Facial Motion Retargeting

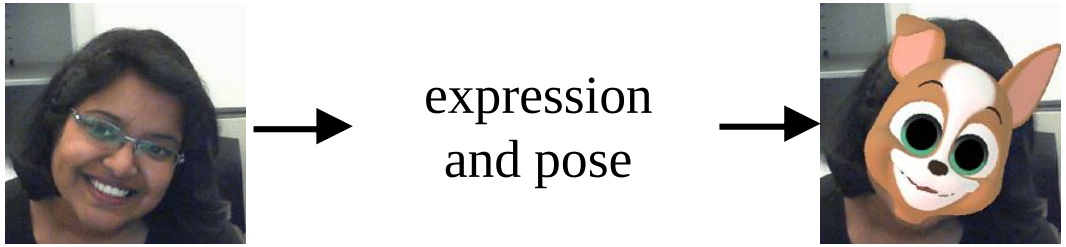
Bindita Chaudhuri

Paul G. Allen School of Computer Science & Engineering, UW
(Linda Shapiro, Alex Colburn, Barbara Mones, Gary Faigin)

Visual Intelligence Group, Microsoft AI&R, Redmond

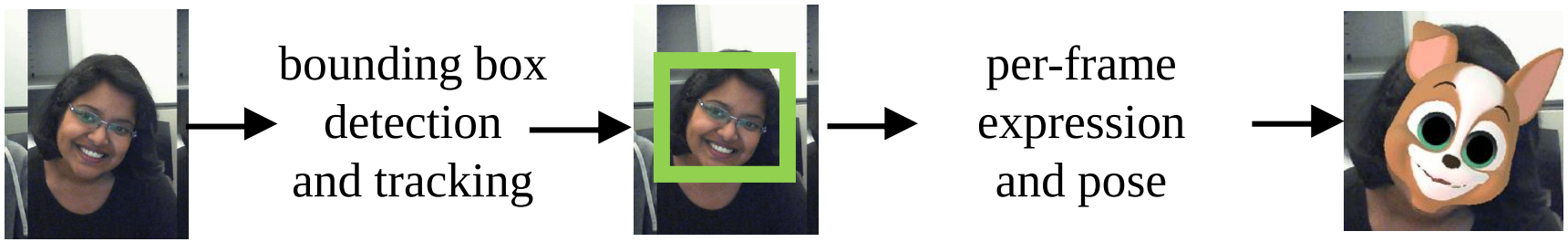
Introduction

- **Goal**: Predict and transfer facial motion from 2D images to 3D models



Introduction

- **Goal**: Predict and transfer facial motion from 2D images to 3D models



- Recent applications include:

