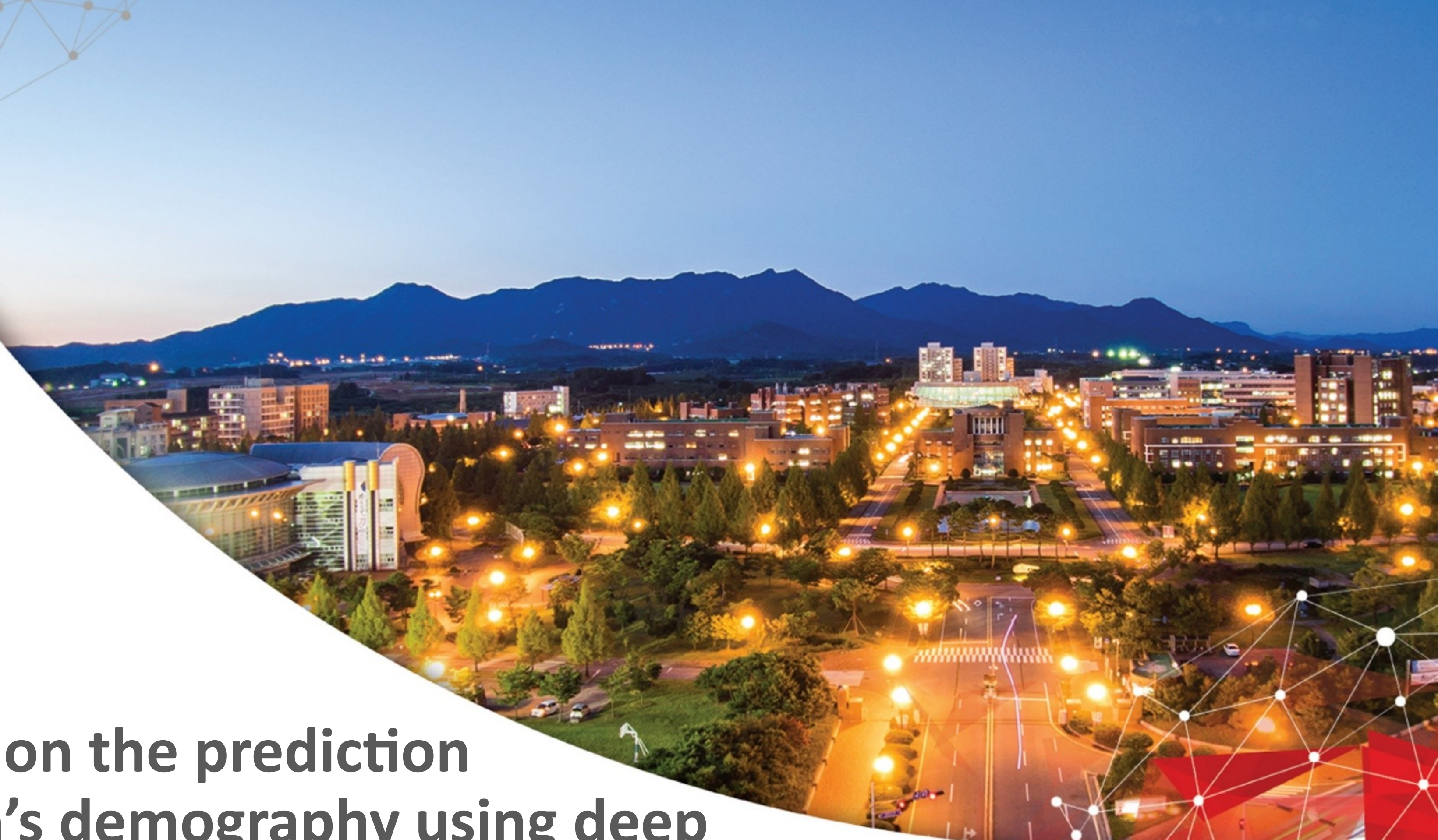
Defense for Master of Science
School of Electrical Engineering and Computer Science



A study on the prediction of Asian's demography using deep neural network

Tran Trung Tin / 20194070 / Advisor: Prof. Moongu Jeon

MLV lab.

GIST

Template from Yechan Kim

I. Introduction

- Age and Gender are two important key features on human face.
- Combine with a face detection system, Age & Gender prediction has some applications such as recommendation system or human-computer interaction.
- With the help of convolutional neural network, and public large-scale face dataset, the Age & Gender prediction accuracy improve significantly.

Source: https://www.kiwi-digital.com/produkty/age-gender-detection

I. Introduction – Problem 1

- Recent large-scale public datasets for Face detection and Age, Gender prediction are available. Such as:
  - IMDB (~460.000 images).
  - Wiki (~62.000 images).
  - CACD, Cross age celebrity data (~163.000 images).
  - UTKFace, Facial, FG-NET.

- All mentioned datasets above are mixed of all human facial characteristic from different country (e.g. America, Europe, Asia,...)

I. Introduction – Problem 1

- Two datasets available for Asian (mostly Chinese):
  - AFAD, Asian Face Age (~165.000 images).
  - AAF, All Age Face (~13.000 images).

