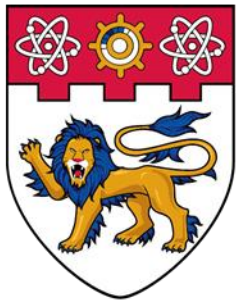
# Multiple Transfer Learning and Multi-label Balanced Training Strategies for Facial AU Detection In the Wild

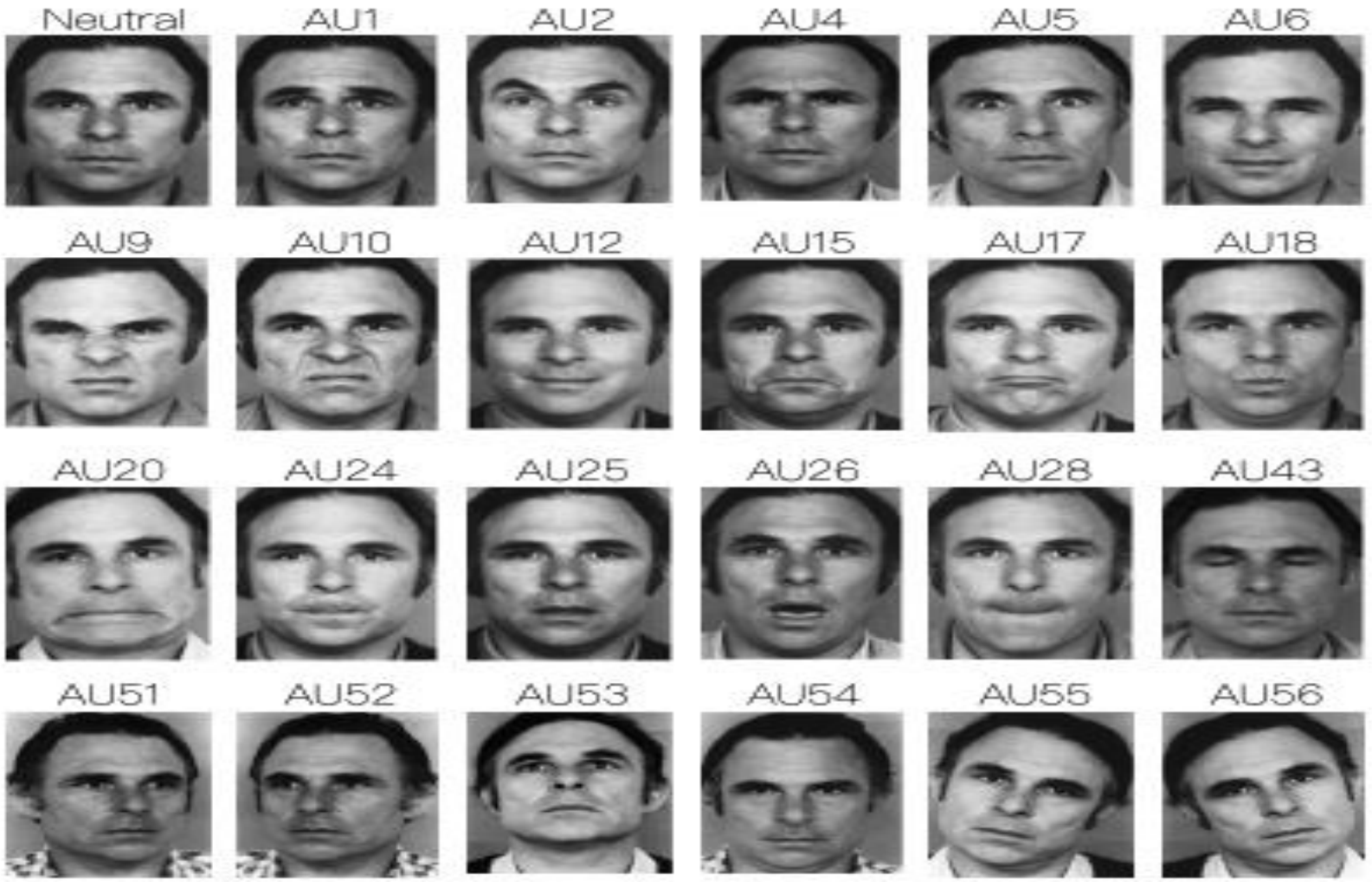
Sijie Ji, Kai Wang, Xiaojiang Peng, Jianfei Yang, Zhaoyang Zeng, and Yu Qiao