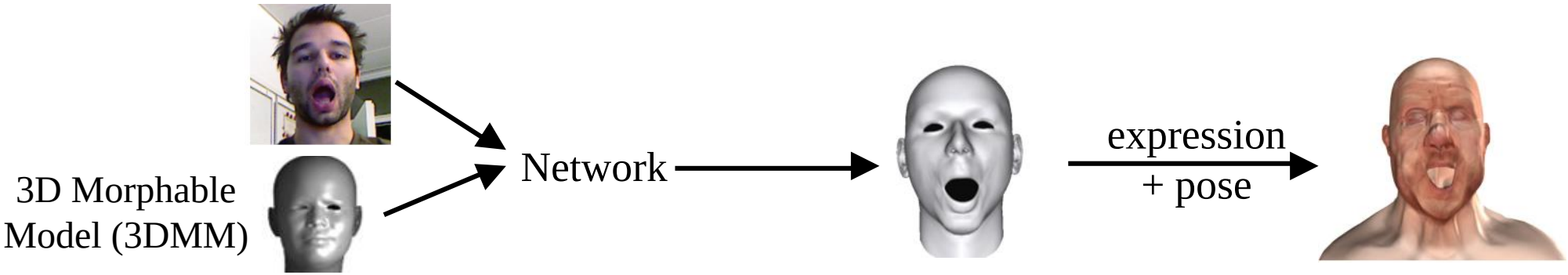
video games

motion capture films

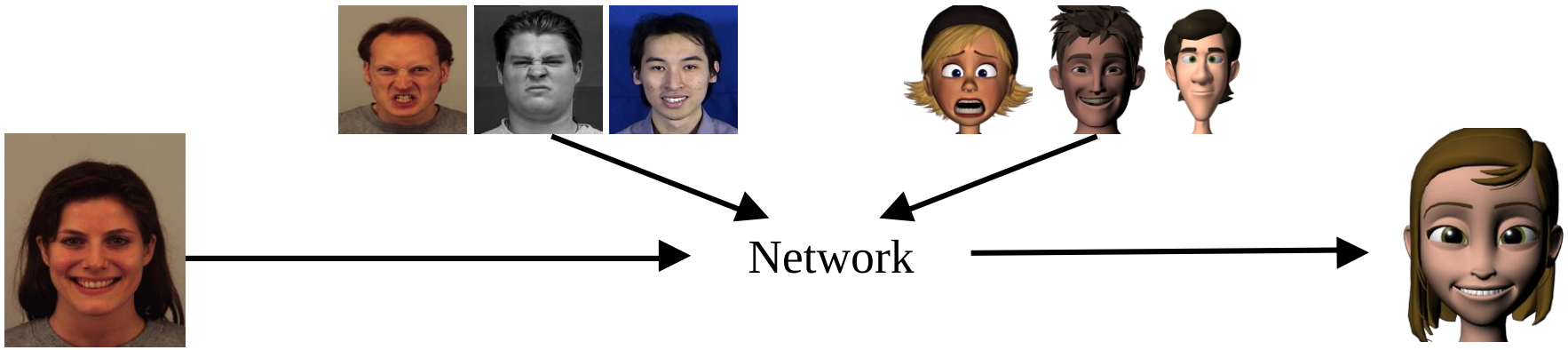
social VR experience

Methodology

- **Blendshape based approach** (*better generalizability to multiple characters*):



- **Example based approach** (*better generalizability to out-of-space expressions*):