- ***→ Problem 1: Lack of method that focused on predicting Age & Gender from Asian's face characteristic.***

I. Introduction – Problem 2

- Most of applications using Age & Gender prediction tent to be used on small and handheld device such as:
  - A Raspberry Pi board embedded with a screen at the entrance to detect and predict user Age & Gender for content recommendation.
  - A fun Age prediction app on mobile using on-device resourses to process instead of server for privacy concern.

- Hence, using a bulky model like FaceNet (~140M params) or VGG16 (~138M params) is not ideal on such devices.

- → ***Problem 2: Having a light-weight model that useable on small devices.***

I. Introduction – Contribution

- (I) Proposed a light-weight model for Age & Gender prediction trained on Asian’s face specifically.
- (II) Making a study to improve the real Age prediction accuracy by proposed a Soft-label Representation of Age mechanism.

II. Related methods – Focal Loss

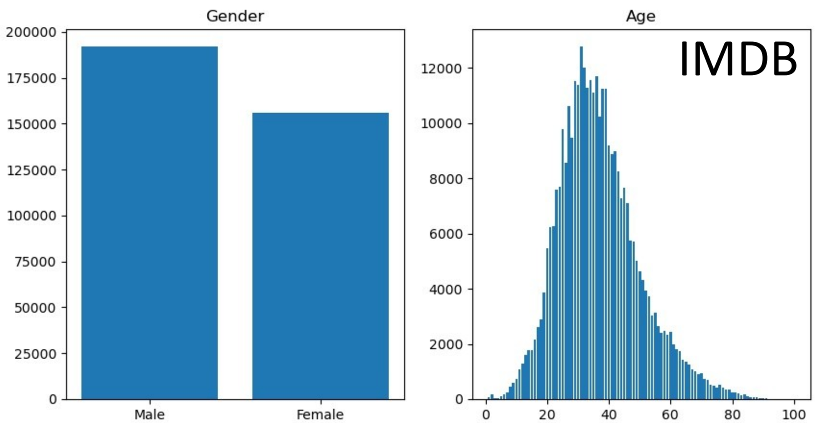
- Proposed by Lin *et al.* [12] in 2018.
- This loss designed to address class imbalance during training tasks e.g. object classification.
- Focal loss add a modulating factor $(1-p_t)^\gamma$ to cross entropy loss. With $\gamma \geq 0$ as a tunable focusing parameter which reduce the relative loss from well-classified examples $(p_t > 0.5)$. The loss defined as:

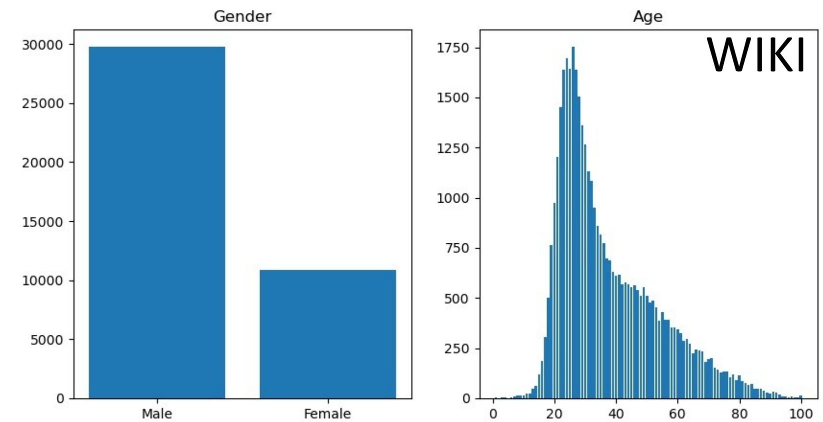
$$FL(p_t) = -(1-p_t)^\gamma \log(p_t)$$

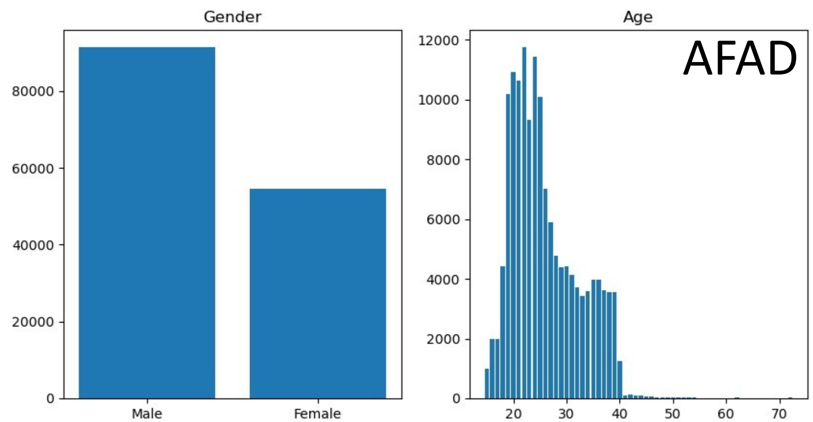
[12] T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Doll´ar, "Focal loss for dense object detection," CoRR, vol. abs/1708.02002, 2017.

II. Related methods – Why Focal Loss ?

- Since most of the available dataset for Age & Gender prediction are having one similar problem: Skew to the younger age.







GIST

II. Related methods – Kullback-Leibler (KL) Divergence

- Measure for the differences between two probability distribution.
- KL-Divergence is defined as:

$$D_{KL}(P\|Q)=\sum_{x\in X}P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

- With $P(X)$ is true distribution and $Q(X)$ is predicted distribution in space X.

