Facial Expression Recognition

By Rita Samir and Sohayla Mohammed

Intro

- Visual non verbal mean to understand human emotion.
- Various applications in different domains.
- Models use
  - Facial Action units
  - Valence-Arousal space

V-A space

- Identify emotion categories according to the value of the emotion dimensions
- Valence is positive or negative affectivity.
- Arousal measures how calming or exciting the information is.

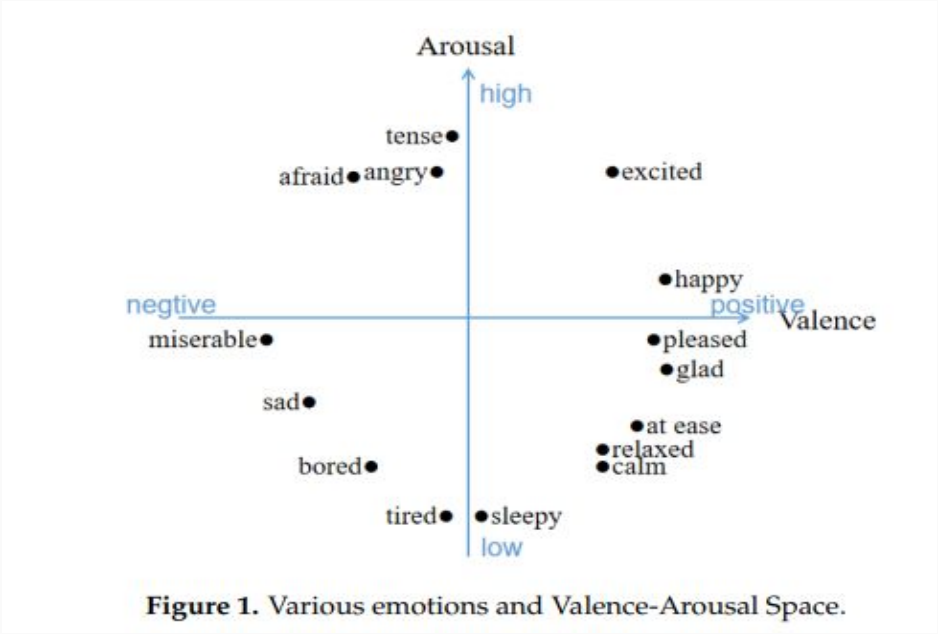
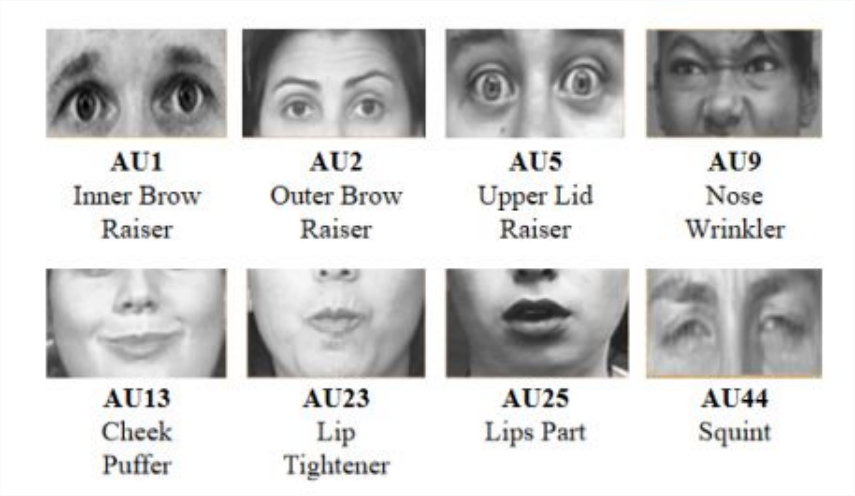


Figure 1. Various emotions and Valence-Arousal Space.

AU units

- The 46 facial action units encode the basic movements of individual or groups of muscles.
- The FER system classifies expression categories by inspecting the combinations of the detected face AUs.



