Facial Expression Recognition

Jiangang Chen

Facial Expression Recognition

- Emotion recognition has attracted a major attention in numerous fields
  - Marketing, Psychology, Surveillance, and Entertainment are some examples

- Facial Expression Recognition (FER) has broad applications in multiple domains
  - Human–Computer Interaction
  - Virtual Reality
  - Augmented Reality
  - Advanced Driver Assistance Systems
  - Education
  - Entertainment

Goal

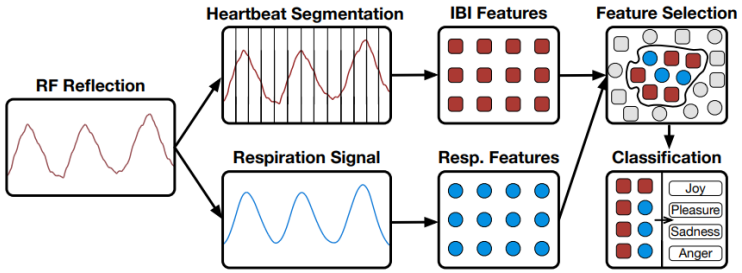
- How we achieve the Facial Expression Recognition
- Introducing Conventional & DL methods
- Defects & Future Works

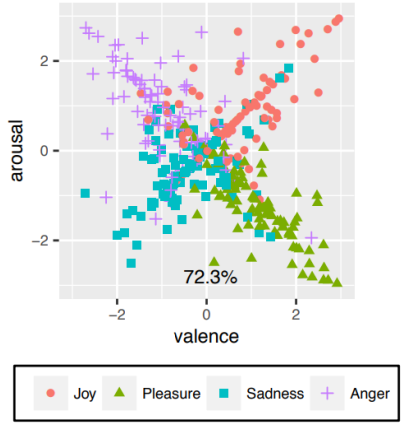
Why is FER important?

- Traditional interaction methods are difficult to achieve natural and harmonious emotional interaction
  - (Human-Computer Interaction) Such as keyboard, mouse, screen, and pattern, which is far from meeting the requirements for artificial intelligence
- Human expression is the most important carrier of inspirational perception and the most direct and obvious way of expressing emotions
- Facial expression is arguably the most natural, powerful and immediate signal to communicate emotional states and intentions
  - Thus, FER has important theoretical significance for improving the emotional interaction ability of computers