Blendshape based Approach

3DMM tensor $T = V \times B_{exp} \times B_{id}$

identity coefficients w_{id}

user-specific blendshapes $T = V \times B_{exp}$

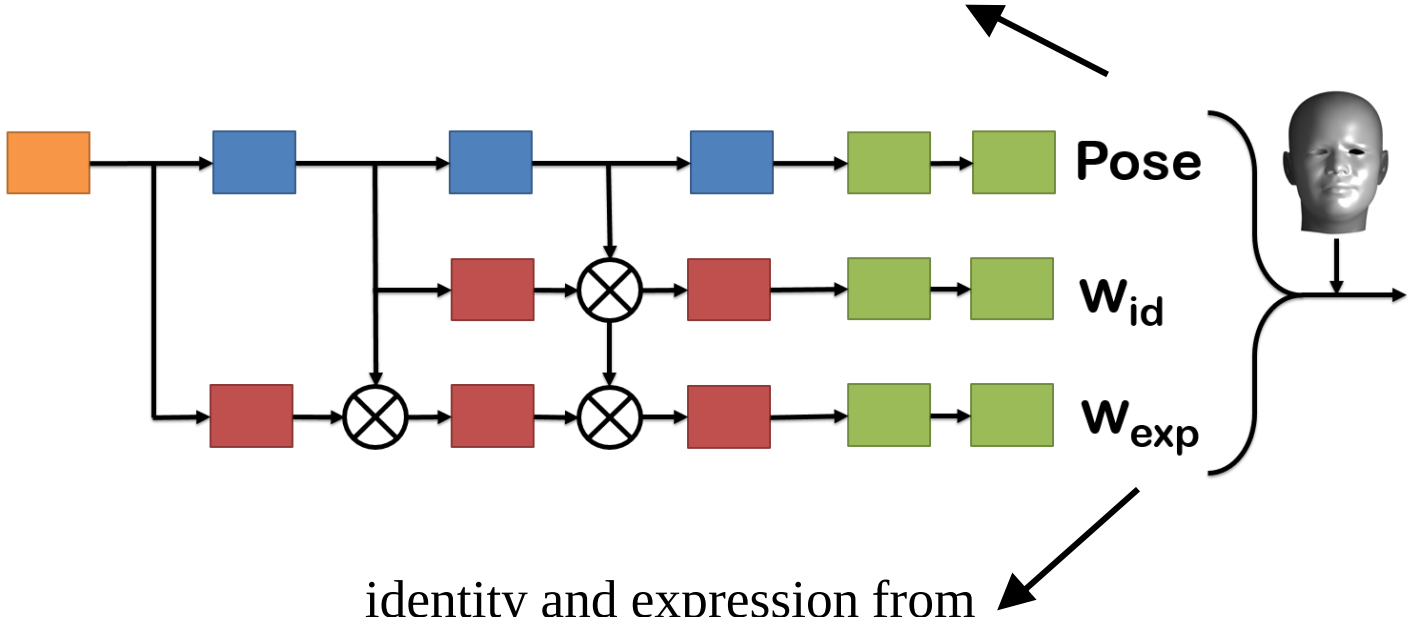
expression coefficients w_{exp}

3D vertex coordinates $M = V$

pose (R, t) and projection (f) parameters

2D landmarks $$\mathbf{P}_{\mathbf{lm}} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \end{bmatrix} [\mathbf{R} * (\hat{\mathcal{M}} * w_{\text{id}} * w_{\text{exp}}) + \mathbf{t}]$$