III. Methodology – Soft-label Represenation of Age

Age : 22

Guessing: maybe around 20-25 ?

III. Methodology – Soft-label Represenation of Age

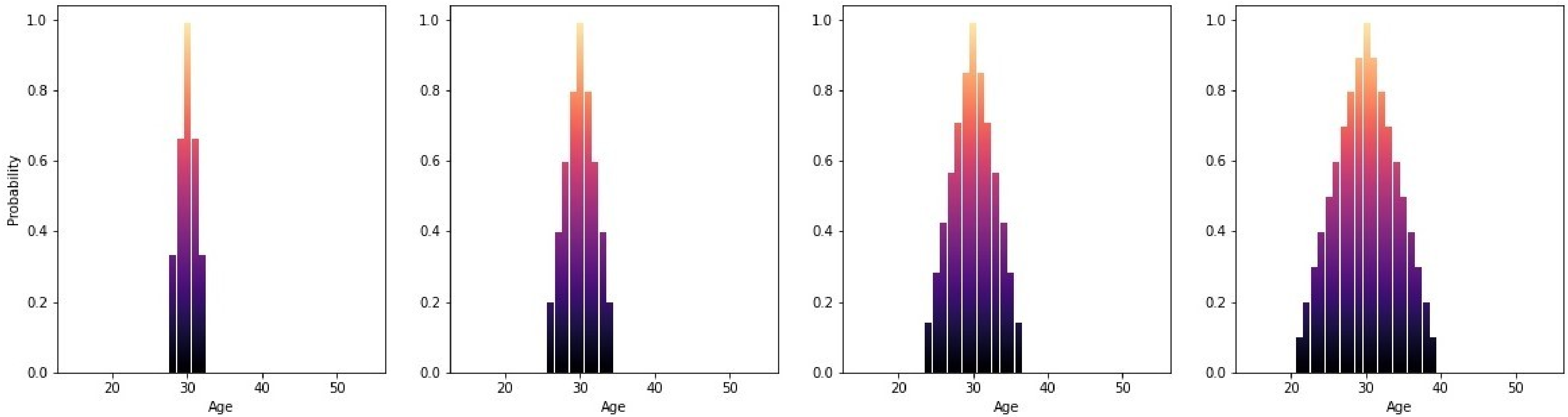
- The regression label (real-age) will be presented as a convex shape with highest probability at the truth label and lower in both size with a gap value.

Real Age: 22

1	0
...	0
20	0.33
21	0.66
22	0.99
23	0.66
24	0.33
...	0
100	0

gap = 3

III. Methodology – Soft-label Represenation of Age



- Probability change with different gap values (3, 5, 7, 10 respectively), age = 30

GIST

III. Methodology – Combine Classification and Regression

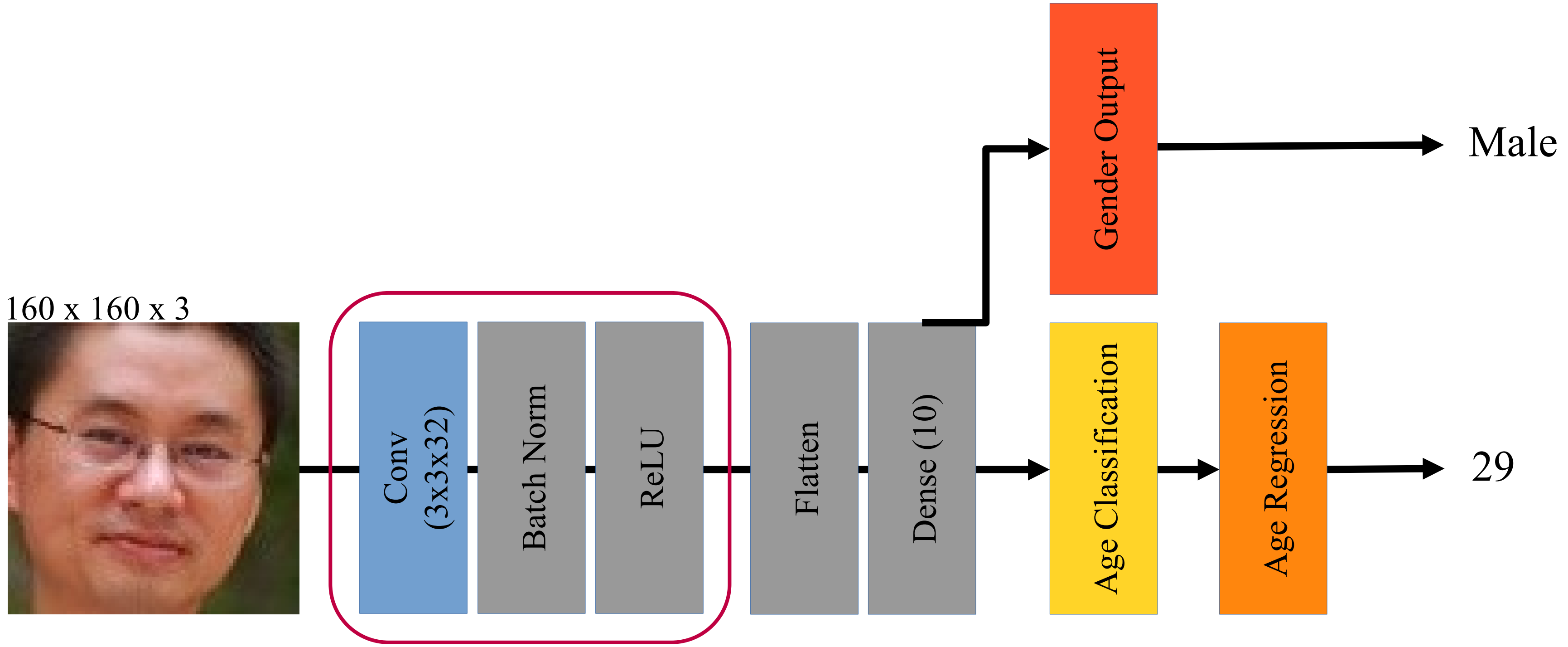
- From the above mechanism, a cascade style of training also implemented.
- With classification label, Focal Loss and KL-Divergence Loss will be applied.
- With regression label, used Mean Absolute Error (MAE Loss). Defined as:

$$L_{reg}(y_n, y'_n) = \sum_n \| y_n - y'_n \|$$

- In training process, two losses will be combined as a total loss:

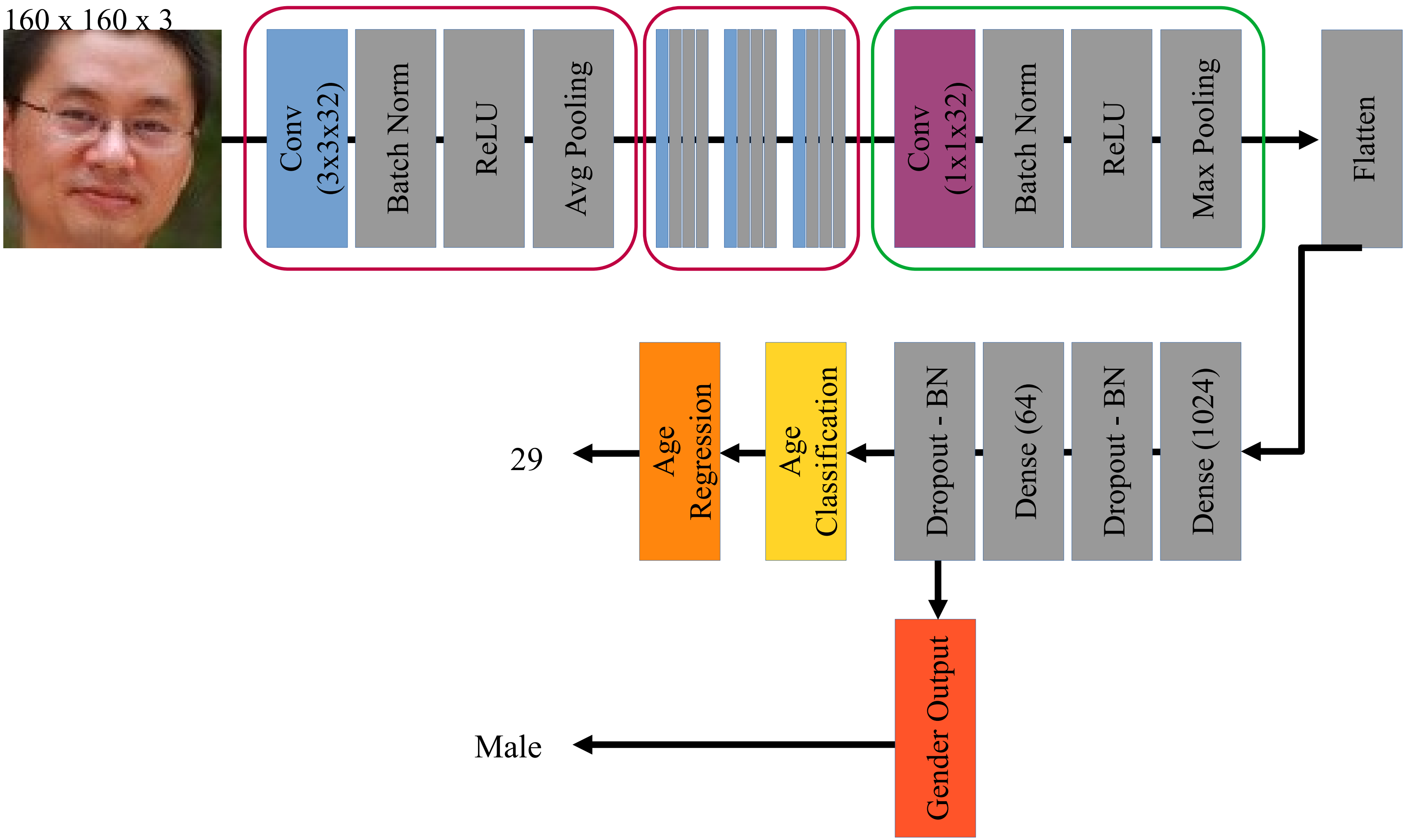
$$L_{combine} = L_{reg} + L_{class}$$