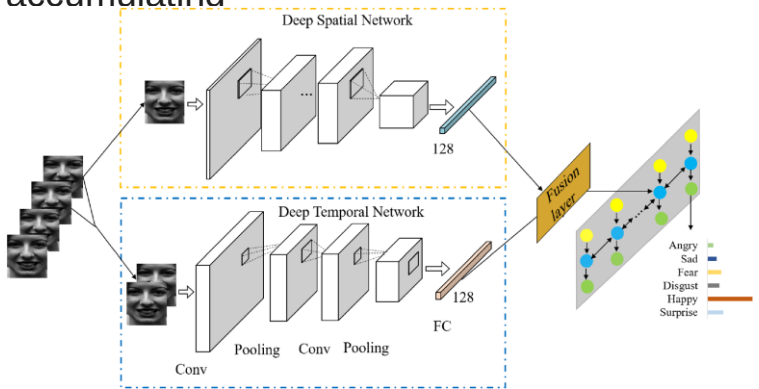
State-of-the-Art Approaches

- RF Signals
- Thermals Measurement



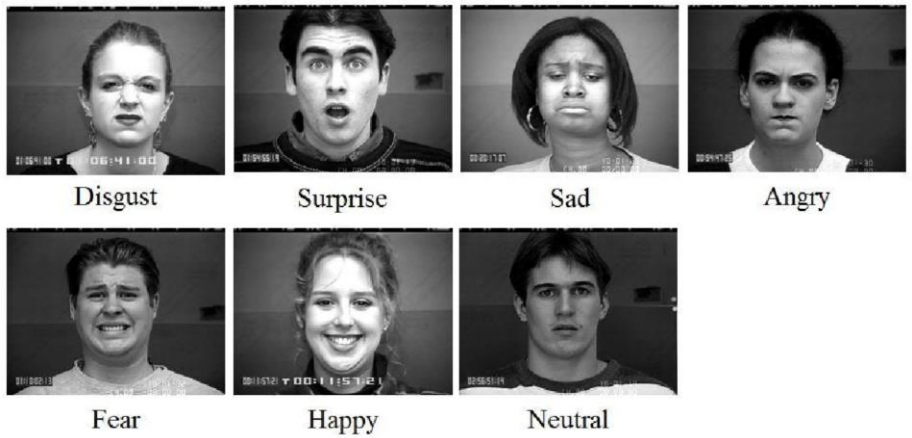


- CV
  - Expression-Specific LBP[1]: better capture the local information of faces on important fiducial points
  - Work [2] presents a new deep neural network with the addition of inception layers which increases the depth and width of the network while keeping the computational budget constant.
  - Paper [3] instead of using the whole face region, three kinds of active regions are applied to classify facial expression through a decision-level fusion strategy
  - Work [4] proposes a deep convolutional fusion network, which addresses the FER task through discriminative spatial features learning and temporal dependencies accumulating



Types of Emotions

- Basic Emotions [1]
  - Happiness, surprise, sadness, anger, disgust, and fear
- Compound Emotions [2]
  - Including 7 basic emotions (6 basic emotions and 1 neutral), 12 compound emotions expressed commonly by humans, and 3 additional emotions (Awed, Appalled, and Hatred).
- Micro Expressions [3]
  - Reveal the true and potential expressions of a person
  - Lasts for only 1/25 to 1/3 s
  - Psychology and Police Investigations



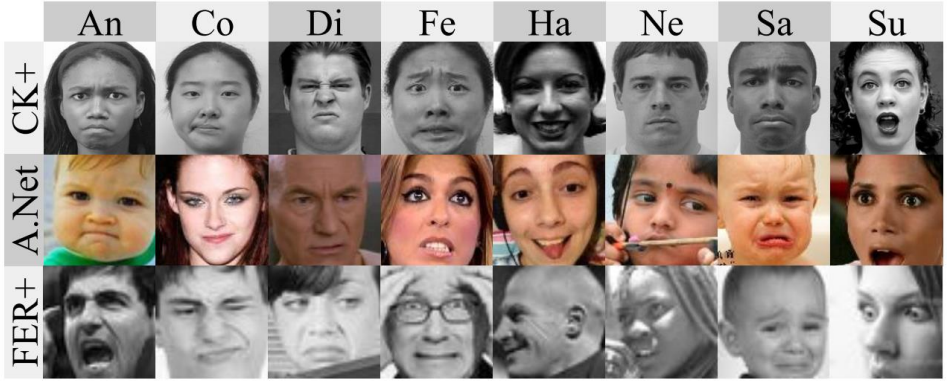
[1] Ekman, Paul. "An argument for basic emotions." *Cognition & emotion* 6.3-4 (1992): 169-200.

[2] Du, Shichuan, Yong Tao, and Aleix M. Martinez. "Compound facial expressions of emotion." *Proceedings of the National Academy of Sciences* 111.15 (2014): E1454-E1462.

[3] Ekman, Paul. "Darwin, deception, and facial expression." *Annals of the new York Academy of sciences* 1000.1 (2003): 205-221.

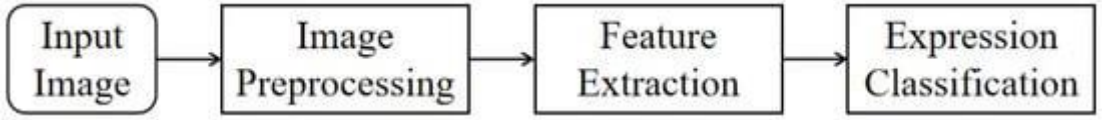
FER Datasets

- In-the-Lab Dataset
  - Extended Cohn–Kanade Dataset (CK+)
    - Recorded from frontal camera-view points
    - Annotated and validated in terms of 30 action units (AUs)
- In-the-Wild Dataset
  - AffectNet
    - The largest dataset of facial expressions in the wild
    - Heavily imbalanced label distribution and strong baselines
  - FER+
    - Created by querying facial images from Google's image search engine using 184 emotion keywords