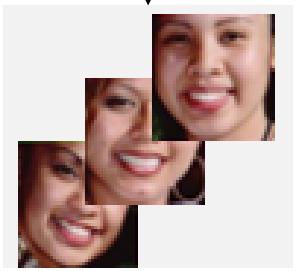
Single Face Network

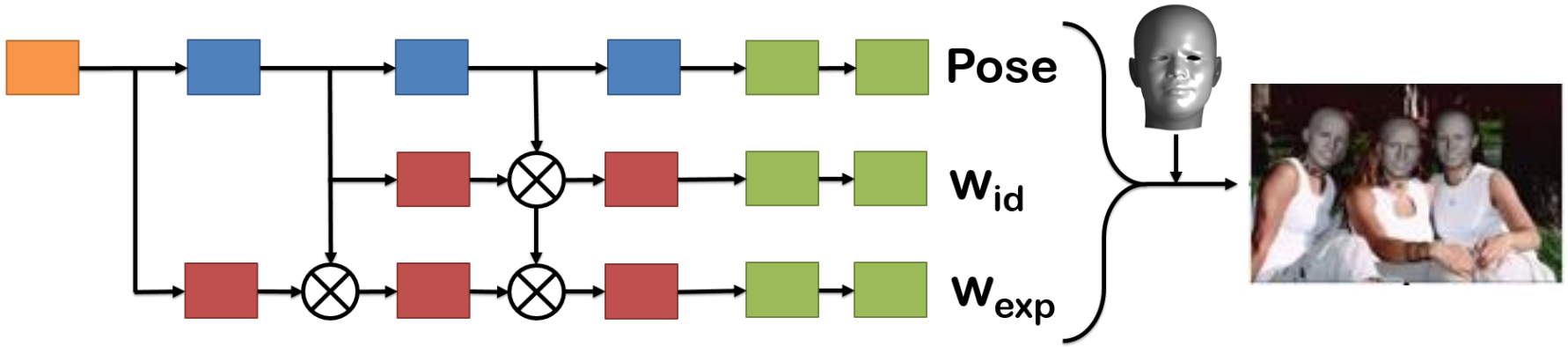


B. Chaudhuri, N. Vesdapunt, B. Wang, Joint Face Detection and Facial Motion Retargeting for Multiple Faces , *CVPR 2019*

Multi Face Network

face detection

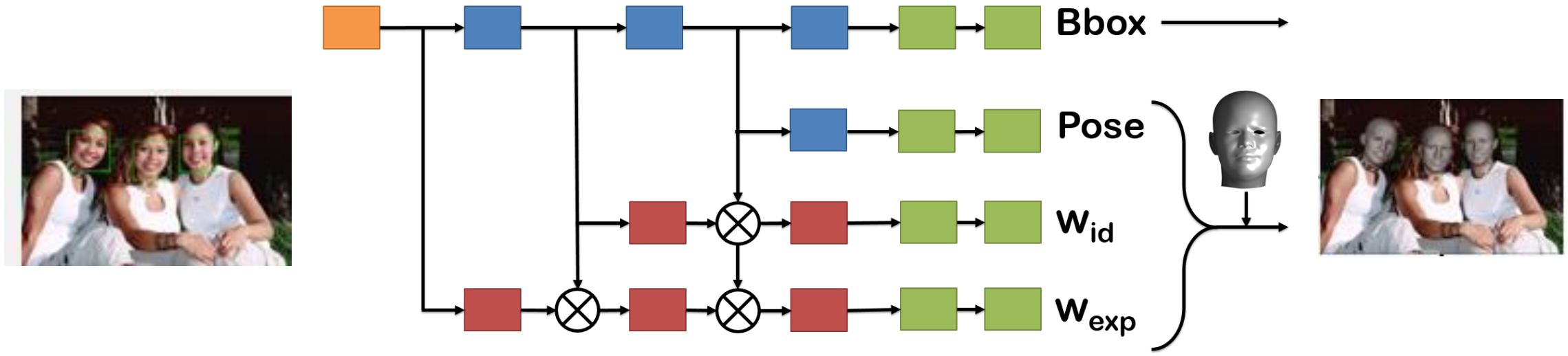




- two sequential networks; memory inefficient
- runtime increases linearly with number of faces

B. Chaudhuri, N. Vesdapunt, B. Wang, Joint Face Detection and Facial Motion Retargeting for Multiple Faces , *CVPR 2019*

Multi Face Network

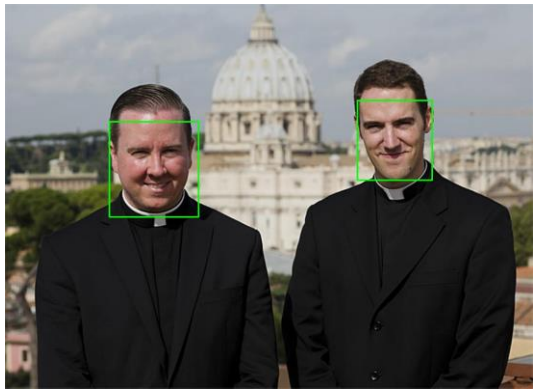


- YOLO loss function; bounding box and 3D face prediction help each other

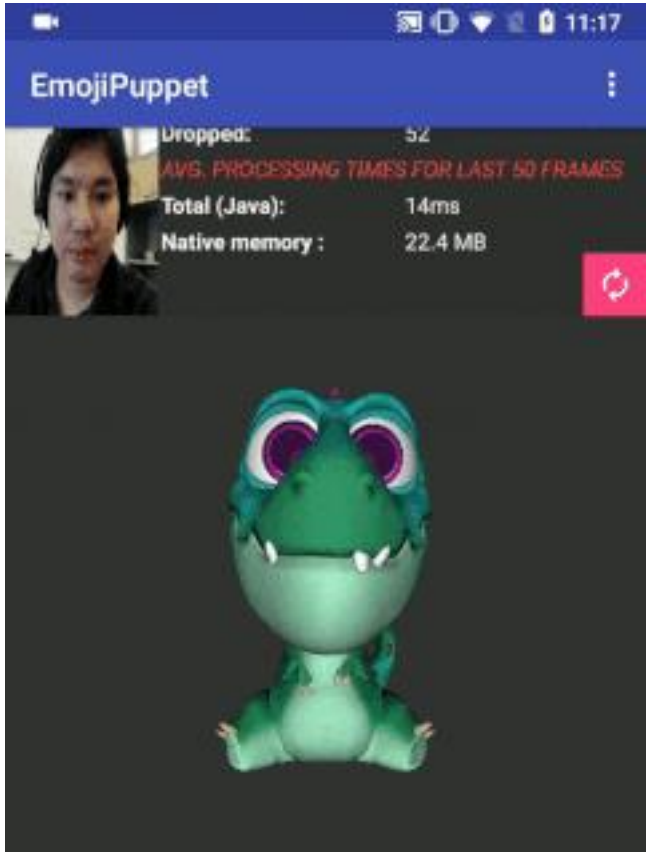
$$b_{\mathrm{lm}_x} = b_x + b_w * b_{\hat{\mathrm{lm}}_x};\ b_{\mathrm{lm}_y} = b_y + b_h * b_{\hat{\mathrm{lm}}_y}$$

