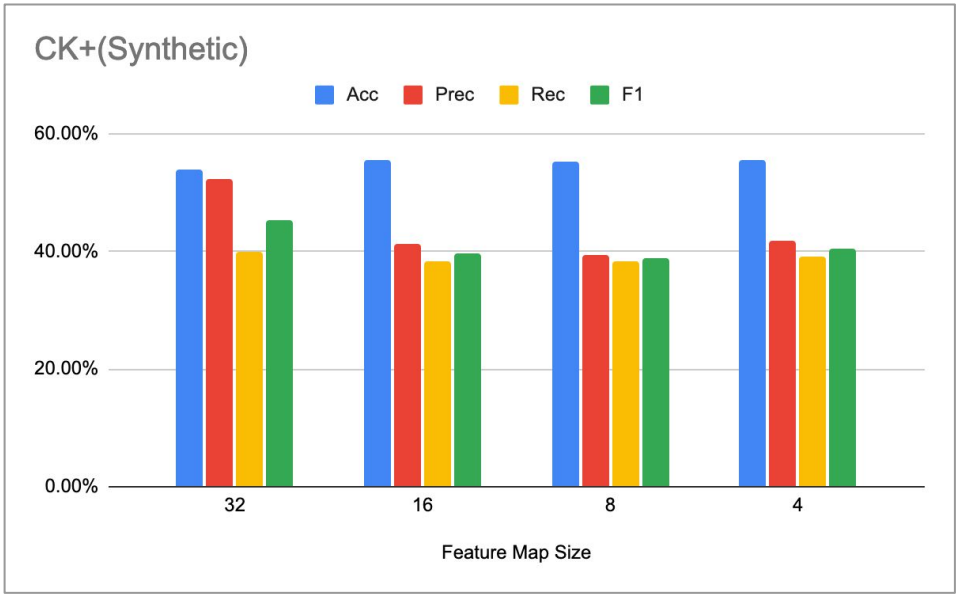


Synthetic FERPlus confusion matrix



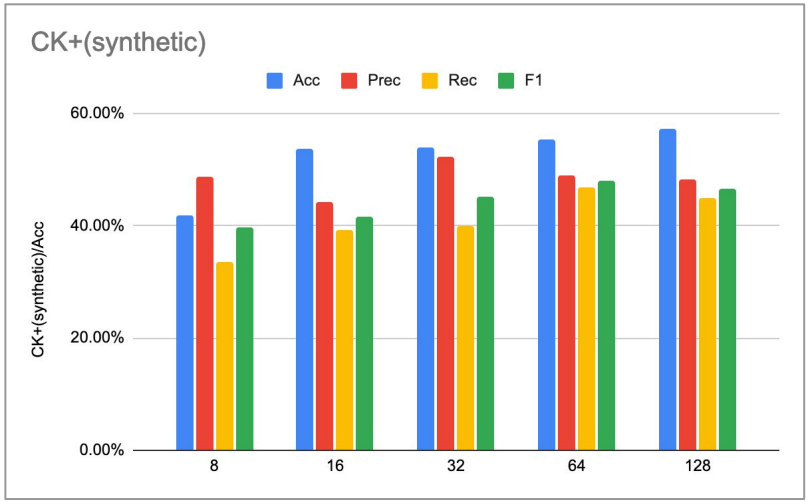
Feature Map Size

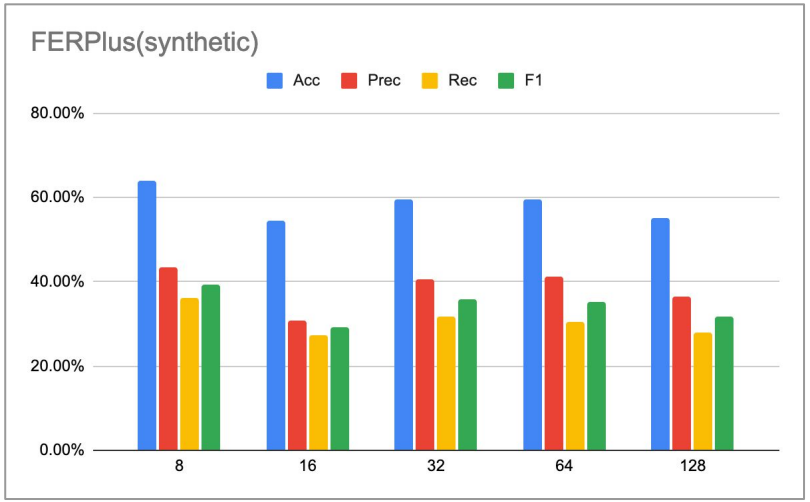
Feature map size = 32 has highest precision and f1 score compared with other options.