FACs

Category	AUs	Category	AUs
Happy	12,25	Sadly disgusted	4,10
Sad	4,15	Fearfully angry	4,20,25
Fearful	1,4,20,25	Fearfully surprised	1,2,5,20,25
Angry	4,7,24	Fearfully disgusted	1,4,10,20,25
Surprised	1,2,25,26	Angrily disgusted	4,25,26
Disgusted	9,10,17	Disgusted surprised	1,2,5,10
Happily sad	4,6,12,25	Happily fearfully	1,2,12,25,26
Happily surprised	1,2,12,25	Angrily disgusted	4,10,17
Happily disgusted	10,12,25	Awed	1,2,5,25
Sadly fearful	1,4,15,25	Appalled	4,9,10
Sadly angry	4,7,15	Hatred	4,7,10
Sadly surprised	1,4,25,26		

The FACS system classifies many human expressions in real life, and is the definitive reference standard for muscle movements in facial expressions today.

Related Work

Related Work

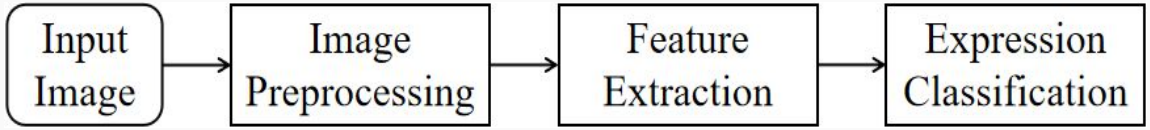
Studies can be divided into two groups according to whether the features are manually extracted or generated through the output of neural networks,

- The conventional FER approaches
- The deep learning-based FER approaches.

The Conventional FER approaches

It is highly dependent on manual feature engineering.

The researchers need to preprocess the image and select the appropriate feature extraction and classification method for the target dataset.



1. **Image Preprocessing** : Noise reduction, Face detection, ...etc
2. **Feature Extraction :** Feature Point Tracking and Haar-like Feature Extraction

3. **Expression Classification :** kNN (k-Nearest Neighbours), SVM (Support Vector Machine), Adaboost (Adaptive Boosting) and etc.

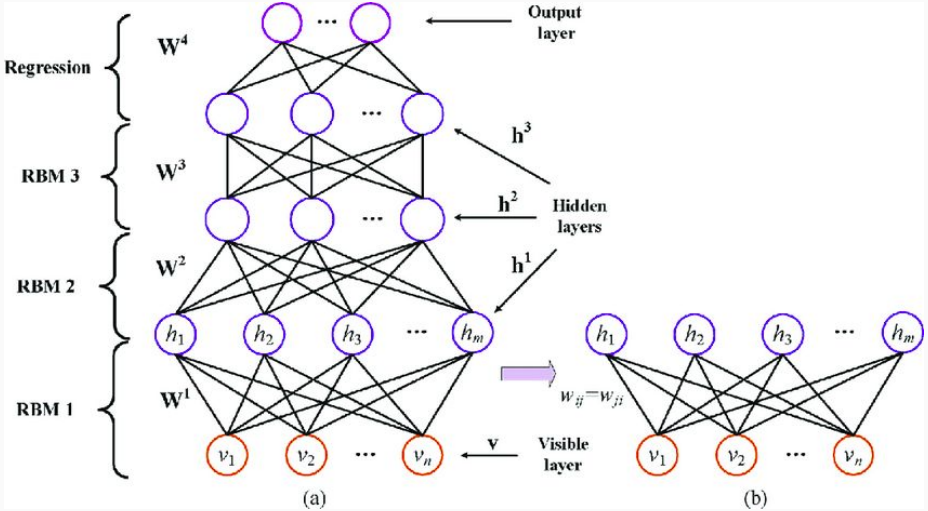
The Deep-learning based FER approaches

In FER, 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

1. Convolutional Neural Network (CNN)
2. Deep Belief Network (DBN)
3. Long Short-Term Memory (LSTM)
4. Generative Adversarial Network(GAN)

The Deep-learning based FER approaches

Deep Belief Network (DBN)