B. Chaudhuri, N. Vesdapunt, B. Wang, Joint Face Detection and Facial Motion Retargeting for Multiple Faces , *CVPR 2019*

Network Performance for Test Images

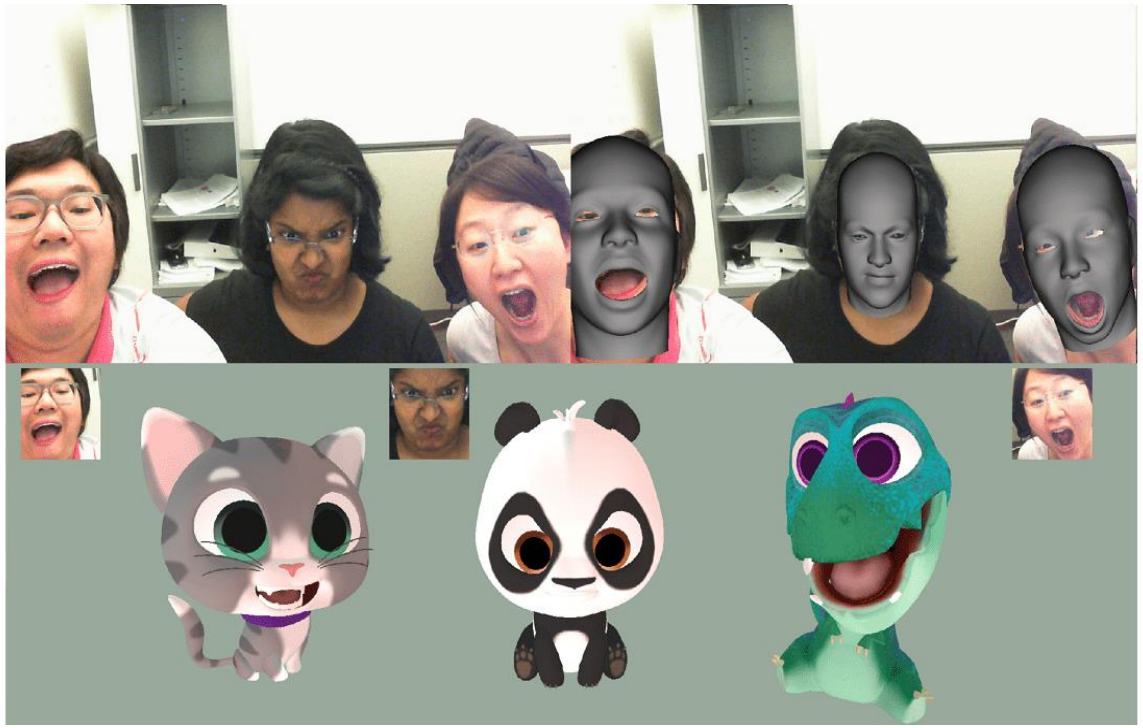


Results for Single Face Based Application

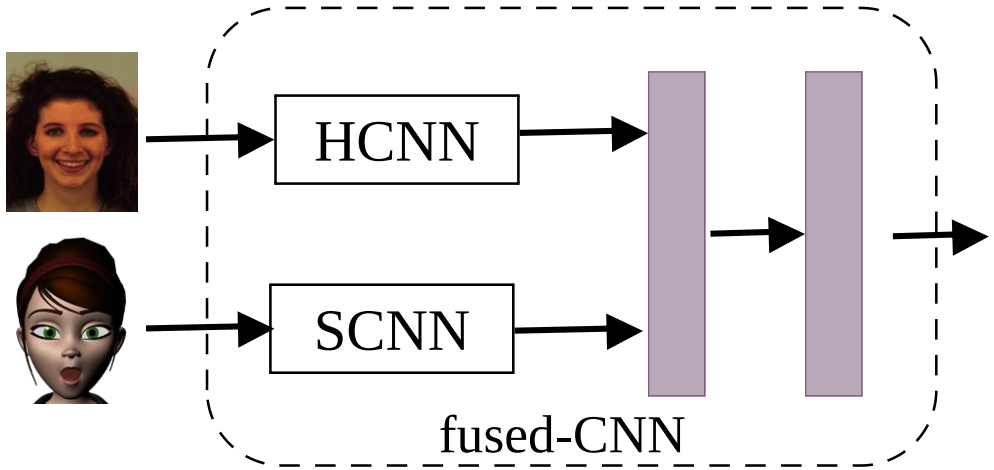


Hardware: Google Pixel 2

Live Performance Capture for Multiple Faces

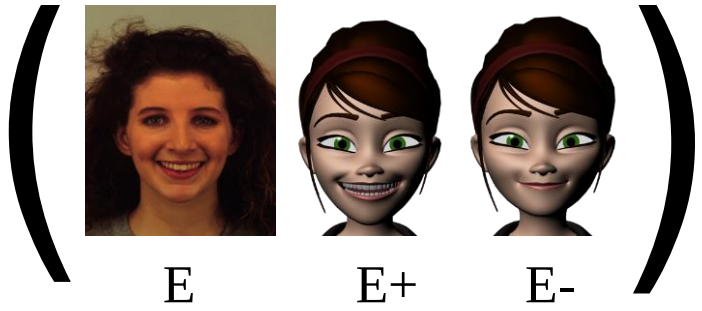


Example based Approach (semi-supervised)