III. Methodology – Base CNN



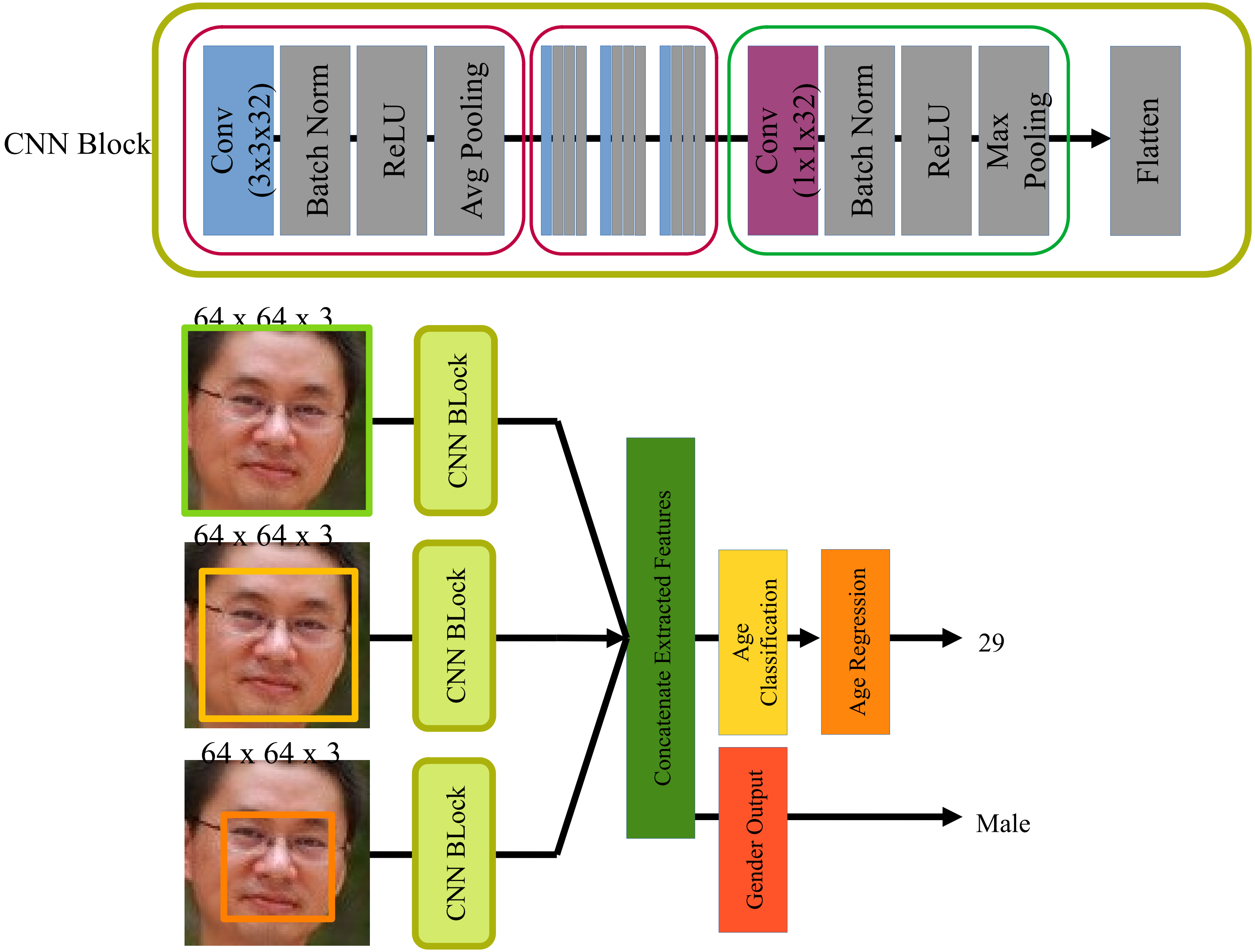
GIST

III. Methodology – Deeper CNN (D-CNN)



GIST

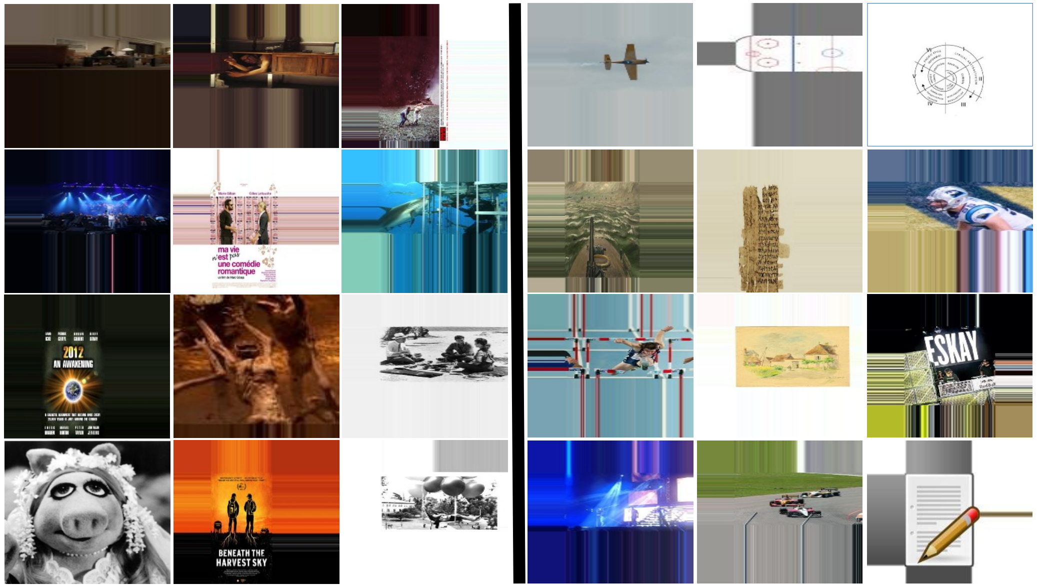
III. Methodology – Multiple Input CNN (M-CNN)