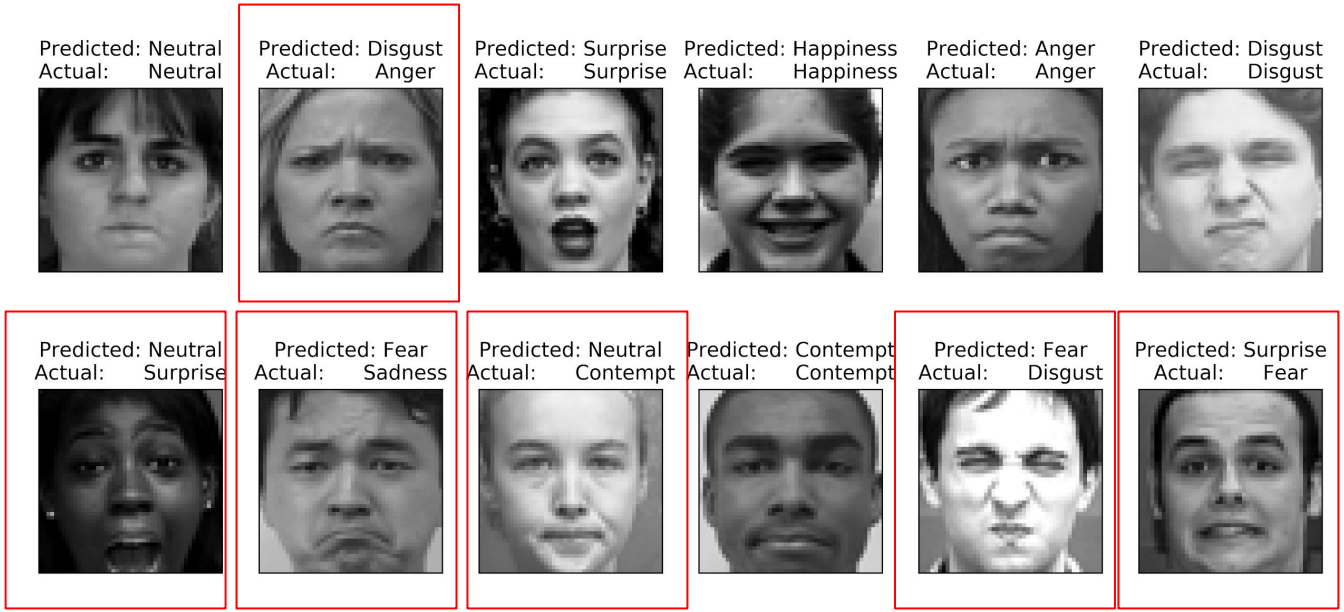
Number of Filters in Reconstruction Module

- Setup: FERAtt+Cls, synthetic dataset
- For CK+, increasing the number of filters leads to better performance
- For FERPlus, number of filters has trivial influence on prediction results