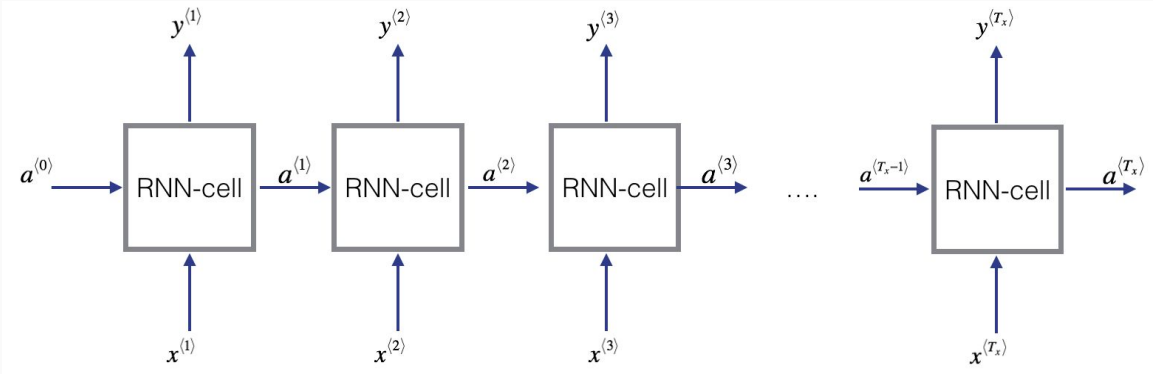


- When trained on a set of examples without supervision, a DBN can learn to probabilistically reconstruct its inputs
- Combined with other components, DBN has proved to be an effective FER approach(DBN+MLP).

The Deep-learning based FER approaches

Long Short-Term Memory (LSTM)

- Well suited for the temporal features extraction of consecutive frames.



The Deep-learning based FER approaches

Generative Adversarial Network(GAN)

- GAN is an unsupervised learning model composed of a generative network and a discriminative network.
- Well suited for posed images.
- The generator frontalises a frontal face image and the discriminator is trained to distinguish and recognise.
- 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