Binary similarity score

Online triplet generation based on distance metric



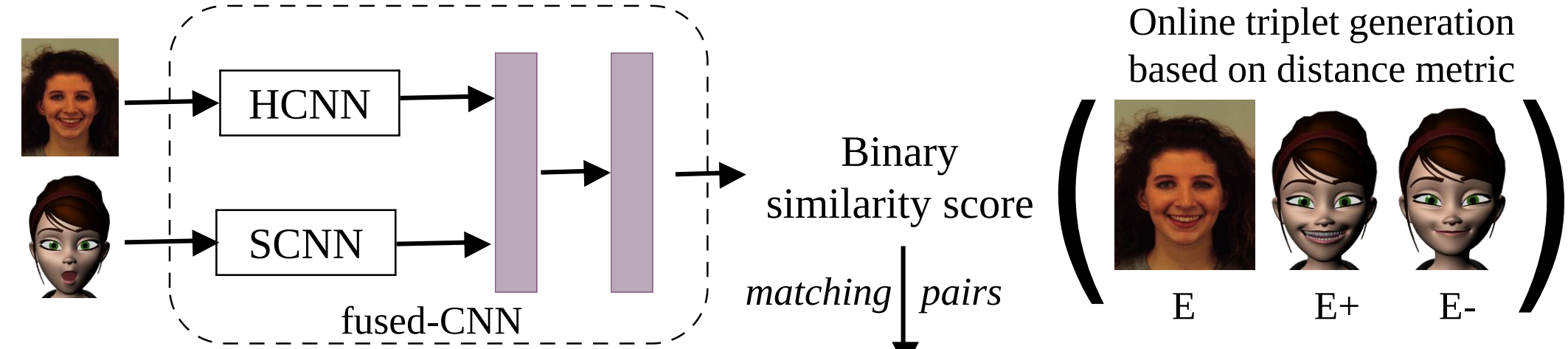
$$\phi_d = \alpha\,|\text{JS Distance}| + \beta\,|\text{Geometric Distance}|$$

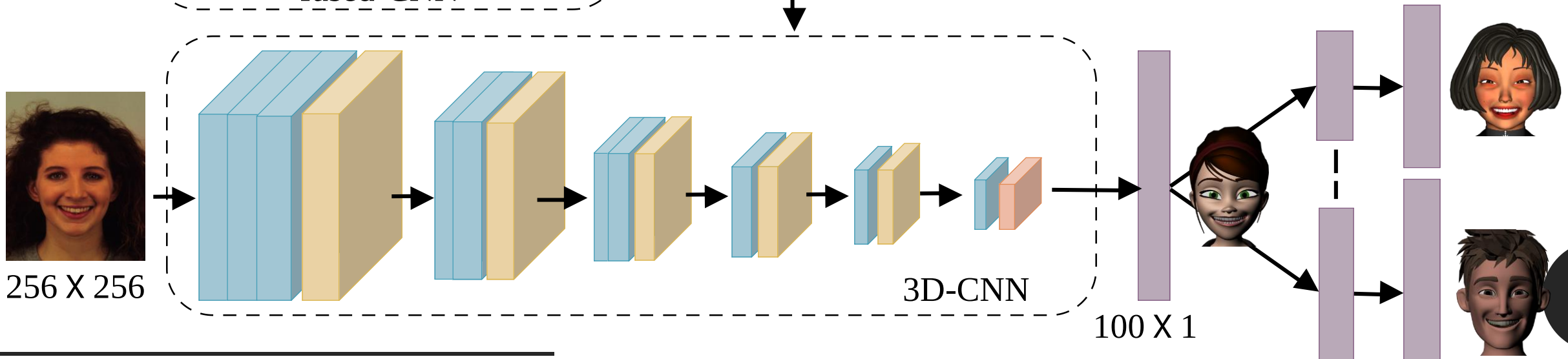
Expression feature vectors

Geometry feature vectors

D. Aneja, B. Chaudhuri, A. Colburn, G. Faigin, L. Shapiro, B. Mones, Learning to Generate 3D Stylized Character Expressions from Humans, *WACV 2018*

