Age Progression/Regression by Conditional Adversarial Autoencoder

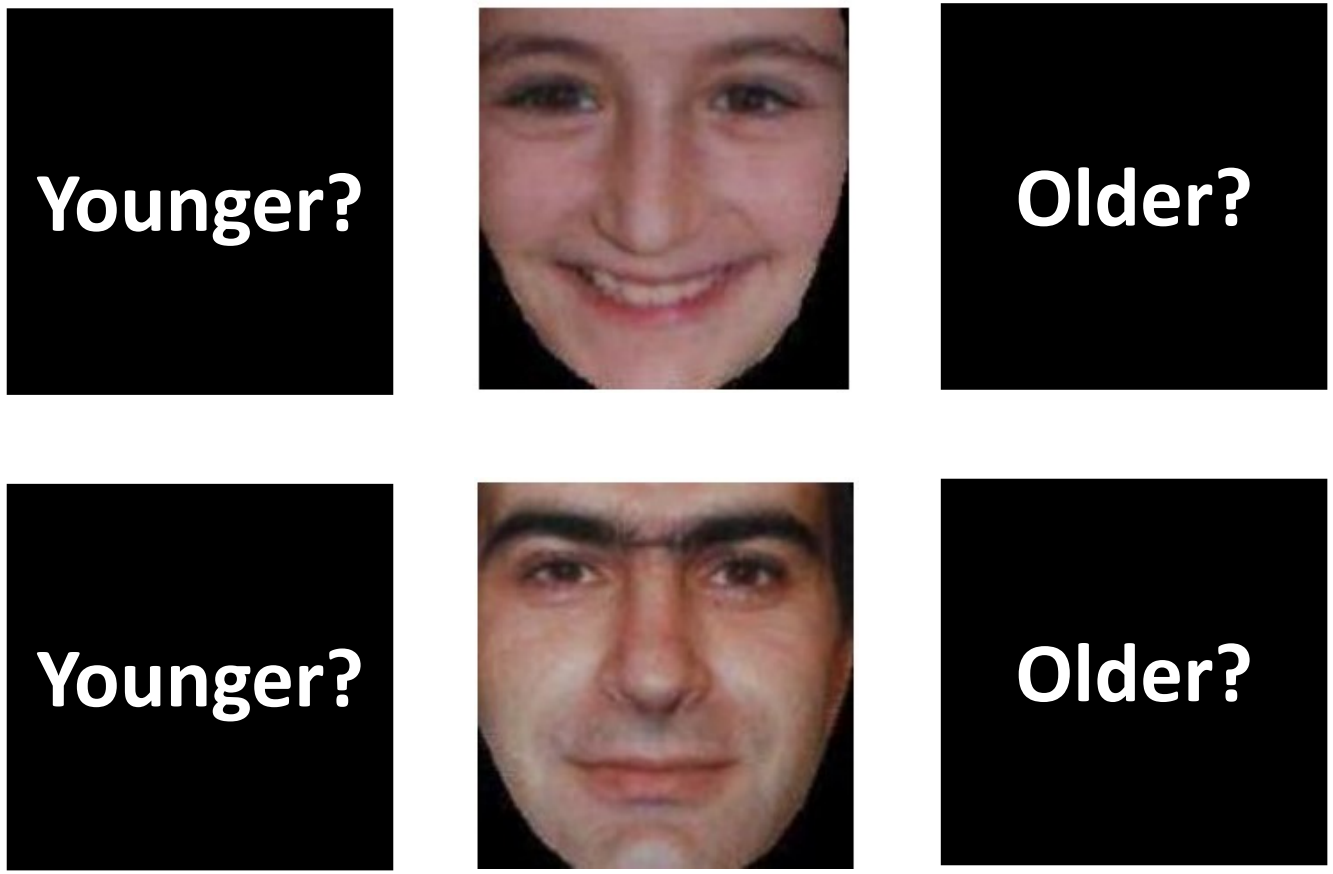
Zhifei Zhang, Yang Song, and Hairong Qi

Outline

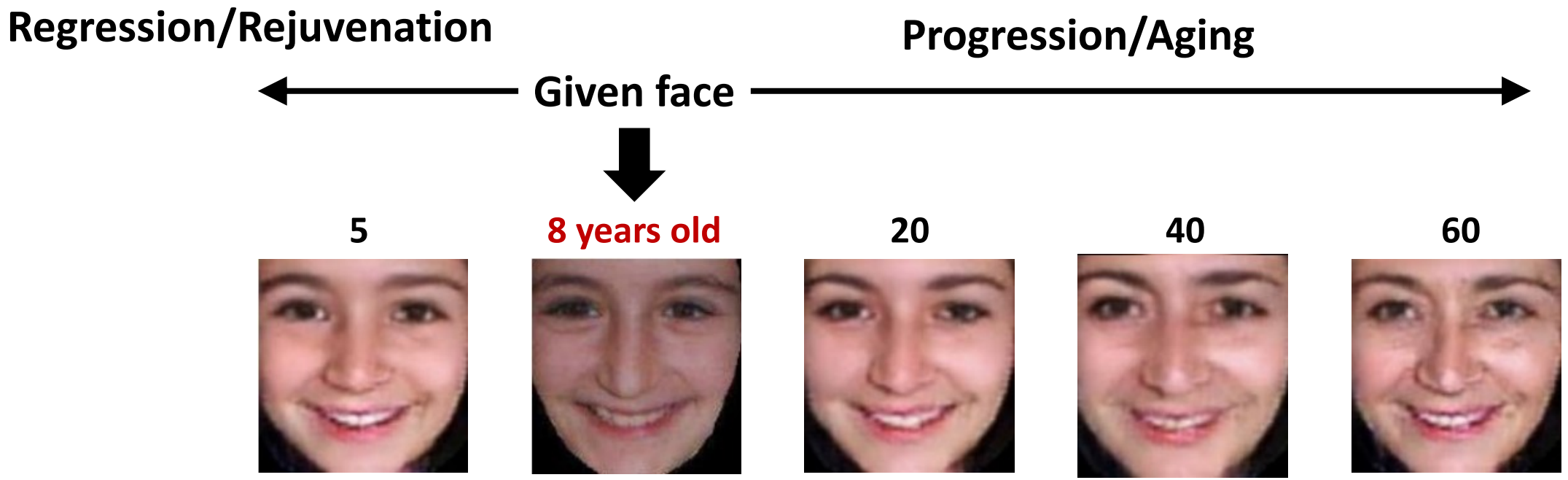
1. Introduction
2. Traversing on the Manifold
3. Approach
4. Experimental Evaluation

Introduction

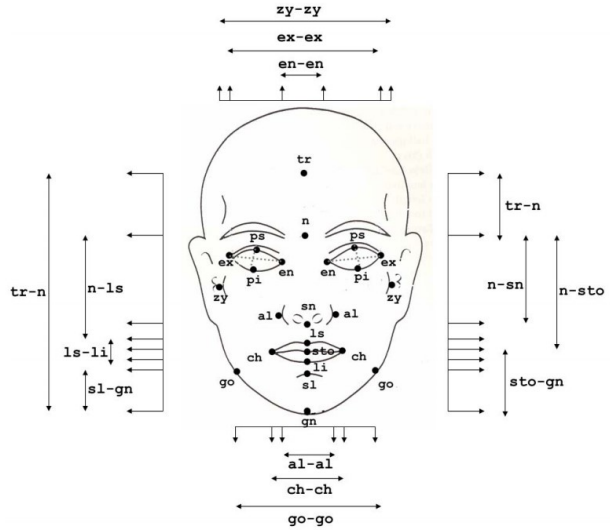
"If I provide you a face image of mine (without telling you the actual