Problem statement

Develop a facial landmarks model that trains on synthetic facial images and evaluates on real facial images.

The model should be able to run inference on consumer devices in real time, taking video frames as input at an FPS of at least 30.

Datasets

Dataset (train & validation): Synthetic 2D human facial images with 70-point 2D landmarks positions.

The Face Synthetics dataset: <https://github.com/microsoft/FaceSynthetics>

Dataset (test): Real-world 2D human facial images with 3D landmarks.

AFLW2000-3D is a dataset of 2000 images that have been annotated with image-level 68-point 3D facial landmarks. <https://www.tensorflow.org/datasets/catalog/aflw2k3d>

<http://www.cbsr.ia.ac.cn/users/xiangyuzhu/projects/3DDFA/main.htm>

Assumptions

Inputs/Outputs

Input: The model takes human facial images without depth information. (Synthetic data in training, real data in inferencing)

Output: The model outputs the probability of a face being present, and the positions of 3D landmarks.

Expected final output: Demo application on desktop CPU that takes human facial images cropped from video frames and predicts 3D landmarks visualized in 3D coordinates.

What I have done so far

1. Face detection module, light weight, fast. Based on BlazeFace (CVPR2019, also used in Mediapipe).
2. Preprocessing: Crop, resize. Based on results from the dace detection module.
3. Train network: preliminary results.
4. Demo: Desktop demo, video input showing FPS and webcam input in real time.

What I will do for the remaining time

1. Determine the minimum viable dataset size for crossing domain gap
2. Identifying which data augmentation techniques offered the most improvements to the model’s accuracy
3. Report, project video and presentation

Ref

3DMM: <https://arxiv.org/pdf/1909.01815.pdf>