GIST

IV. Experiments – Datasets

- **IMDB**: Contains 460.723 face images from 20.284 celebrites crawled from IMDB with the age range from 1-100.
- **WIKI**: Contains 62.328 celebrity's face images from Wikipedia.
- → Since these two datasets are collected large-scale on Internet, there are lots of noise images such as non-person, low-quality images.
- → These datasets will used for pre-train the model.



Examples of noise images from IMDB (Left) & WIKI (Right).

IV. Experiments – Datasets

- **AFAD (Asian Face Age Dataset)**: Contains 160.000 images with age and gender label. This dataset collected from Chineses Social Network, which contained uploaded photos from students and more. The author of [8] manually filtered out the noise data.
- → This dataset will use for fine-tuning the model.

Example images from AFAD dataset with the age of 18, 24, 40, 67

[8] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua, "Ordinal regression with multiple output cnn for age estimation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

IV. Experiments – Pipeline

- (I) All images went thought a face detection to filtered out most of noise images.
- (II) Resized to [160x160x3] & [64x64x3] includeding corresponding Age & Gender label.
- (III) All datasets used for training will be splited into train/test as [80:20].
- (IV) Run and collect the results from all proposed models with different losses.

IV. Experiments – Training Configuration

- (I) ***batch_size=512*** will keep the same in every experiments.
- (II) ***epochs=100***.
- (III) ***learning_rate=0.001***, decreased by factor of 0.1 with patience epochs=10 to min=0.00001.
- (IV) gap value ***g=3*** as default.
- (V) ***Adam Optimizer*** was applied.

IV. Experiments – List of experiments

- (I) Collect the results across all the proposed models.
- (II) Select the best performance model and tuning its parameters.

IV. Experiments – 1st experiment

Table 4.1: Evaluation results on IMDB-WIKI and AFAD dataset with three proposed models and different losses configuration. The results formatted as *MAE/Gender Acc* (Lower/Higher is better, respectively).

Losses	Model	IMDB-WIKI	AFAD
KL	Base	10.014/0.660	4.670/0.890
	D-CNN	7.689/**0.770**	**3.505**/**0.945**
	M-CNN	7.832/0.632	5.588/0.416
FL	Base	9.698/0.676	4.521/0.910
	D-CNN	**7.521**/0.766	3.843/0.943
	M-CNN	9.425/0.743	4.898/0.781

KL: Kullback-leibler divergence
FL: Focal Loss
D-CNN: Deeper-CNN
M-CNN: Multiple-CNN
MAE: Mean Absolute Error

GIST

III. Experiments – 1st experiment

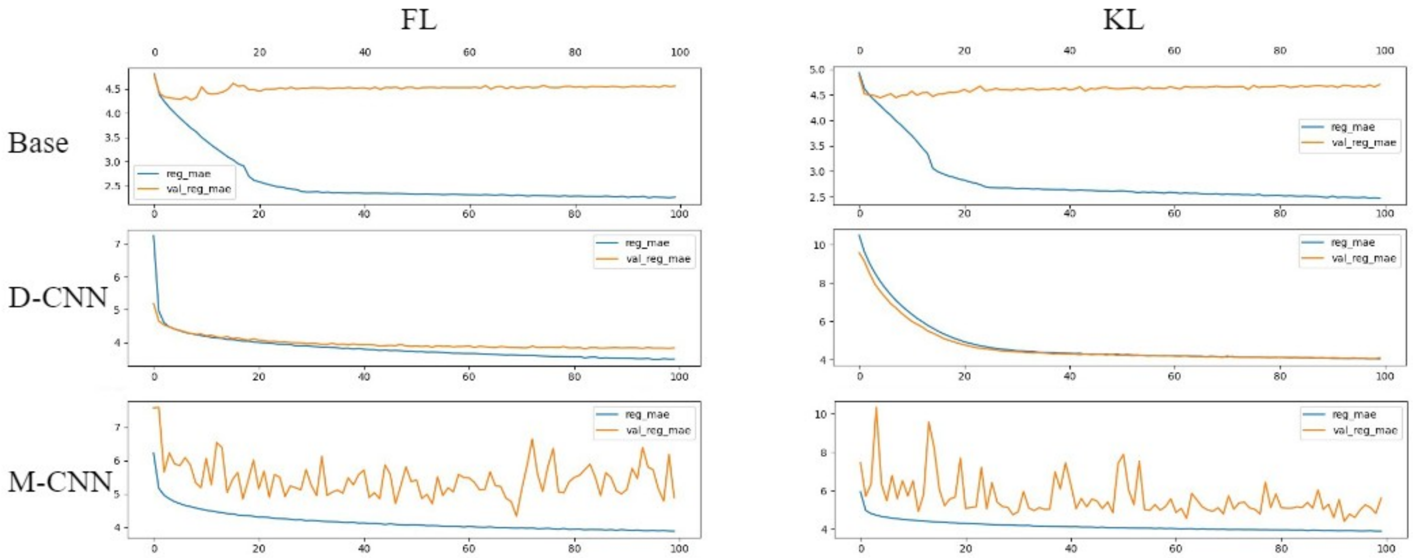


Figure 4.4: Comparison of MAE values from each proposed model with each loss on Train/Valid set. These results were collected after 100 epochs of training with AFAD dataset.

