ARG (Age Race Gender) Detection Using Transfer learning based on FaceNet Pretrained Model

Abstract

Focus: vision, face recognition, feature extraction, and classification
Base Line: Pre trained model, VGGFace2 [1] with UTKFace Dataset [2]. Race/Gender Prediction from cropped images [3] Age/gender recognition [4]
Architecture and Algorithm: Tensorflow, Adam Optimizer, Transfer Learning, Multi-Task Learning, YOLO
Contribution: 1- Age Race Gender detection (ARG) 2- Online Face detection and tracking for ARG 3- Investigate network performance by changing hyper parameters.
Training Parameters: 0.63, 0.375, 0.142 Cross-entropy loss for age, race, and gender prediction respectively, after 50 Epochs of training

UTKFace Dataset

- Consists of **20k+ face images** in the wild (only single face in one image)
- provides the correspondingly **aligned and cropped faces**
- Images are labelled by **age**, **gender**, and **ethnicity**
- Age span: from 0 to 116 years old
- 4741 samples were set aside for **validation**.

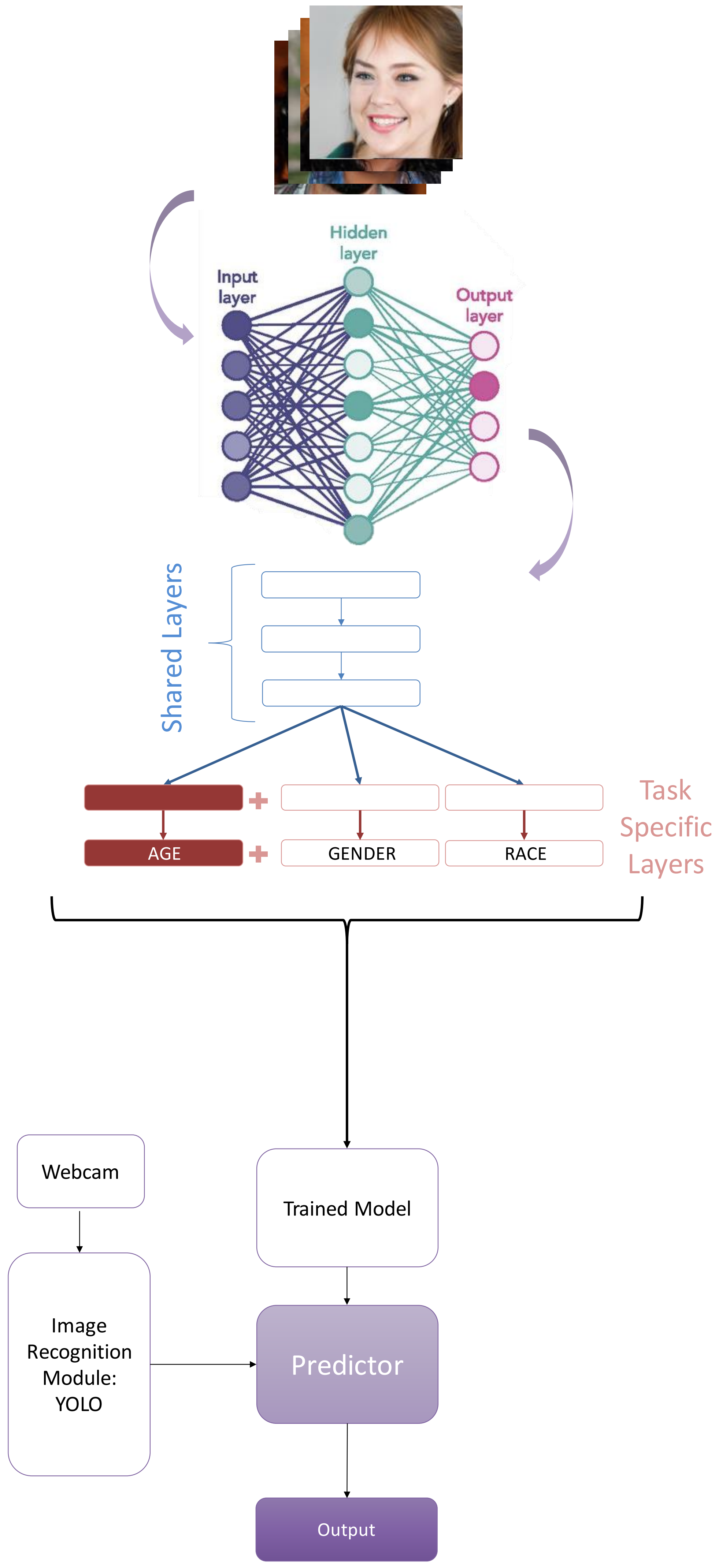
Pretrained FaceNet

- **Transfer learning** techniques have been applied on **pretrained FaceNet** in order to initiate the weights.

- The **CASIA-WebFace** dataset has been used for training.
- The best performing model has been trained on the **VGGFace2** dataset consisting of ~3.3M faces and ~9000 classes
- Data set is more complex compared to UTKFace, thus ideal for transfer learning
- Trained using **softmax loss** with the **Inception-Resnet-v1** model. The datasets has been aligned using **MTCNN**.
- The accuracy achieved for CASIA-WebFace and **VGGFace2** datasets are 0.9905 and 0.9965, respectively.

Online Prediction Module

- Uses **YOLO** for face detection and generating the bounding boxes.
- Recognizes and track faces real-time using a webcam
- Captures multiple images during network feeding phase
- Crops and Aligns face images to generate compatible inputs for the trained model
- Runs each image through multiple threads and takes the average over the results to generate the final prediction
- Inference for multiple images is done in parallel.