中国科学院深圳先进技术研究院
SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES



NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

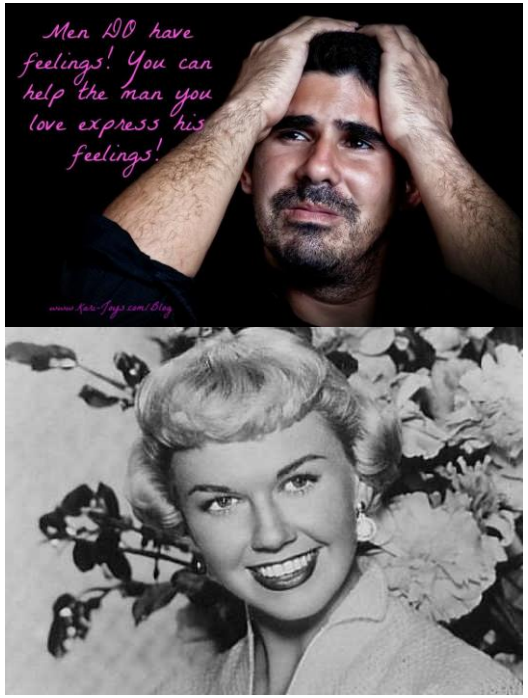
# Facial Action Units(AU)




Picture from: http://cbcsl.ece.ohio-state.edu/enc-2020/index.html

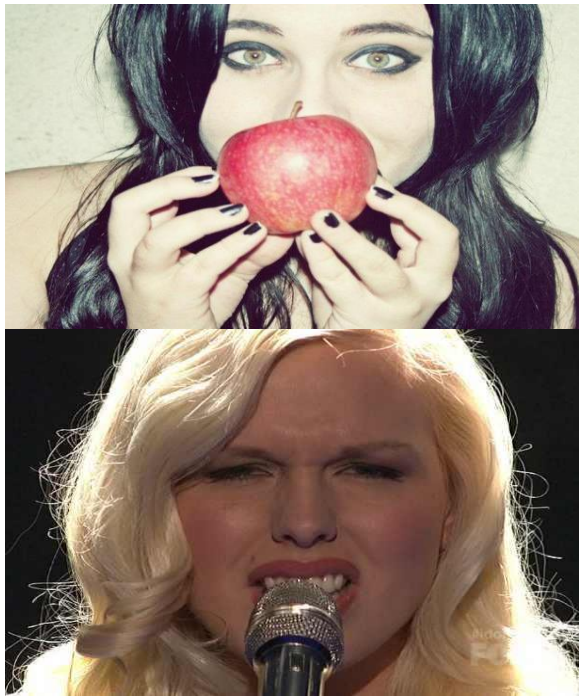
# Challenge: unconstrained heterogeneity of in the wild facial images