GIST

III. Experiments – 1st experiment

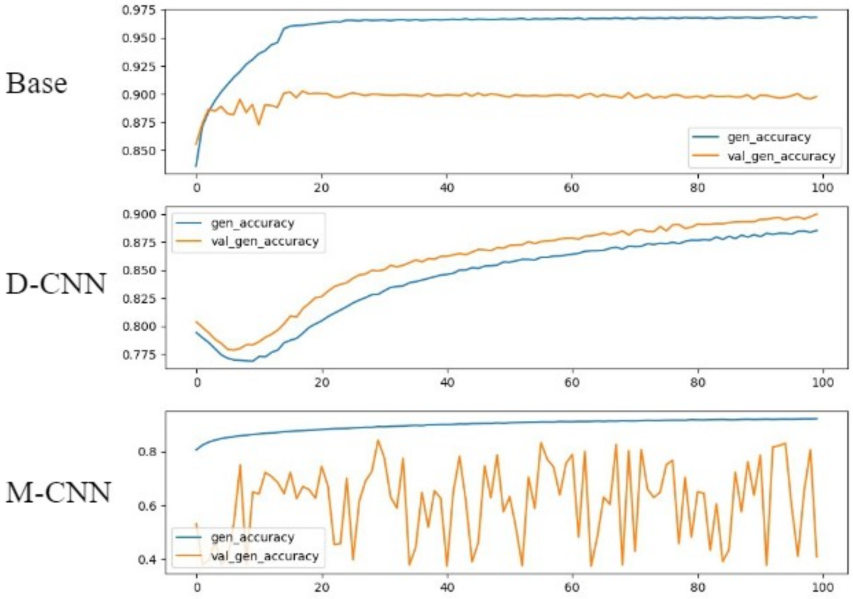
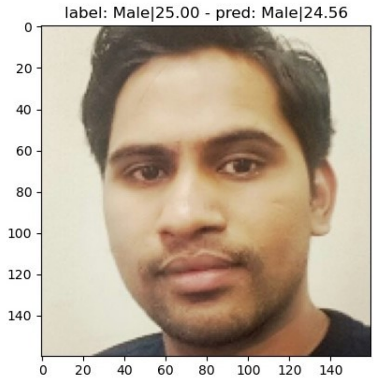
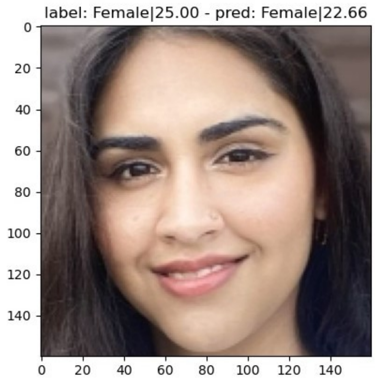


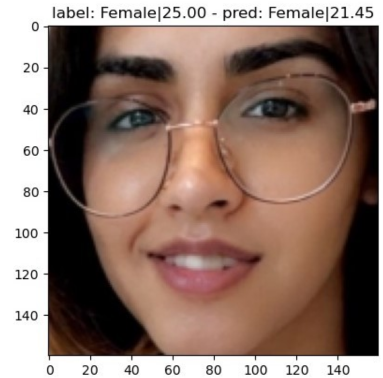
Figure 4.5: Comparison of Gender accuracy from each proposed model. These results were collected after 100 epochs of training with AFAD dataset.

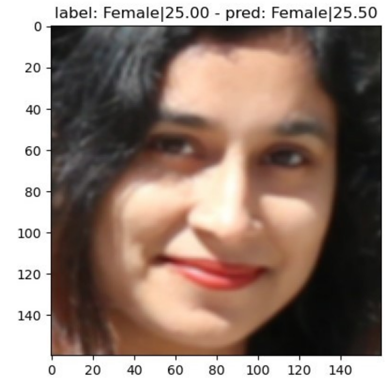
GIST

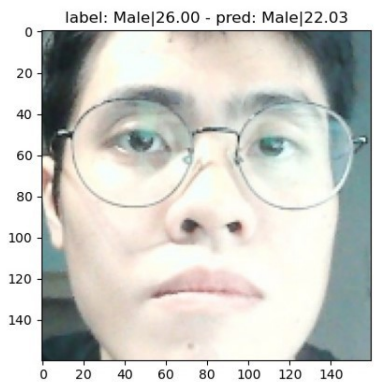
III. Experiments – 1st experiment

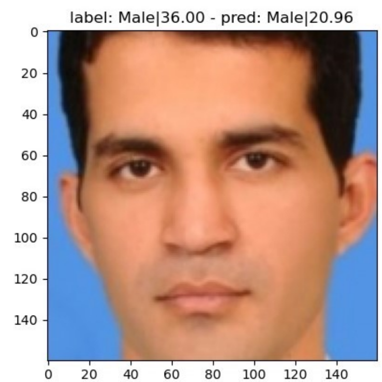












III. Experiments – 2nd experiment

Table 4.2: Evaluation results on AFAD dataset with three CNN based model CNN-LSVR, MR-CNN, D-CNN (Lower is better).