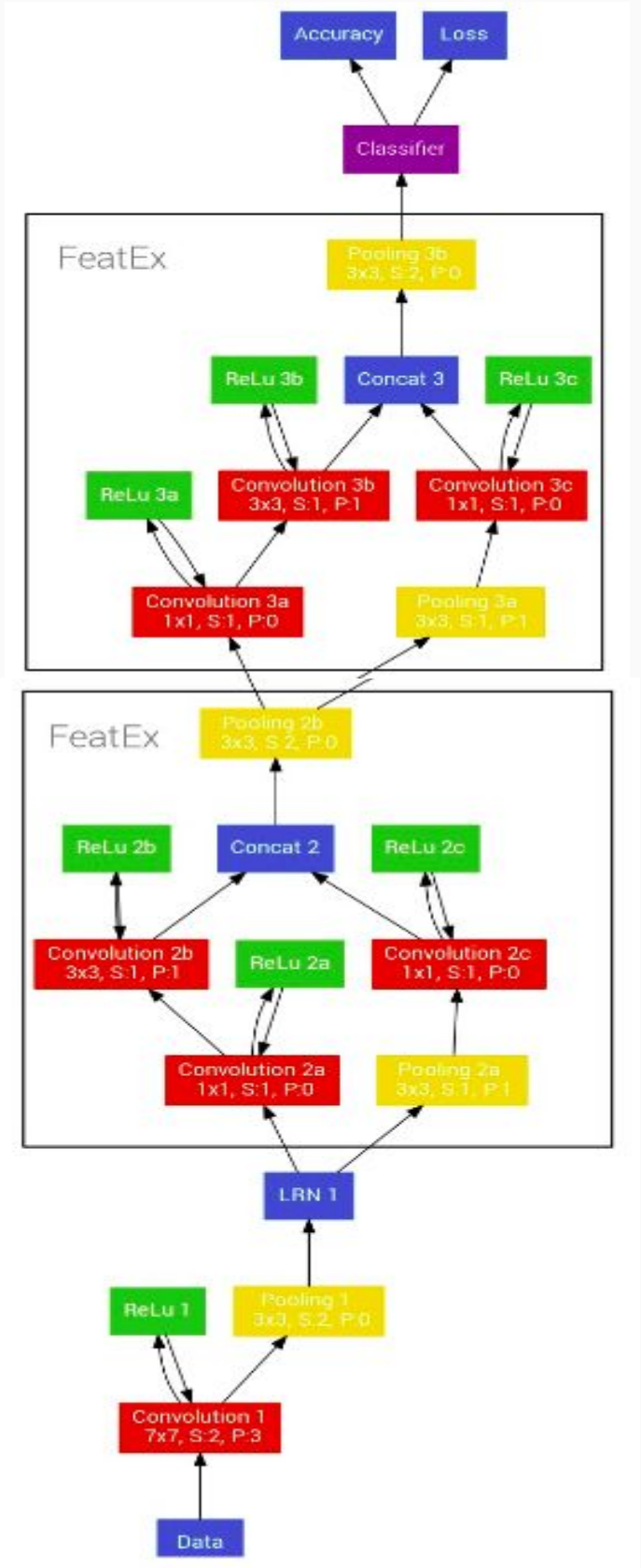
Types of emotions

- Two Different types of emotions are represented in FER datasets.
- Basic Emotions
  - Six basic human emotions, i.e., happiness, surprise, sadness, anger, disgust, and fear.
  - Some datasets add a seventh emotion (i.e neutral, contempt…).
- Compound Emotions
  - Combinations of two basic emotions.

Datasets used

- Ck
  - 48x48 resolution.
  - Unique
  - Posed
  - 6 BEs & 1 Neutral.
- FER2013
  - 48x48 resolution.
  - Wild
  - Posed and spontaneous.
  - 6 BEs & 1 Neutral.

Proposed model


Accuracy
Loss
Classifier
FeatEx
Pooling 3b
3x3, S:2, P:0
ReLu 3b
Concat 3
ReLu 3c
ReLu 3a
Convolution 3b
3x3, S:1, P:1
Convolution 3c
1x1, S:1, P:0
Convolution 3a
1x1, S:1, P:0
Pooling 3a
3x3, S:1, P:1
FeatEx
Pooling 2b
3x3, S:2, P:0
ReLu 2b
Concat 2
ReLu 2c
Convolution 2b
3x3, S:1, P:1
ReLu 2a
Convolution 2c
1x1, S:1, P:0
Convolution 2a
1x1, S:1, P:0
Pooling 2a
3x3, S:1, P:1
LRN 1
ReLu 1
Pooling 1
3x3, S:2, P:0
Convolution 1
7x7, S:2, P:3
Data

DeXpression

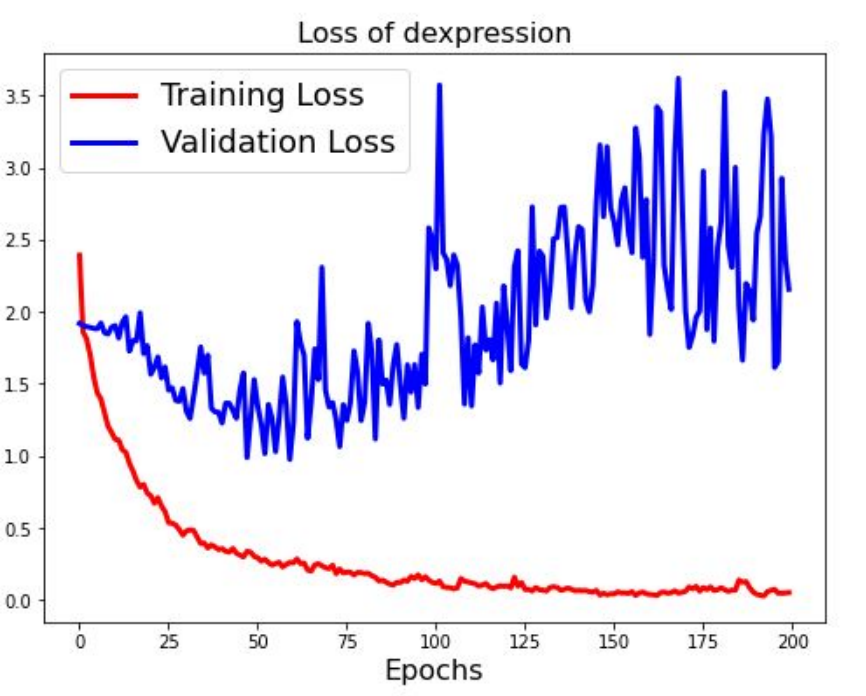
- The proposed deep Convolutional Neural Network architecture consists of four parts
    - The first part automatically preprocesses the data (Convolution, Pooling and LRN)
    - The next steps are the two FeatEx (Parallel Feature Extraction Block) blocks.
    - The features extracted by these blocks are forwarded to a fully connected layer, which uses them to classify the input into the different emotions.