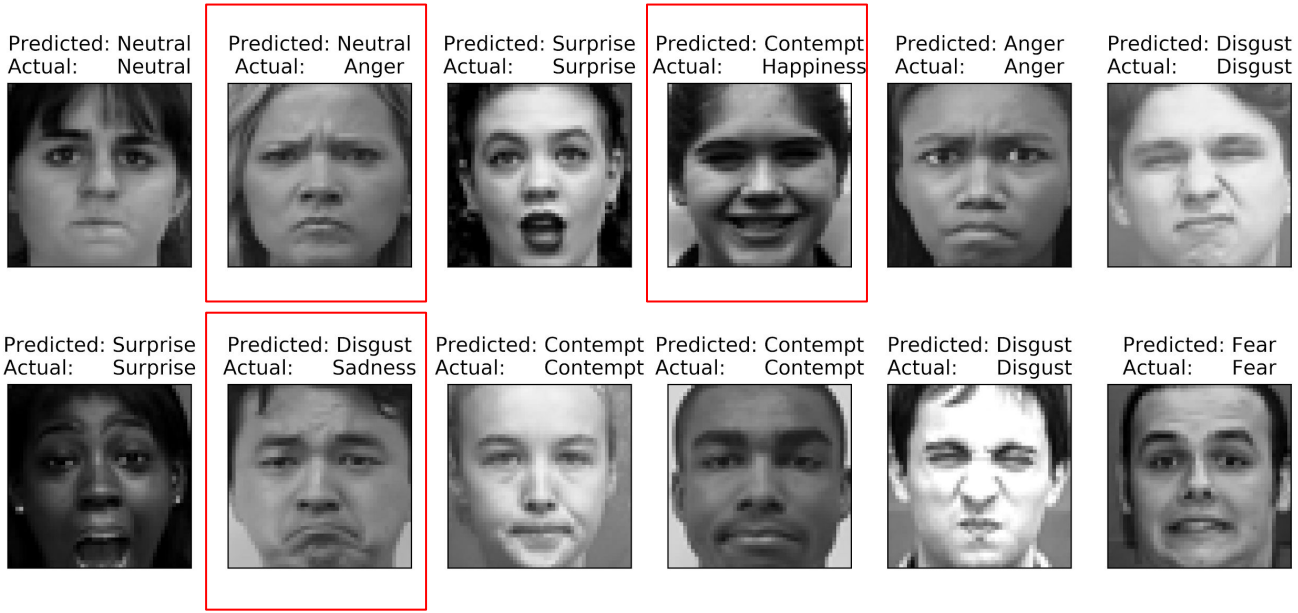




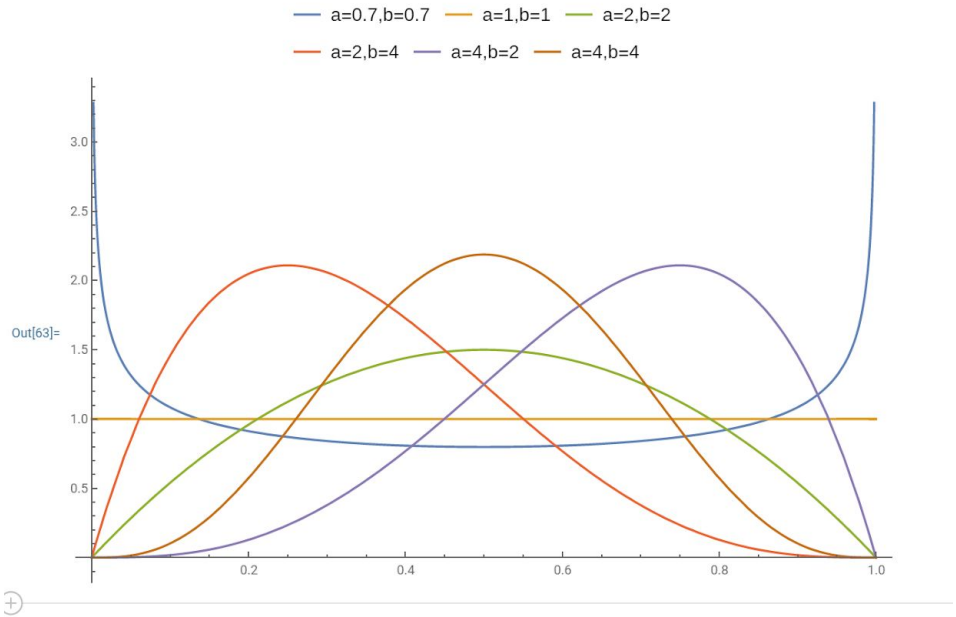
FERAtt+cls


Predicted: Neutral
Actual: Neutral
Predicted: Disgust
Actual: Anger
Predicted: Surprise
Actual: Surprise
Predicted: Happiness
Actual: Happiness
Predicted: Anger
Actual: Anger
Predicted: Disgust
Actual: Disgust
Predicted: Neutral
Actual: Surprise
Predicted: Fear
Actual: Sadness
Predicted: Neutral
Actual: Contempt
Predicted: Contempt
Actual: Contempt
Predicted: Fear
Actual: Disgust
Predicted: Surprise
Actual: Fear

Improved FERAtt+cls



The beta distribution



Different Distributions

- To test the distribution, we ran 10 epochs of each distribution five times. The minimum of each epoch was recorded along with the epoch number. The loss of an untrained model was about 4.939.

Data taken over 5 runs	$\alpha = 2, \beta = 4$	$\alpha = 4, \beta = 2$	$\alpha = 1, \beta = 1$	$\alpha = .7, \beta = .7$	$\alpha = 4, \beta = 4$
Average loss	4.751	4.771	4.703	4.735	4.925
Average lowest epoch	4.2	4	4.6	4.2	3.6
Minimum loss	4.694	4.700	4.599	4.578	4.857
Minimum loss epoch	4	4	4	5	8

Conclusion

- The proposed model has better performance on Large datasets, such as FERPlus, when compared with small dataset(CK+)
- In small datasets, with the number of filters in reconstruction module increasing, the accuracy and f1-score of proposed model improves.
- Synthetic data can boosts the performance on real facial data. The accuracy in CK+ boosts to 86.10%, increased by 28.80%. In FERPlus, the accuracy is 81.60%(17.5% up).
- The regularization of Gaussian manifold loss is not effective

Future Work

- The synthetic datasets are a step on the way to recognizing emotion out in the wild, but actual pictures of people in their surroundings could be used.
- Different kinds of pooling layers could be tried in reconstruction module .

Reference

[1]Deep Facial Expression Recognition: A Survey

[2]FERAtt: Facial Expression Recognition with Attention Net

[3]Deeply Learning Deformable Facial Action Parts Model for Dynamic Expression Analysis