Joint Face Detection and Facial Motion Retargeting for Multiple Faces
Bindita Chaudhuri¹, Noranart Vesdapunt², Baoyuan Wang²
¹University of Washington, ²Microsoft Corporation

Objective

- A novel top-down approach to jointly learn
  - bounding box locations
  - 3D Morphable Model (3DMM) parameters
- Multi-scale representation learning to disentangle the 3DMM parameters
- An end-to-end real-time memory-efficient system for practical applications with multi-face images (26 fps on Google Pixel 2)

Methodology

3DMM: $M = V \times b_u \times b_p$
Projected 68 2D facial landmarks:

$\mathbf{p}_u = \frac{1}{J} \sum_{j=1}^{J} \mathbf{R} \left( \mathbf{b_u} \mathbf{w}_u + \mathbf{v}_u + \mathbf{w}_u \right)$

- Conv + BatchNorm + ReLU
- Fire module² + Squeeze-Excite module² + Maxpooling
- Fire module¹ (stride 2)
- Fully-connected layer
- Channel-wise concatenation

Quantitative Evaluation

- Lower 2D landmark error
- Higher alignment accuracy than state-of-the-art
- Improved face detection due to 3DMM constraints
- Constant runtime
- Faster compared to separate networks

Qualitative Results

- Single Face Retargeting (custom test set)
- Multi Face Retargeting (network outputs for AFW and WIDER test set)
- Multi Face Retargeting (live performance capture using webcam and CPU)

Contributions

- A novel top-down approach to jointly learn
  - bounding box locations
  - 3D Morphable Model (3DMM) parameters
- Multi-scale representation learning to disentangle the 3DMM parameters
- An end-to-end real-time memory-efficient system for practical applications with multi-face images (26 fps on Google Pixel 2)