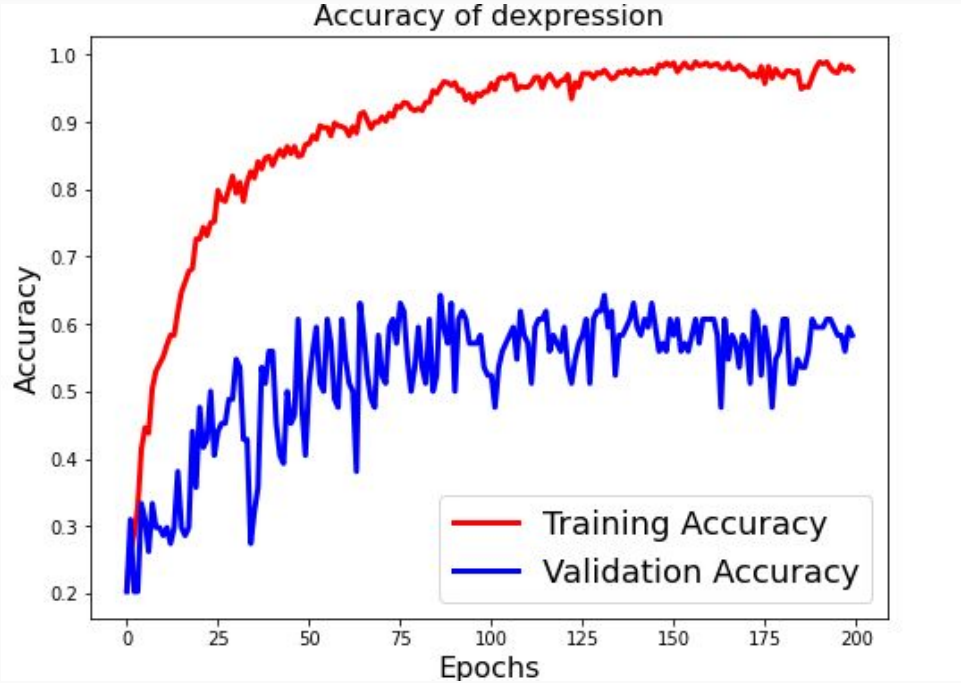
Deploying the original model

First Steps

- Problem with reading CK+ dataset, so we reimplemented the code using keras to have a better grasp on loading the dataset.
- The accuracy obtained when running the CK+ dataset was ~ 62%, we tried fine tuning till we can get a better accuracy or one that was close to the published accuracy in the paper.
- After some parameters fine tuning, the training accuracy was in the [94-97]% range, but huge overfitting occured