Model	AFAD (MAE)
CNN-LSVR [22]	5.560
MR-CNN [8]	3.510
D-CNN (Proposed)	**3.505**

III. Experiments – 2nd experiment

Table 4.3: Evaluation results on AFAD dataset with three different g values (Lower is better).

Model	g	AFAD (MAE)
D-CNN	1	4.0630
	3 (default)	**3.505**
	5	3.7339
	7	3.7562

III. Experiments – 2nd experiment

Table 4.4: Evaluation results on AFAD dataset with (i) additional weight loss for L_{class}, L_{reg}, L_{gen} equal to 30, 1, 10 respectively (WL); (ii) remove the loss from soft-label representation of age (No class); (iii) default configuration of D-CNN; (Lower is better).

Model	Tuned	AFAD (MAE)
	$WL = [30, 1, 10]$	4.0630
D-CNN	No class	4.0651
	Default	**3.505**

IV. Conclusion

- The combination of Classification and Regression in Age prediction did show its affect to the final real-age prediction.
- In this particular case of Age Classification, Focal Loss did not show the affect of counter skew dataset.
- Higher resolution image can lead to better result instead of multiple small images.

V. Disscussion

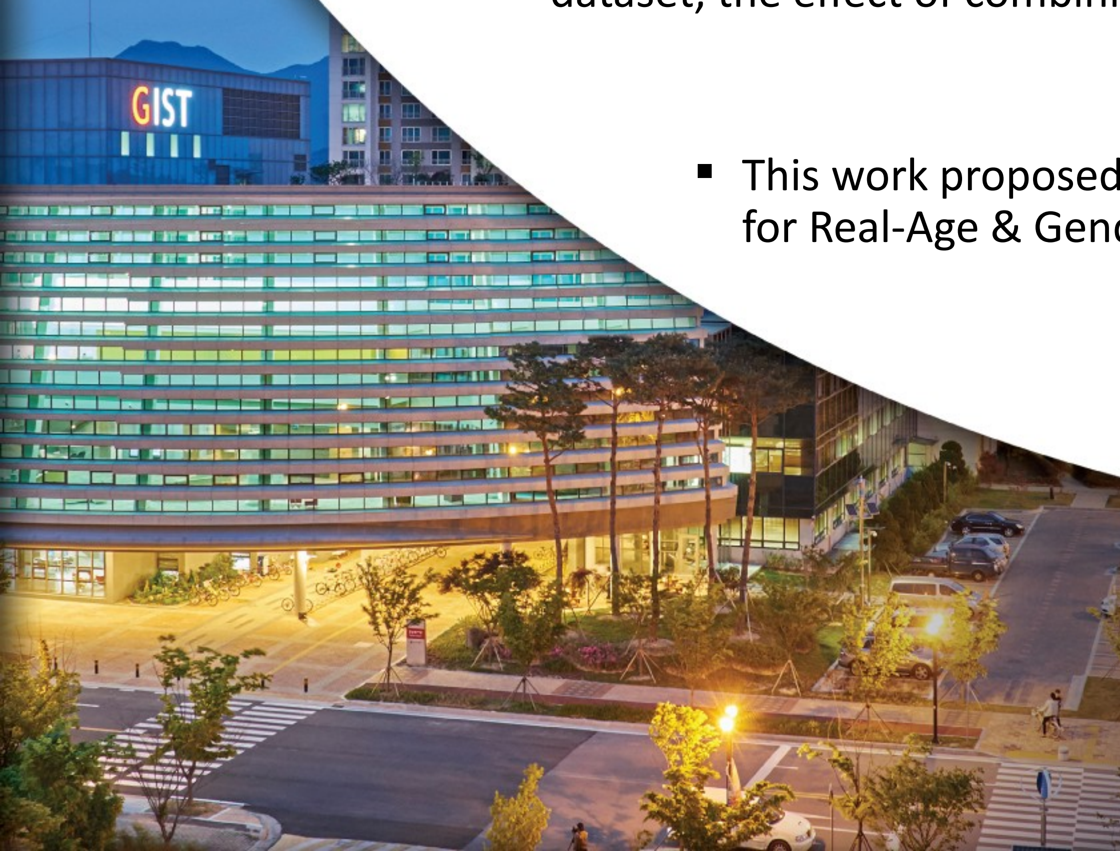
- Facial Plastic Surgery problem → Collect more sources of data instead of just Face image(e.g. heart rate, short movement capture) ?
- An image interceded by beauty filters, photoshop ?

Summary

- This work proposes a new Soft-Label Representaion of Age that help as a classification information for final real-age result.

- This work studies the effect of Focal Loss on imbalance IMDB/Wiki dataset, the effect of combining Classification and Regression Label.

- This work proposed a new light-weight CNN based model for Real-Age & Gender prediction.



Thank you for listening
Q&A