Conventional Methods

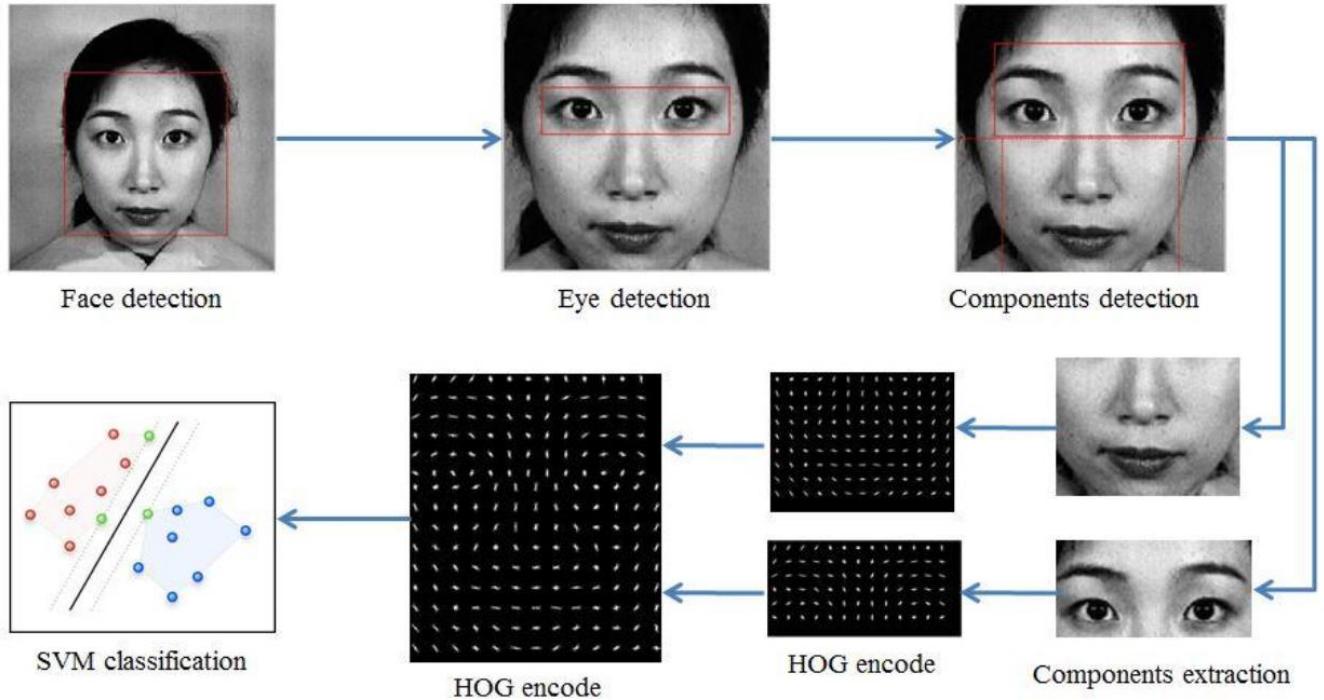
- Notable characteristic: Manual Feature Engineering
- The conventional FER methods can be divided into three major steps



- Image Preprocessing
  - Noise reduction, Face detection, Normalization of the scale and grayscale, Histogram equalization
- Feature Extraction
  - Gabor Feature Extraction | Local Binary Pattern | The Active Shape Model | Optical Flow | Haar Feature Extraction | Feature Point Tracking
- Expression Classification
  - KNN | SVM | Naive Bayes | AdaBoost (with decision trees) | SRC | PNN

Example

- Face detection: Viola-Jones face detector
  - Searching for these Haar-like features
- Encode face component: HOG
  - Sensitive to object deformations
- Classifier: SVM



Conventional Methods

- Accuracy
  - CK+: 88.7%

	AN	CO	DI	FE	HA	SA	SU
AN	0.84	0.04	0.07	0.00	0.02	0.00	0.02
CO	0.06	0.61	0.00	0.11	0.11	0.11	0.00
DI	0.02	0.00	0.95	0.00	0.03	0.00	0.00
FE	0.08	0.04	0.00	0.72	0.12	0.00	0.04
HA	0.01	0.03	0.00	0.00	0.96	0.00	0.00
SA	0.07	0.04	0.00	0.04	0.00	0.82	0.04
SU	0.00	0.01	0.00	0.00	0.00	0.00	0.99