Various pose




Various illumination

Various resolution



Different kind of occlusion

Picture from Emotinet Dataset

# Solution: jointly face detection and alignment

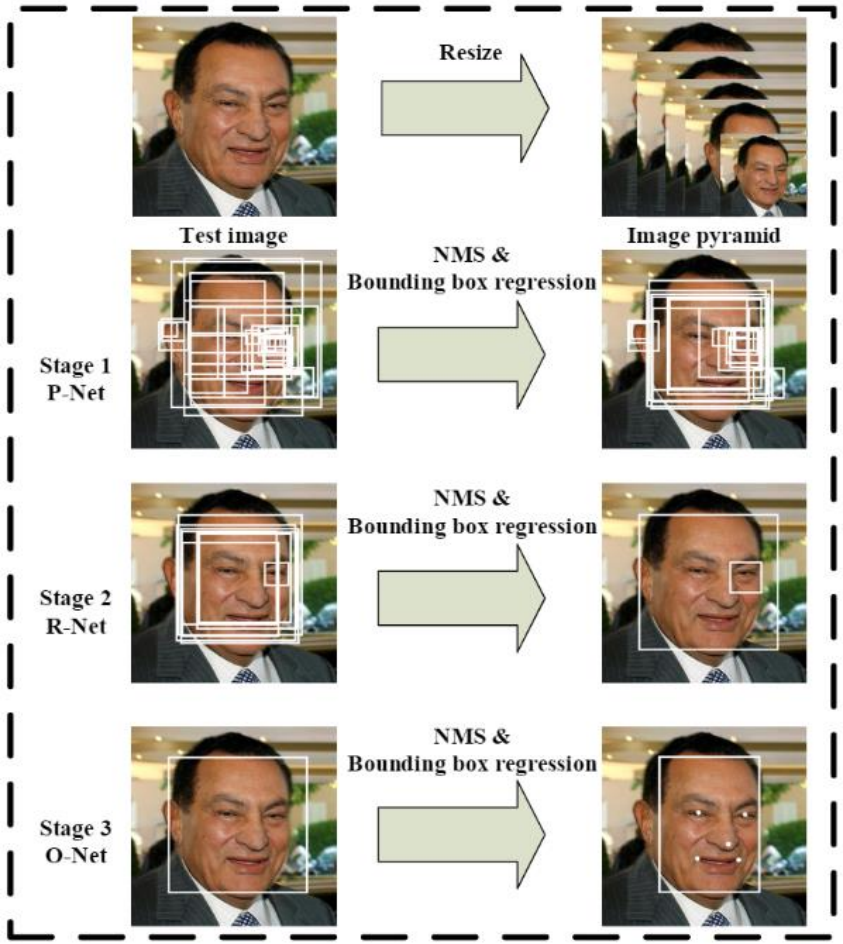
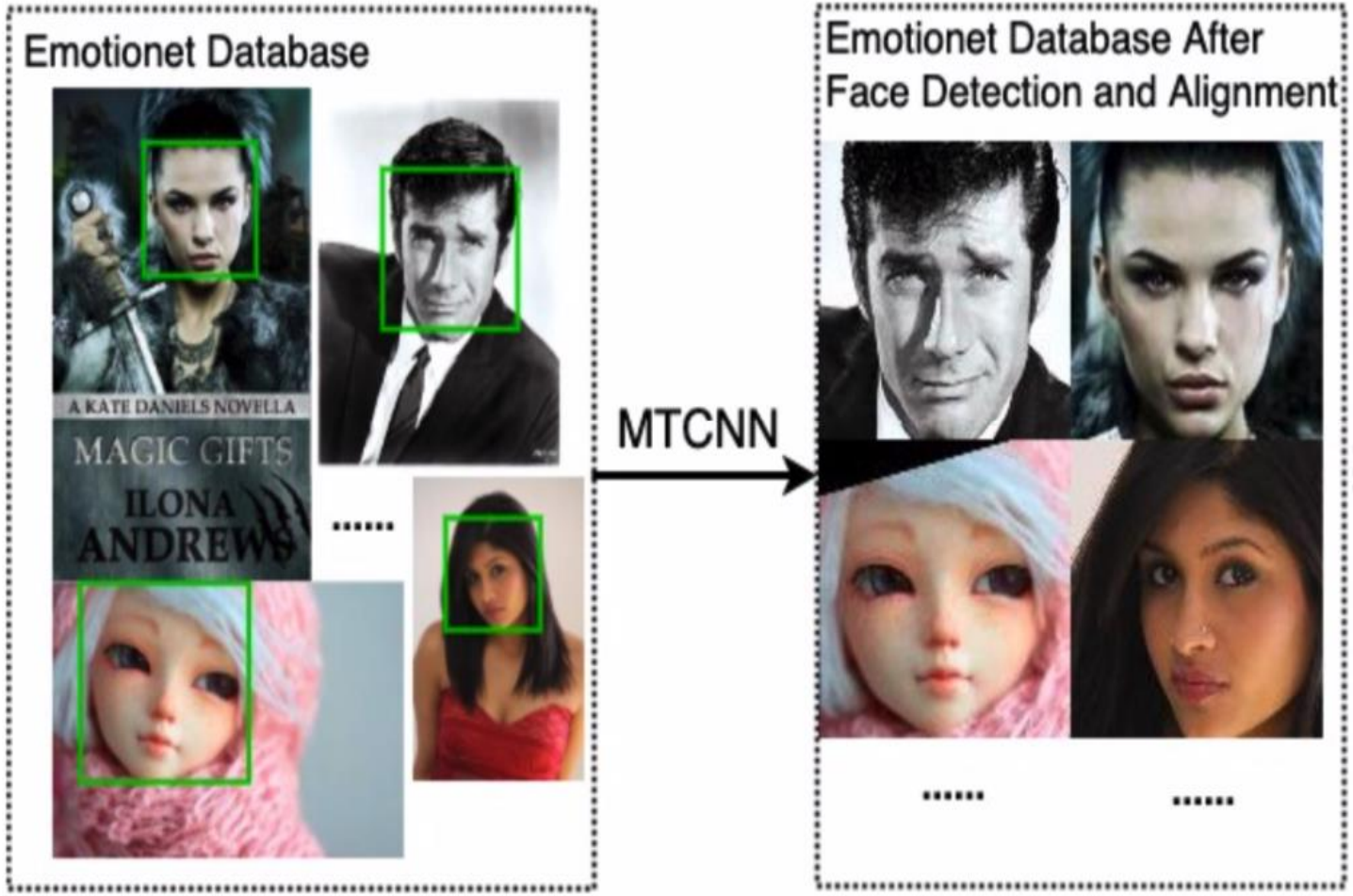