Overall Framework

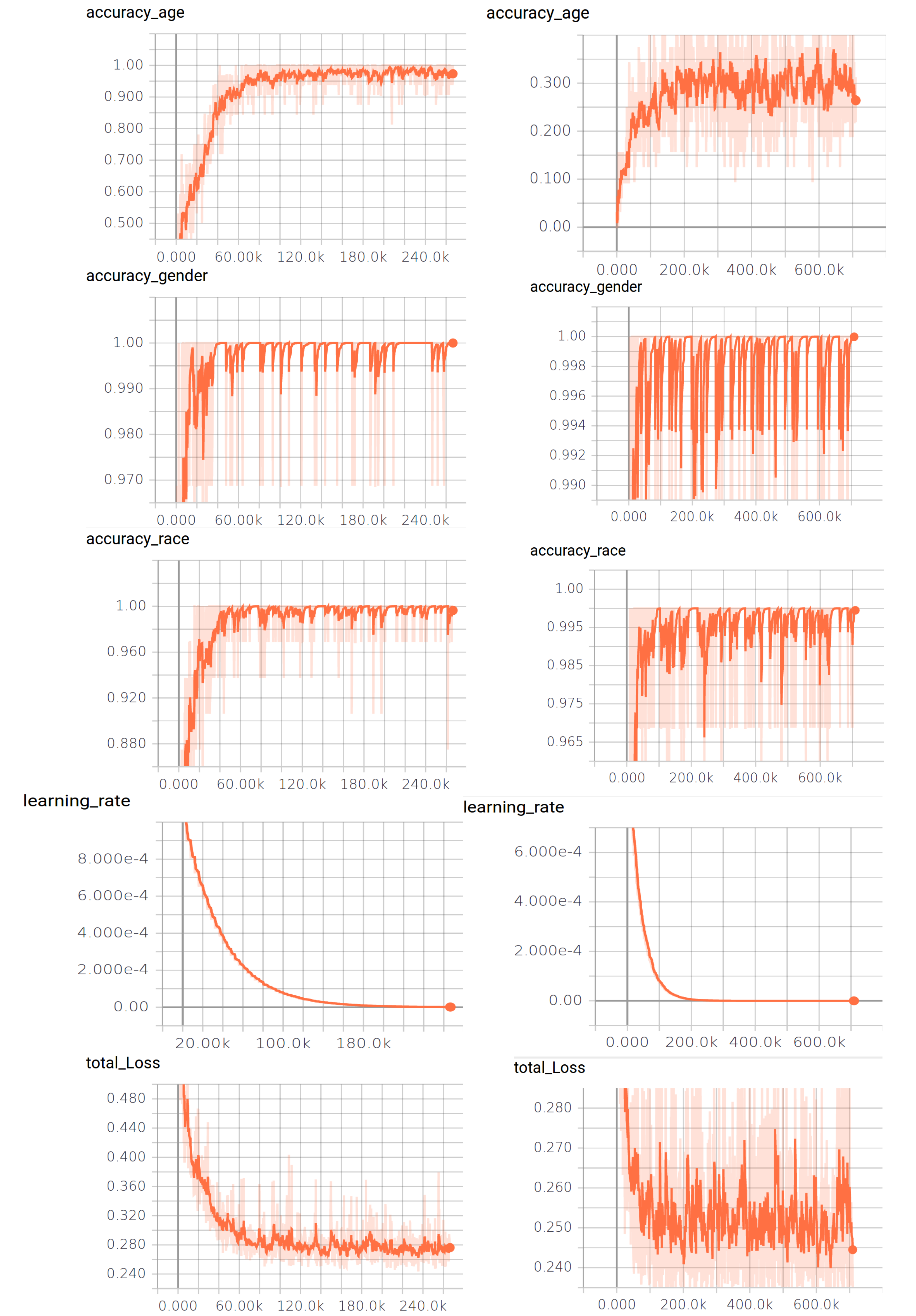
General Structure

- The code for data set extraction and label generation was modified to include age labels.
- The data set augmentation was done by shuffling and flipping the images.
- The trainer and test modules were modified to include age as an output node in the network.
- The age classification was done in 2 different variations.
    - Precision was traded for accuracy.
- The network was trained to generate new weights.
- The first round of results were extracted, then the learning rates were changed, and the second round of results were generated.
- The online prediction module was added to the main framework and tested

Main features:

- TensorFlow-GPU 150 epochs ~ 200 k steps ~ 7 hours run
- Image augmentation to balance the data set
- Used the pertained structure with latest checkpoint, excluding the output nodes which are not useful
- Adaptive learning rate (escape local minimum)
- extract performance from log data with TensorBoard

		Exact Age Prediction	Binning
Accuracy	Age	0.014	0.13
	Race	0.84	0.82
	Gender	0.93	0.92



Discussion

Figures on the left present the training results for the case of ARG prediction using a binning method and 9 classes for age. The figures on the right, show the case of considering 63 age classes. The network trained on 9 classes shows a much better performance in training and evaluation. This results show one of the disadvantages of conventional multi-task learning methods. Larger variations in the number of classes for each task will deteriorate the performance in multitask model.

Contacts

Simon Fraser University
BC, Canada

References

1. https://github.com/davidsandberg/facenet
2. https://susanqq.github.io/UTKFace/
3. https://github.com/zZyan/race_gender_recognition
4. https://github.com/BoyuanJiang/Age-Gender-Estimate-TF
5. Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
6. Jiang, Huaizu, and Erik Learned-Miller. "Face detection with the faster R-CNN." 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE, 2017.
7. Li, Haoxiang, et al. "A convolutional neural network cascade for face detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.