age when I took the picture) and a large amount of face images that I crawled (containing labeled faces of different ages but not necessarily paired), can you show me what I would look like when I am 80 or what I was like when I was 5?"

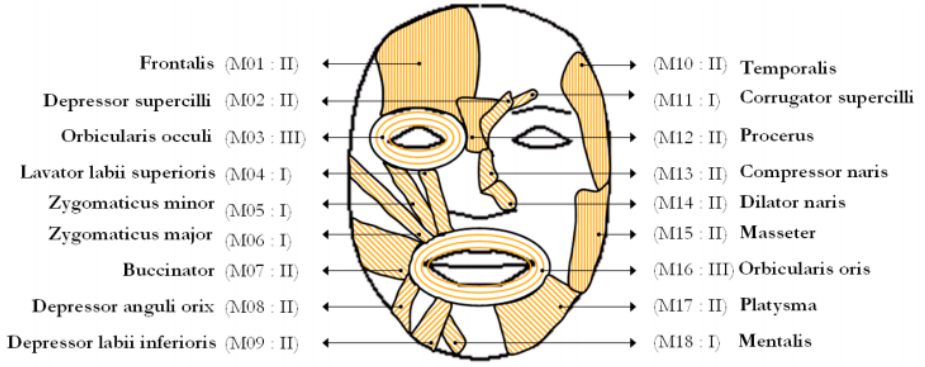


Introduction



Introduction





Face Aging

Physical Model

Ramanathan, et al., 2006 (CVPR), 2008 (AFGR)

Suo, et al., 2010 (PAMI)