Figure of MTCNN[1]

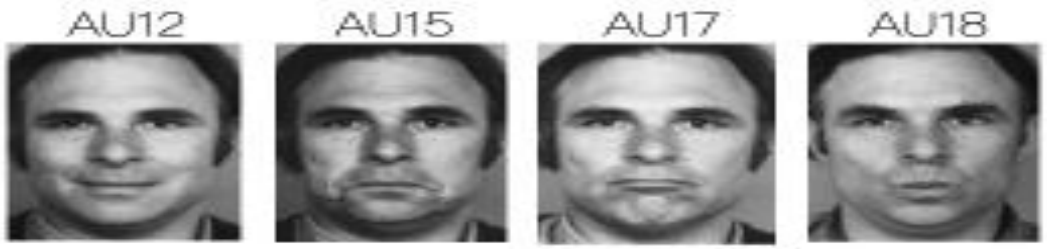



Result demonstrate

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10):1499–1503.

# Challenge: discriminative feature learning

- Tiny face muscle change leads to different AU




- Sub-face level changes requires network to learn deep fine-grained representative features
- Lack of labelled data(25K) have a risk of overfitting

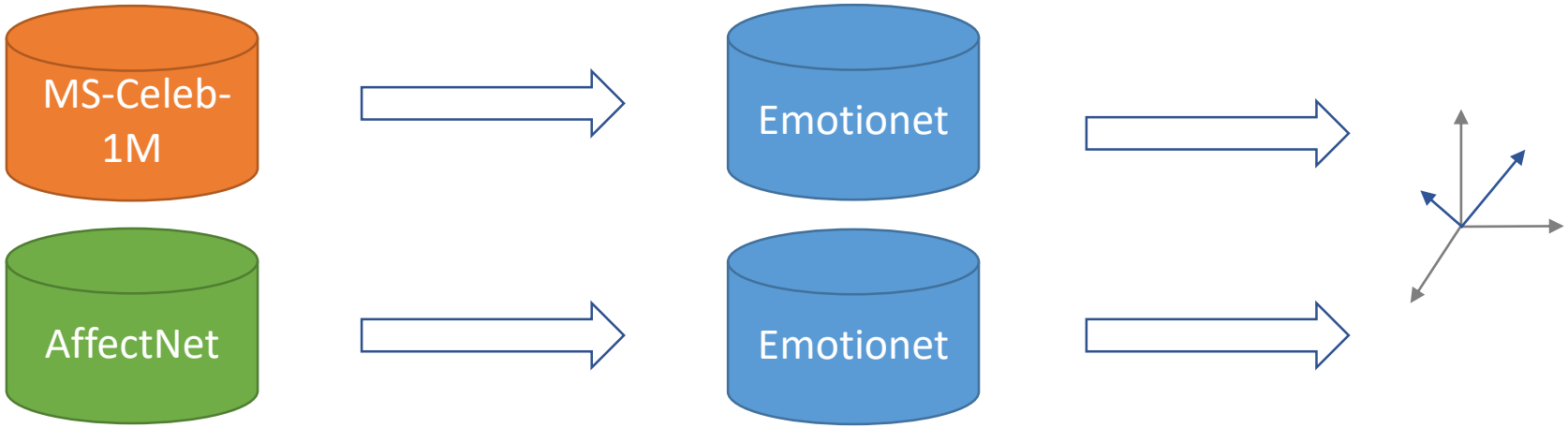
# Solution: transfer knowledge from multi relevant tasks

Opportunity：

- Most location of AU happens on the face landmark position

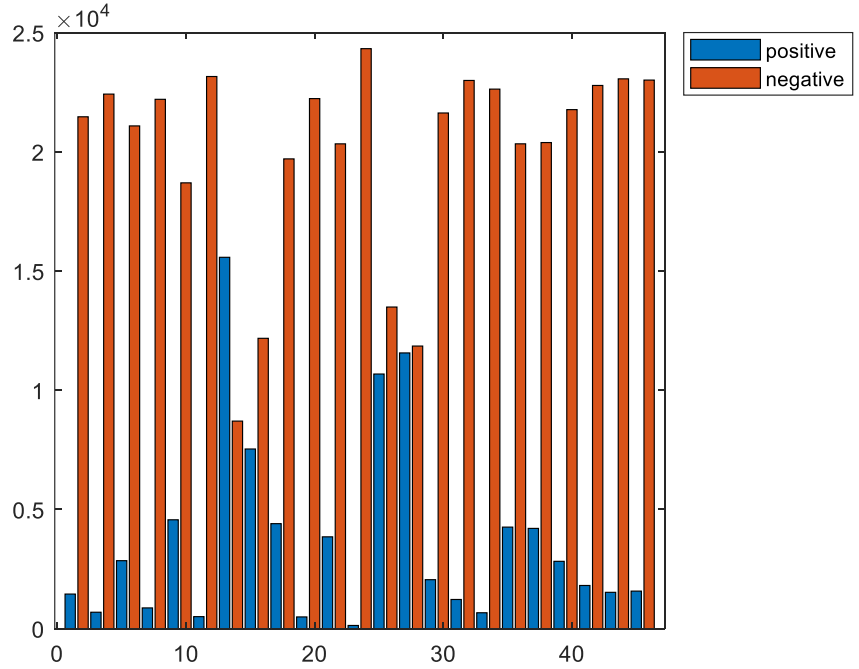
  Transfer knowledge from face recognition task
- Some AU will occur simultaneously to form a specific emotion