- Dlib + HOG + SVM
  - Dlib: 68 points face landmarks
  - FER+

Features	7 Emotions	5 Emotions
HOG Features	29.0%	34.4%
Face Landmarks	39.2%	46.9%
Face Landmarks + HOG	48.2%	55.0%
Face Landmarks + HOG on sliding window	50.5%	59.4%

Deep Neural Network

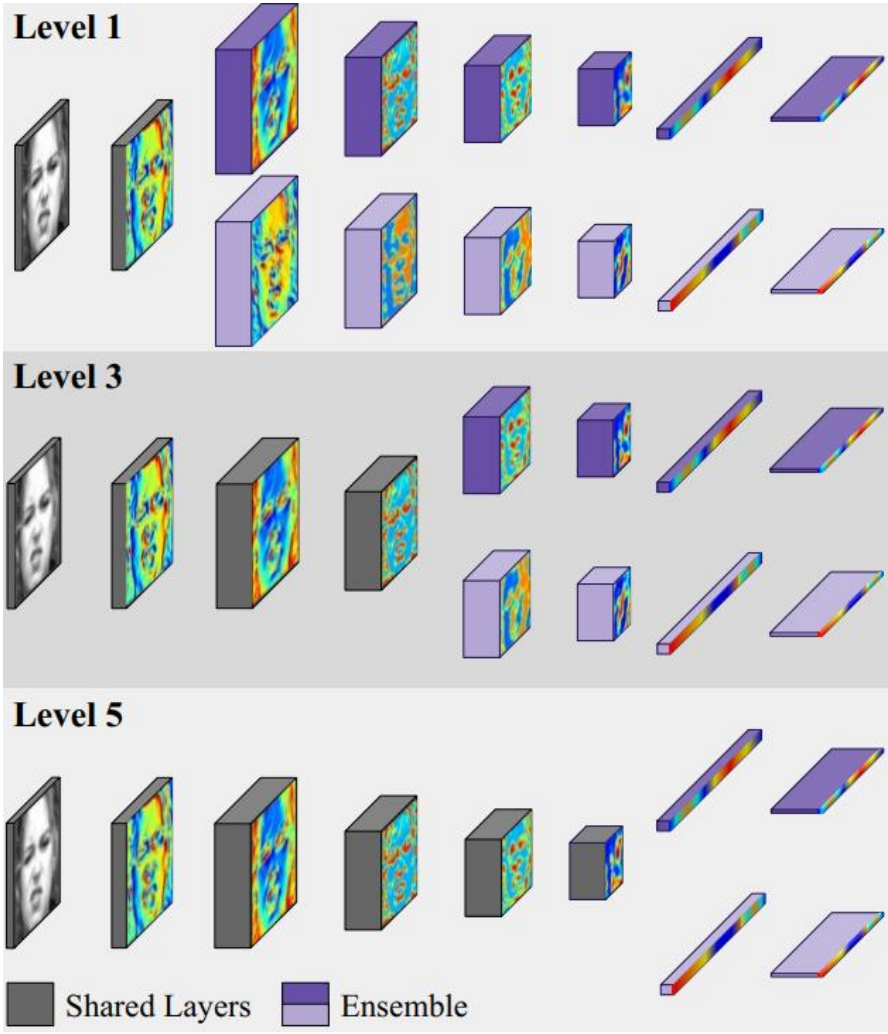
- Why Neural Network?
  - Automatically learn features from data, hand-feature engineering was left out
  - Can learn a broader range of facial features
  - Previously learned can be transferred among related tasks

- Ensemble methods
  - A set of models where an inference is made collectively based on individual predictions
  - Reduce the remaining residual generalization error
  - More accurate than any single model in the ensemble