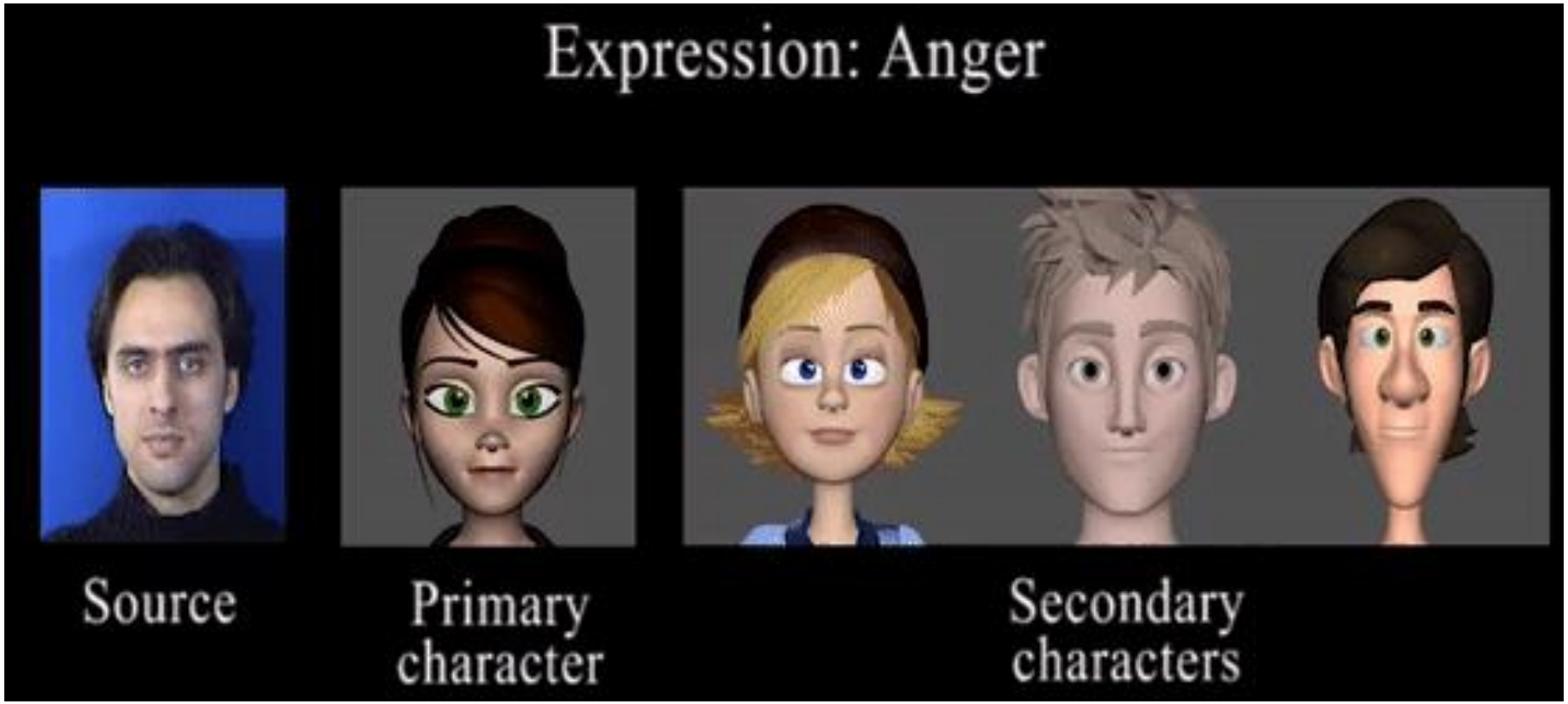
Example based Approach (semi-supervised)





D. Aneja, B. Chaudhuri, A. Colburn, G. Faigin, L. Shapiro, B. Mones, Learning to Generate 3D Stylized Character Expressions from Humans, *WACV 2018*

Results for Videos



Frame-by-frame transfer; jitter removed by temporal smoothening using Savitzky-Golay filter