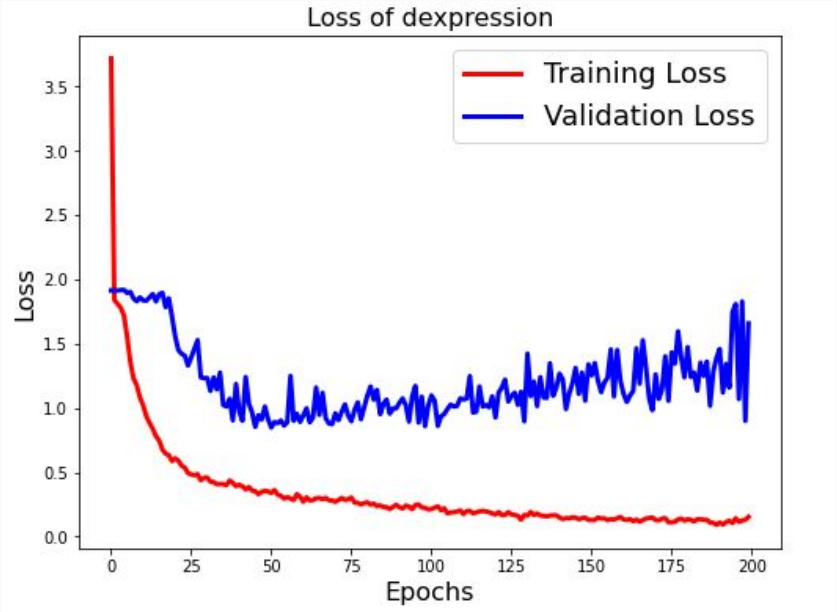
First Steps

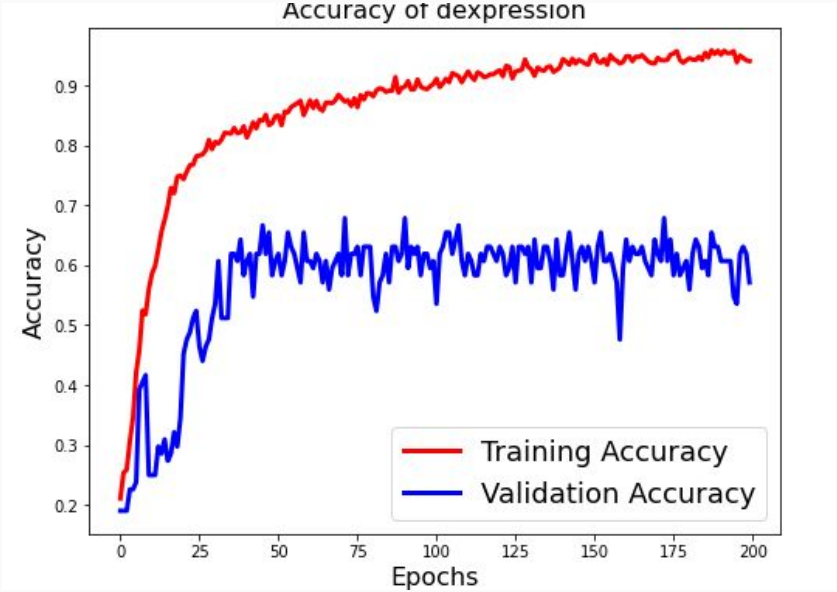

Loss of dexpression
Training Loss
Validation Loss
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
25
50
75
100
125
150
175
200
Epochs


Accuracy of dexpression
Accuracy
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
25
50
75
100
125
150
175
200
Epochs
Training Accuracy
Validation Accuracy

After that

- To overcome overfitting, we added data augmentation which didn't help much :



