Light-Weight Facial Landmark Prediction Challenge

Chien-Yi Wang
Face Science Team
Microsoft AI R&D Center, Taiwan
chiwa@microsoft.com

Face Science Team at Microsoft Taiwan

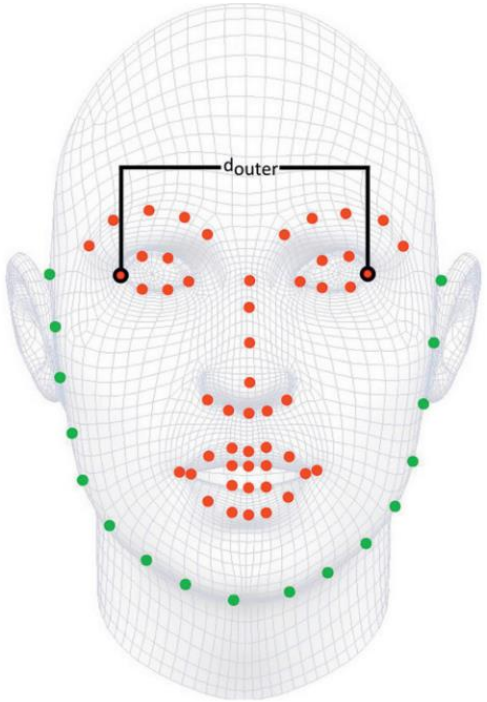
- Topics: Face recognition, Face anti-spoofing, Face synthesis, …
- We are hiring full-time research interns in every semester

Outline

- Introduction
- Dataset
- Challenge Protocol
- Evaluation Metric
- Grading
- Awards

Introduction

- We aim to build a model to predict 68 2D facial landmarks in a single cropped face image with high accuracy, high efficiency, and low computational costs.



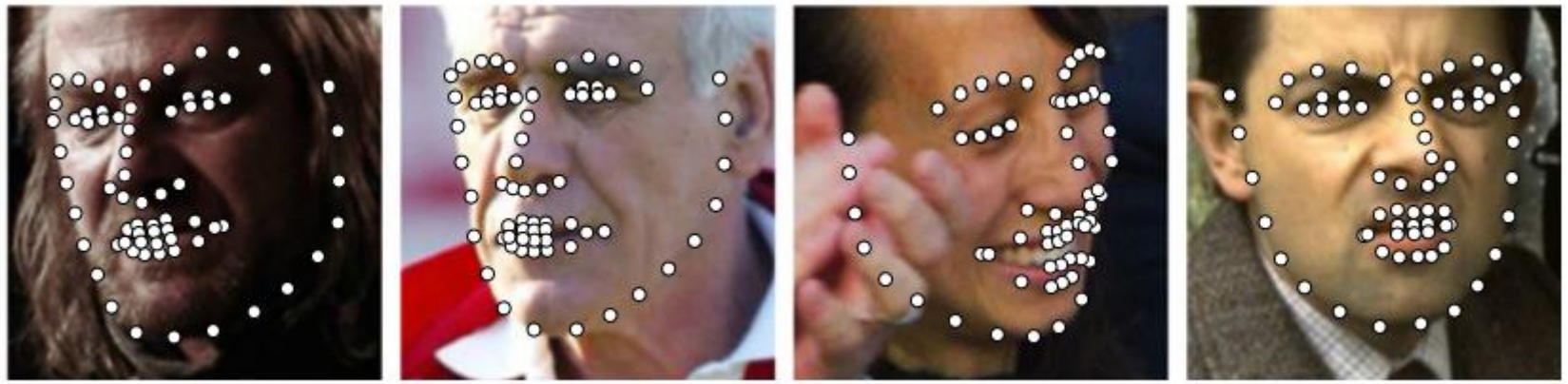


Image Source

Applications

Facial Motion Retargeting

Talking Head Generation

Training Datasets – Microsoft FaceSynthetics

- 100k diverse synthetic facial images with 68 2D landmark coordinates

Accurate ground truth labels can be generated by our synthetic engine
(only the 68 2D facial landmarks can be used in this challenge)