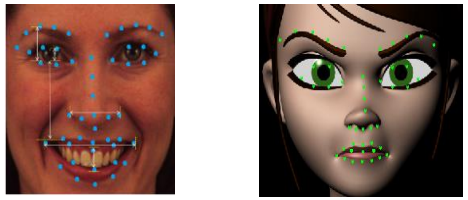
Example based Approach (unsupervised)

Aim:

- Use single network that directly regresses 3D vertices of character
- Generalize to a broader range of expressions

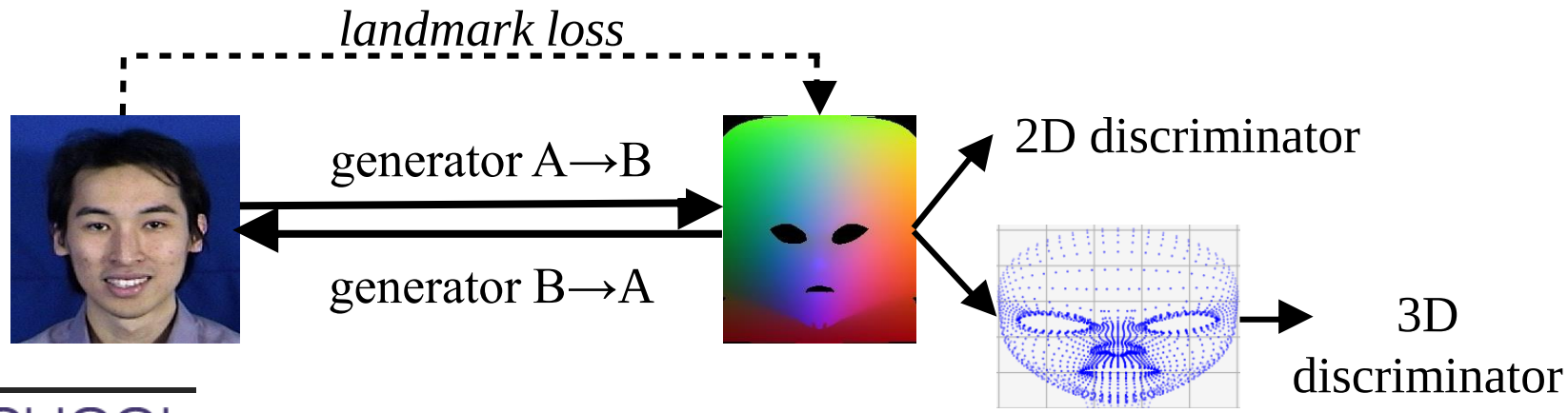
Example based Approach (unsupervised)

- Compute facial landmarks:

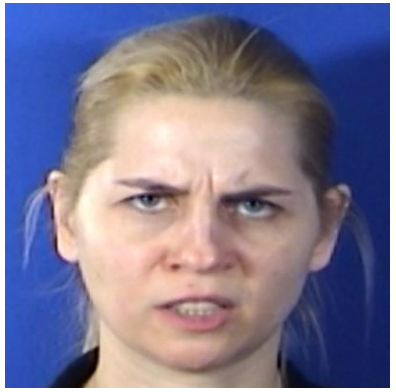
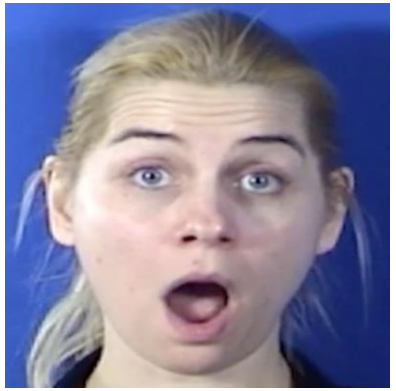


- Convert 3D model to 2D position map:

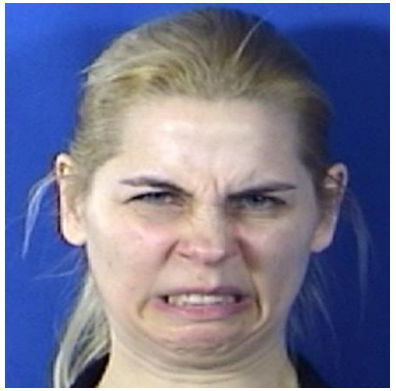
- Train CycleGAN:



Results

Input	Blendshape based	Semi-supervised Example based	Unsupervised Example based
			
			

Results

Input	Blendshape based	Semi-supervised Example based	Unsupervised Example based
			
			

Thank you