- Requires high computational power

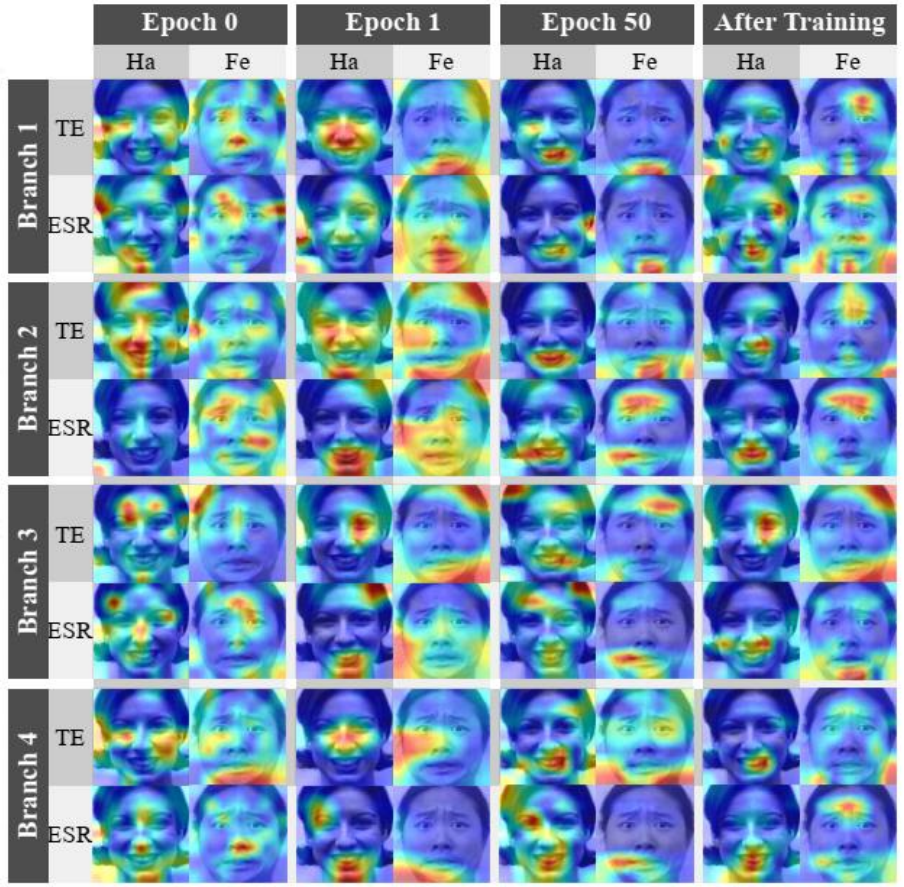
Ensembles with Shared Representations

- ESR consists of two building blocks (Implicit)
- Base of the network
  - Convolutional layers for low- and middle-level features
  - Reduction of redundancy, training, and inference time
  - Low level features learned are shared with the ensemble
- Independent convolutional branches (Explicit)
  - Constitutes the ensemble
  - Carries the diversity of the ensemble.
- The level to start the ensemble of branches
  - Too early: high redundancy of low-level facial features
  - Too late: decrease diversity in the ensemble



Training at Different Branch Level

- TE needs more training epochs than ESR to learn informative facial features
  - (Branch 2) ESR learned after the first update that the region around the mouth is relevant for recognizing happy facial expressions
  - TE took around 50 epochs to discover the same pattern.



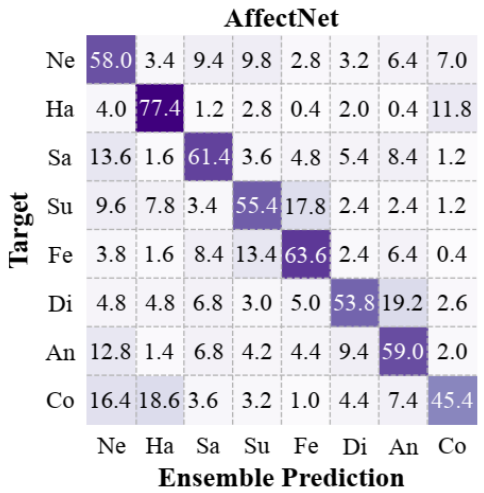
- Diversity Analysis
- Facial Action Coding System
  - Happy
    - AU-6
    - AU-12
  - Fear
    - AU-1
    - AU-20

Result

- CK+
  - Less parameters to train

Approach	#	Accuracy
Single Network	**131.208**	$85.5 \pm 3.5\%$
Traditional Ensemble	524.832	$89.2 \pm 1.2\%$
ESR-4 Lvl. 3	355.104	$\mathbf{89.4 \pm 2.2\%}$
ESR-4 Lvl. 4	243.936	$88.5 \pm 3.8\%$