E. Wood et al., "Fake it till you make it: face analysis in the wild using synthetic data alone", in ICCV 2021

Validation/Testing Datasets – AFLW2000-3D

- The fitted 3D faces of the first 2000 AFLW samples, which can be used for 3D face alignment evaluation
  - The 3D-2D projection is used for the evaluation of 68 2D landmarks in this challenge

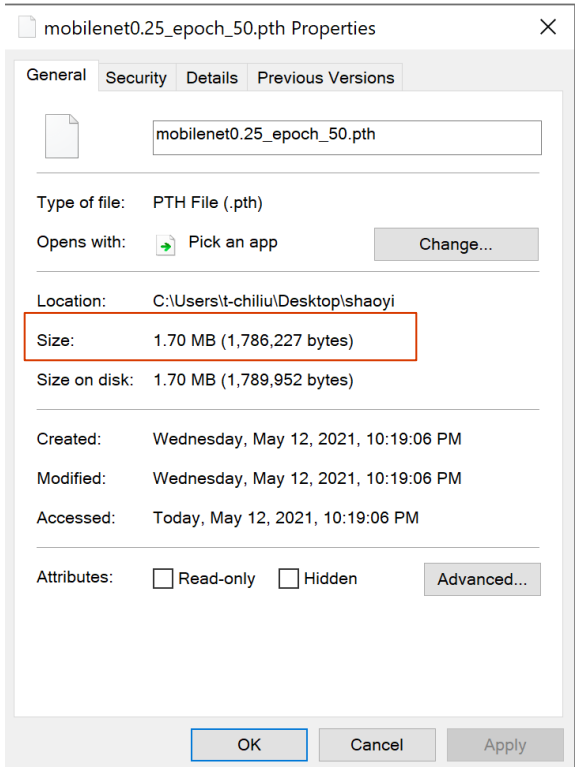
X. Zhu et al., “Face Alignment Across Large Poses: A 3D Solution”, in CVPR 2016

Challenge Protocol

1. We randomly split the AFLW2000 dataset into 10% and 90% as the validation and testing dataset in this challenge
2. You can only train on our training set.
3. The validation set (with labels) can help you validate the results.
4. We will evaluate your method on the testing set (without labels)
5. Face Detector is not allowed during training
6. Please **DO NOT** use external landmark datasets for training your models

Model Constraints

- The upper bound of model size is **15MB**
- The model **should** contain weights (such as neural network)
- We target on **float32** solutions.
  (Float16, int8 or any other quantization methods are **not** allowed.)
- You can **ONLY** train a single model. You **cannot** train an ensembled model.



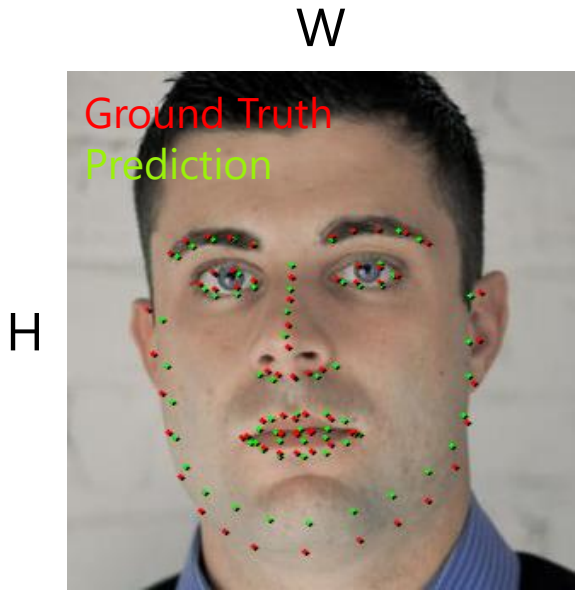
Evaluation Metric

- Under model constraints, we evaluate the results on the testing dataset with the commonly used **Normalized mean error (NME)**

For a matched prediction and the bounding box with width **W** and height **H**, we calculate the NME as follows,

$$NME = \frac{1}{N}\sum_{i=1}^{N} \frac{\sqrt{\Delta x_i^2 + \Delta y_i^2}}{d},$$

where N = 68 and d = $\sqrt{W \times H}$



NEW!