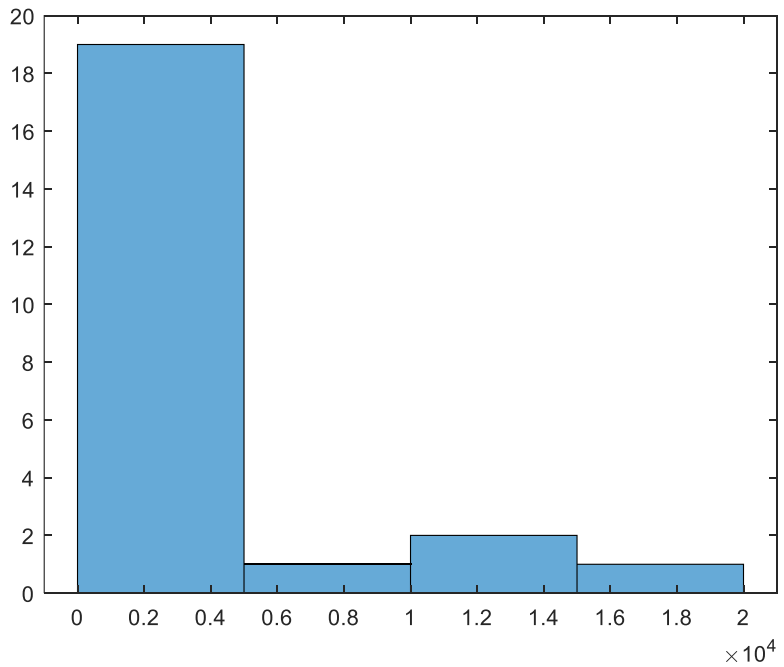
  Transfer knowledge from facial expression recognition task



# Challenge: imbalanced AU distribution

- Most of the Action Unit is labeled negative, negative samples are outnumbered positive samples
- Some AU occurs frequently while some occurs seldomly



# Solution: multi-label balancing strategies

- Balance Sampling

  Data processing phase
- Selective Learning

  Training phase
- Soft Thresholding

  Post optimization phase

# Results

Table 1. Result on Preliminary Backbone Selection

| Backbone | Acc | F1 | Avg |
|---|---|---|---|
| ResNet-18 | 0.901 | 0.364 | 0.633 |
| ResNet-50 | 0.905 | 0.409 | 0.657 |
| ResNet-152 | 0.905 | 0.474 | 0.690 |

Table 2. Ablation Results

| Method | Acc | F1 | Avg |
|---|---|---|---|
| F_face | 0.921 | 0.384 | 0.652 |
| F_emotion | 0.925 | 0.429 | 0.677 |
| F_face + F_emotion | 0.925 | 0.494 | 0.710 |
| Align + F_face + F_emotion | 0.927 | 0.527 | 0.727 |
| Align + F_ face + F_emotion + balancing | 0.915 | 0.552 | 0.734 |

2020 CHALLENGE

| Group | Mean Accuracy | F1 | Final Score |
|---|---|---|---|
| TAL | .9147 | .5465 | .7306 |
| University of Magdeburg | .9124 | .5478 | .7301 |
| SIAT-NTU | .9013 | .4410 | .6711 |
| USTC-alibaba | .8609 | .3497 | .6053 |

2020 VALIDATION TOP-3

| Group | Mean Accuracy | F1 | Final Score |
|---|---|---|---|
| TAL | .9200 | .5720 | .7460 |
| University of Magdeburg | .9198 | .5706 | .7452 |
| SIAT-NTU | .9195 | .3531 | .6363 |

# Conclusion

- Conduct targeted solution to each challenge
  - Jointly face detection and alignment
  - Multiple Transfer Learning
  - Multi-label balancing strategies
- Our solution demonstrates robustness during three phase of Emotionet Challenge 2020 and achieve the 3$^{rd}$ place