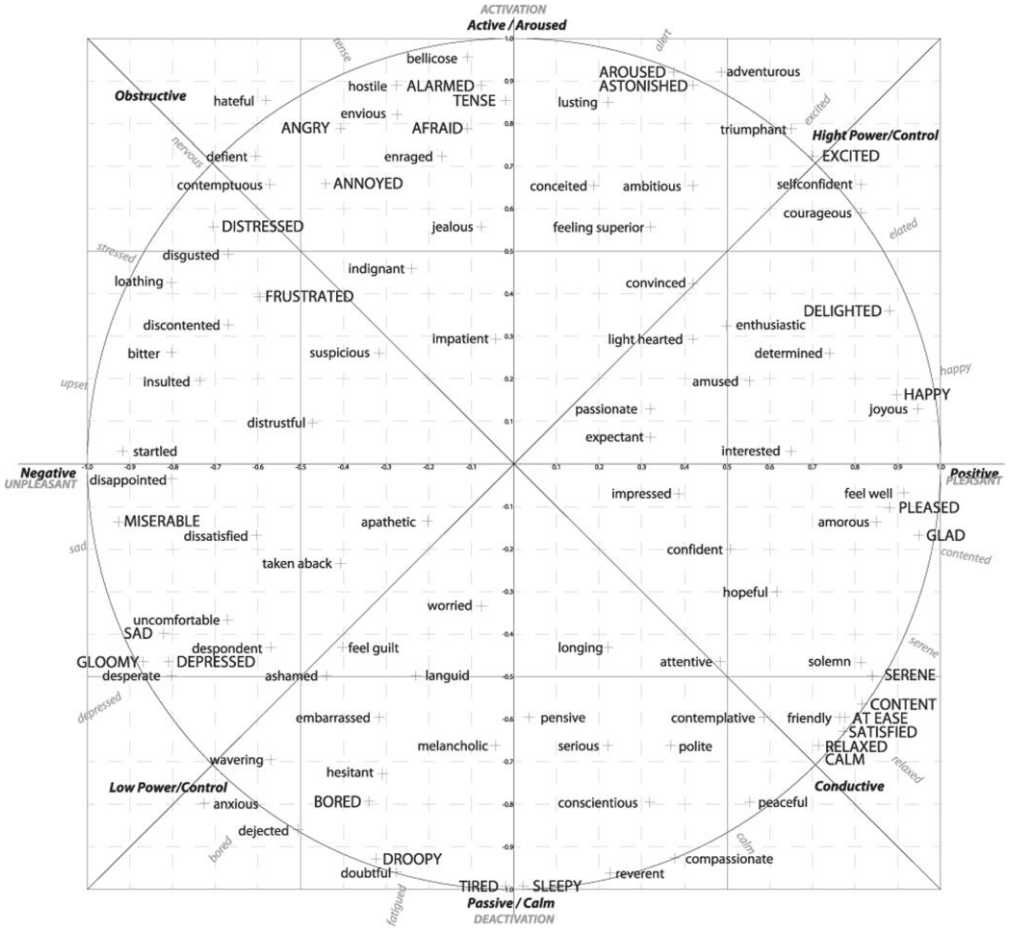
- AffectNet: 59.3%



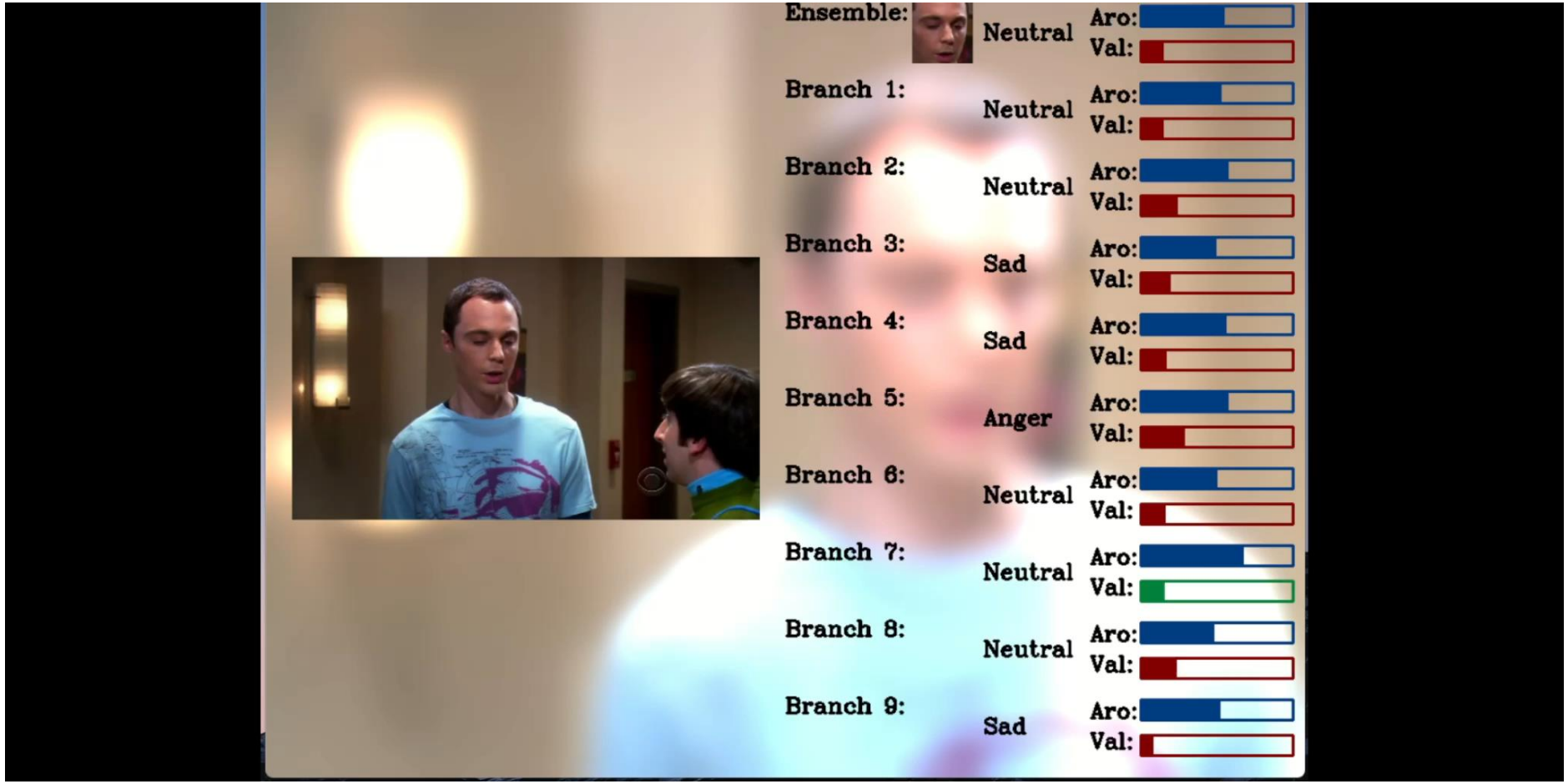
- FER+: 87.15%

Demo

- Valence–Arousal space



Demo


Ensemble:
Neutral
Aro:
Val:
Branch 1:
Neutral
Aro:
Val:
Branch 2:
Neutral
Aro:
Val:
Branch 3:
Sad
Aro:
Val:
Branch 4:
Sad
Aro:
Val:
Branch 5:
Anger
Aro:
Val:
Branch 6:
Neutral
Aro:
Val:
Branch 7:
Neutral
Aro:
Val:
Branch 8:
Neutral
Aro:
Val:
Branch 9:
Sad
Aro:
Val:

Defects & Future Works

- Wild Environmental Conditions
  - Occlusion and pose-variation
- The Pressure of High-Volume Data Processing

- Multi-Modal Affect Recognition
  - Including sound, Face Depth Map, IR
  - Vision Privacy

Reference

- [1] Chao, Wei-Lun, Jian-Jiun Ding, and Jun-Zuo Liu. "Facial expression recognition based on improved local binary pattern and class-regularized locality preserving projection." *Signal Processing* 117 (2015): 1-10.
- [2] A. Mollahosseini, D. Chan and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1-10, doi: 10.1109/WACV.2016.7477450.
- [3] Sun, Ai, et al. "Facial expression recognition using optimized active regions." *Human-centric Computing and Information Sciences* 8.1 (2018): 1-24.
- [4] Liang, Dandan, et al. "Deep convolutional BiLSTM fusion network for facial expression recognition." *The Visual Computer* 36.3 (2020): 499-508.