Dataset format

- Download link : https://drive.google.com/file/d/1hhcsXxGehgf_wf2QJKSuwB7e3xxrTYn9/view?usp=sharing

data/

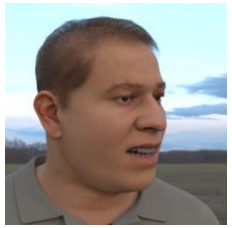
synthetics_train/

annot.pkl

000000.jpg

000001.jpg

……



aflw_val/

annot.pkl

image00013.jpg

……,

NEW!

Dataset format

- How to load the ground truth pickle file?

```
import pickle
with open('your path', 'rb') as f:
    annot = pickle.load(f)
    X, Y = annot
```

X is a list containing image names.

Y is a list of list, and each element is a list containing 68 ground truth landmark (x,y) coordinates.

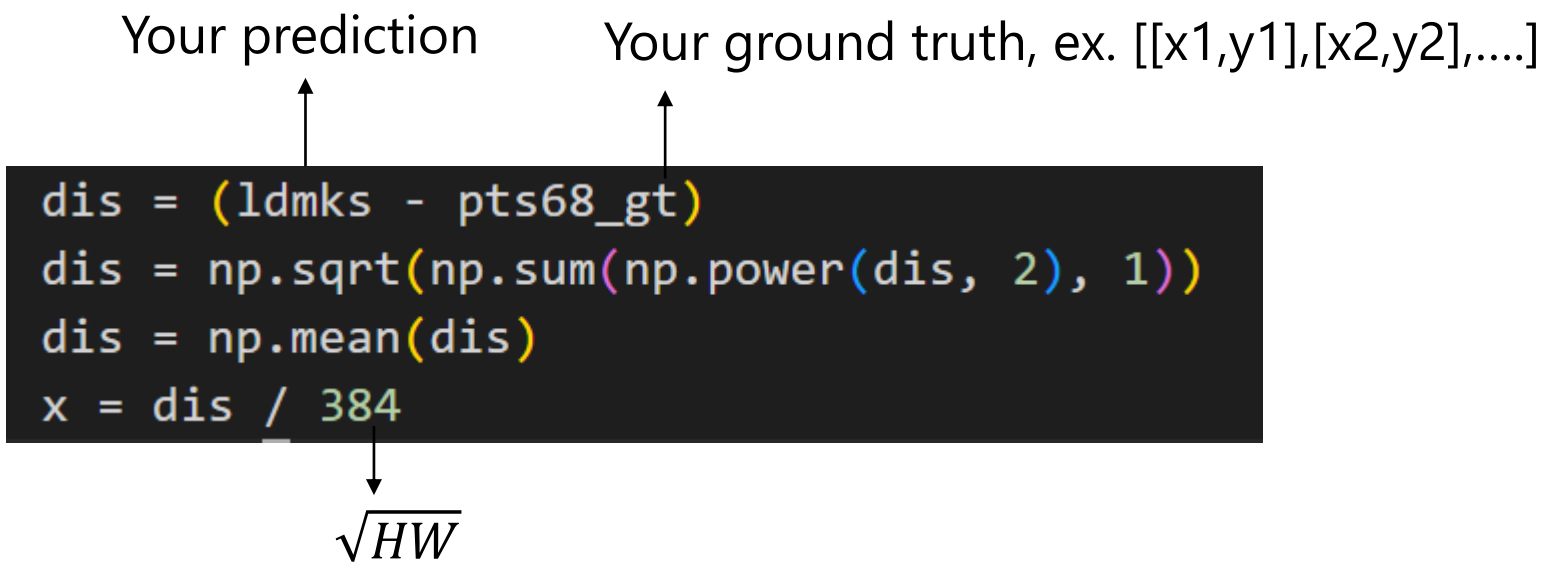
Ex. If we print the first element:

```
print(X[0])
print(Y[0])
```

```
image02408.jpg
[(111.59376, 182.81862), (116.13023, 208.07193), (124.482155, 230.69882), (131.35452, 252.12373), (141.70364, 273.31085), (154.61208, 290.53793), (165.997, 301.28796), (182.60072, 310.43527), (206.86557, 311.82864), (233.0994, 302.87238), (254
.72496, 288.61755), (271.5397, 273.8399), (284.583, 252.41061), (290.10254, 228.27419), (290.91284, 205.87512), (292.11243, 182.21095), (290.8931, 157.12468), (114.360985, 156.241), (121.548225, 149.99855), (132.0784, 146.11295), (142.52617, 1
43.1792), (152.90811, 140.92203), (195.95723, 135.28134), (207.4049, 131.23265), (220.66728, 129.4109), (235.43791, 128.7139), (250.00038, 133.73232), (178.41107, 161.37543), (177.87138, 177.33313), (177.42969, 194.58379), (178.1808, 208.63658
), (169.74551, 222.30959), (175.17937, 222.24596), (183.7246, 222.6826), (193.31168, 220.80333), (201.22157, 219.13063), (131.96628, 167.29024), (136.33342, 159.93718), (146.66042, 157.46002), (157.50896, 163.65761), (148.9875, 168.21695), (13
8.85498, 170.50003), (206.21097, 155.15405), (212.75107, 145.31543), (223.15895, 145.14531), (234.53879, 151.05911), (225.13979, 155.05875), (214.31537, 156.92178), (163.45567, 261.2568), (167.67874, 251.91907), (175.46483, 243.55487), (182.99
959, 244.69962), (190.2535, 242.16667), (206.98692, 248.48268), (222.47176, 258.56342), (213.36182, 271.5478), (203.61232, 279.2393), (191.35901, 282.38696), (179.65742, 281.66153), (172.34691, 274.74518), (165.565, 261.21948), (175.90991, 254
.35864), (184.42104, 252.42915), (194.9466, 252.16464), (220.89227, 259.1861), (201.33879, 271.22598), (190.87549, 273.54205), (181.48886, 272.5461)]
```