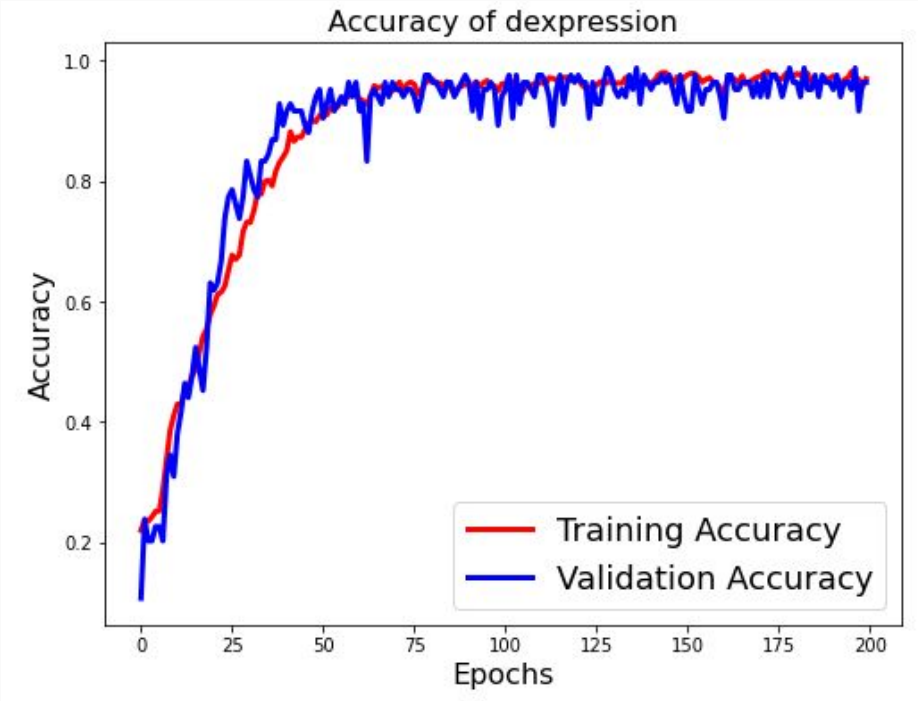


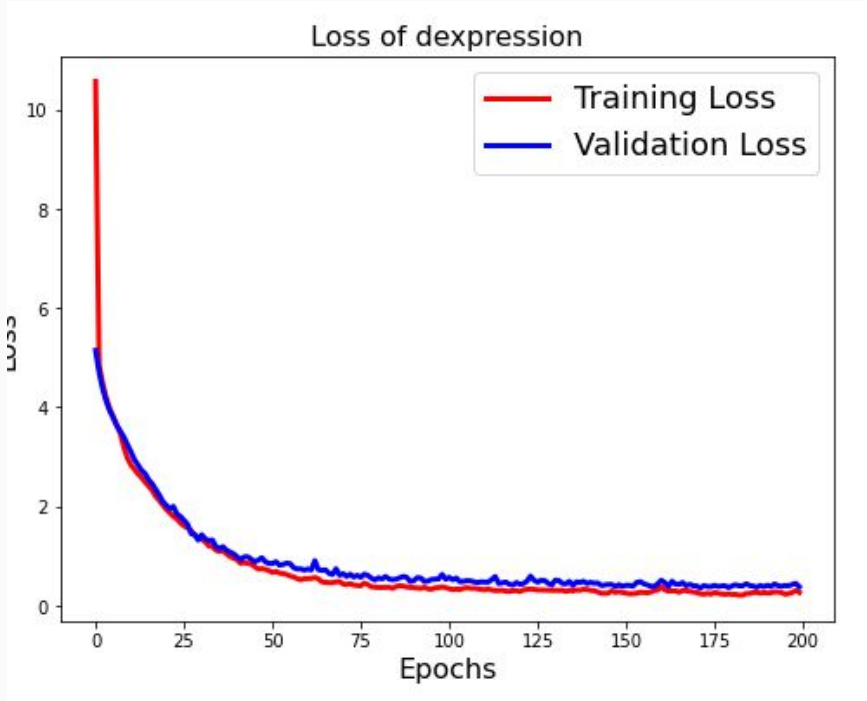
Finally

- We then added to the architecture different regualizers (drop out with different probability in each layer, L2 regularizer), changed strides values in the convolution layers. All that combined with data augmentation, produced better results :

```
Testing Loss = 0.33681220943863327
Testing Accuracy = 0.972973
```

Finally


Accuracy of dexpression
Accuracy
Epochs
Training Accuracy
Validation Accuracy
1.0
0.8
0.6
0.4
0.2
0
25
50
75
100
125
150
175
200


Loss of dexpression
Epochs
Training Loss
Validation Loss
10
8
6
4
2
0
0
25
50
75
100
125
150
175
200

Next steps

Next steps

- 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

- Enhancing the accuracy results with FER2013 dataset by trying and tuning different architectures.

Timeline

Tasks	Due Date
Running the FER2013 dataset on the proposed model	31st March 2020
Running the FER2013 dataset on another arch01(ensemble CNN)	10th April 2020
Adding pose invariance to arch01.(arch02)	22nd April 2020
Milestone ||	23rd April 2020
Tuning our chosen model to get highest accuracy with it	14th May 2020

Online resources/papers

Online resources/papers

- Facial Expression Recognition: A Survey Yunxin Huang
- Deep Facial Expression Recognition: A Survey
- Article of Facial Expression Recognition: A Survey
- Facial Action Coding System (FACS) - A Visual Guidebook
- Training Deep Networks for Facial Expression Recognition with Crowd-Sourced Label Distribution

Online resources/papers

- MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Face Images
- Suppressing Uncertainties for Large-Scale Facial Expression Recognition
- Facial expression databases

Thank you ^ ^.