NEW!

How to calculate the NME by yourself



NEW!

Schedule

- **Evaluation Server Open**
  2022/06/01 12:00 GMT+8 (中午十二點)
- **Evaluation Server Close**
  2022/06/15 12:00 GMT+8 (中午十二點)

- Other details will be released soon.

NEW!

Grading (100%)

- Quantitative (50%)
  - We will evaluate the model with NME on the private testing set.

- Presentation (50%)
  - Novelty and technical contribution (20%)
  - Completeness of experiments (20%)
    (comparison with different models, ablation studies, visualization, etc)
  - Presentation (10%)

Note : We will choose 10 teams for the final presentation based on the quantitative results.
For other teams, the other 50% will based on your report.

NEW!

Contact

- 劉致廷
- jackieliu@media.ee.ntu.edu.tw

Challenge Awards

- Top **3** teams of this challenge will receive cash prizes sponsored by Microsoft AI R&D center
  - 1st: NTD $5k
  - 2nd: NTD $3k
  - 3rd: NTD $2k

References

- Synthetic Facial Landmark Datasets:
  - [ICCV 2021] Fake It Till You Make It
- Methods
  - [ECCV 2020] Towards Fast, Accurate and Stable 3D Dense Face Alignment
  - [CVPRW 2019] Accurate 3D Face Reconstruction with Weakly-Supervised Learning
  - [ICCV 2017] How far are we from solving the 2D & 3D Face Alignment problem?