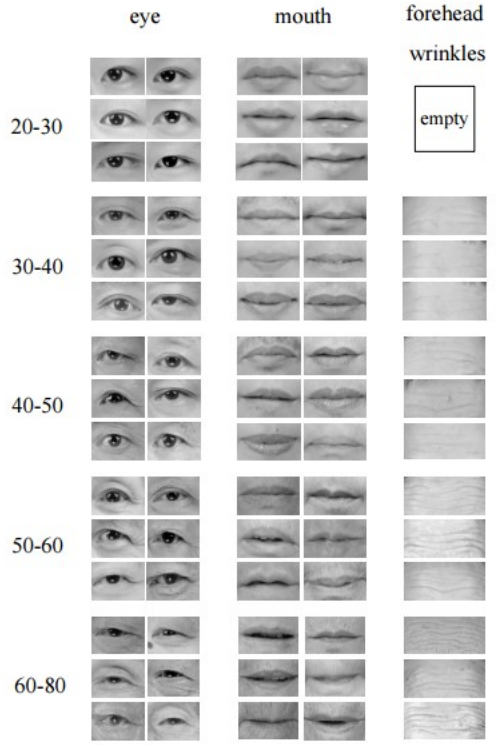
Prototype-based

Kemelmacher, et al., 2014 (CVPR)

Shu, et al., 2015 (ICCV)

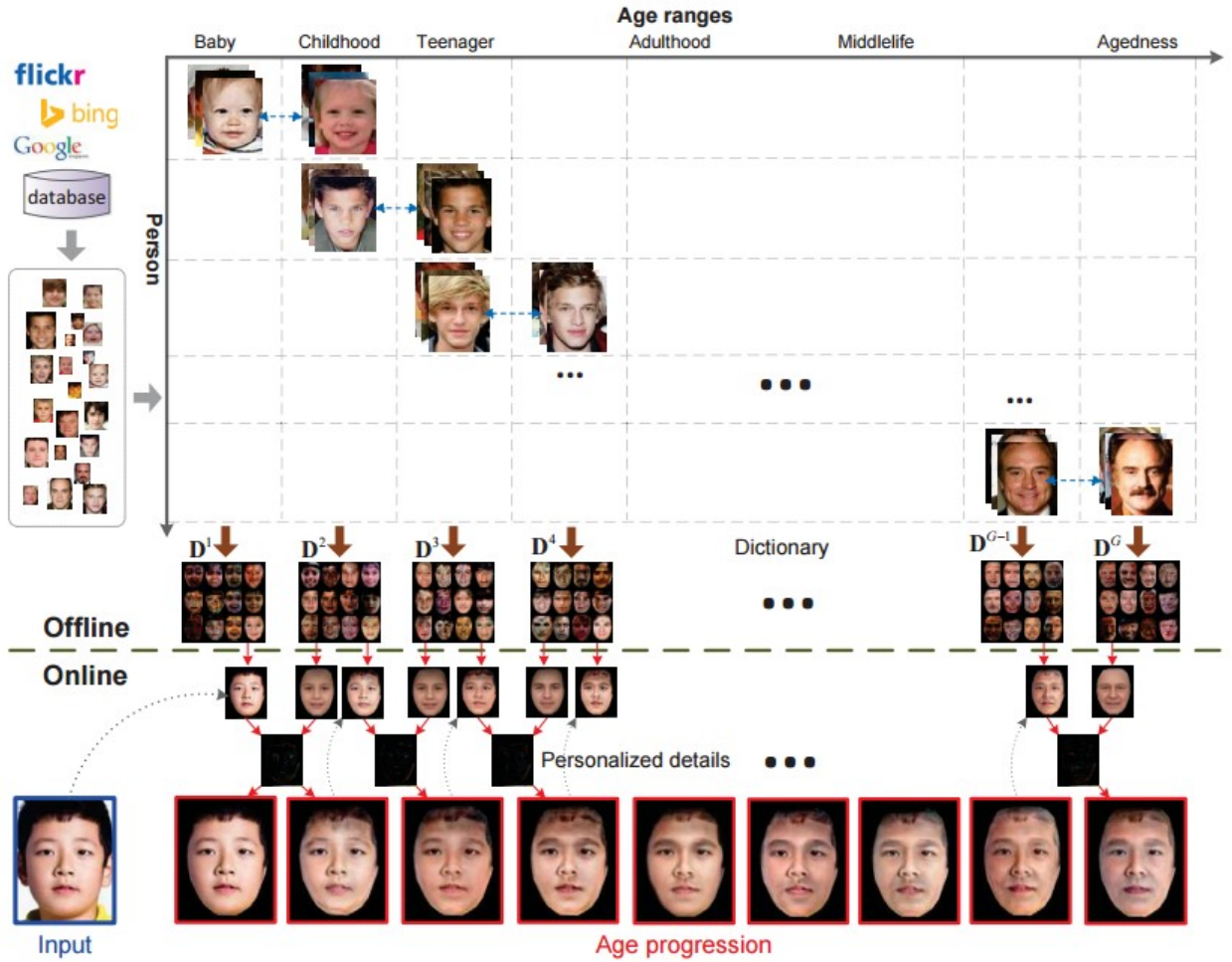
Deep learning

Wang, et al., 2016 (CVPR)



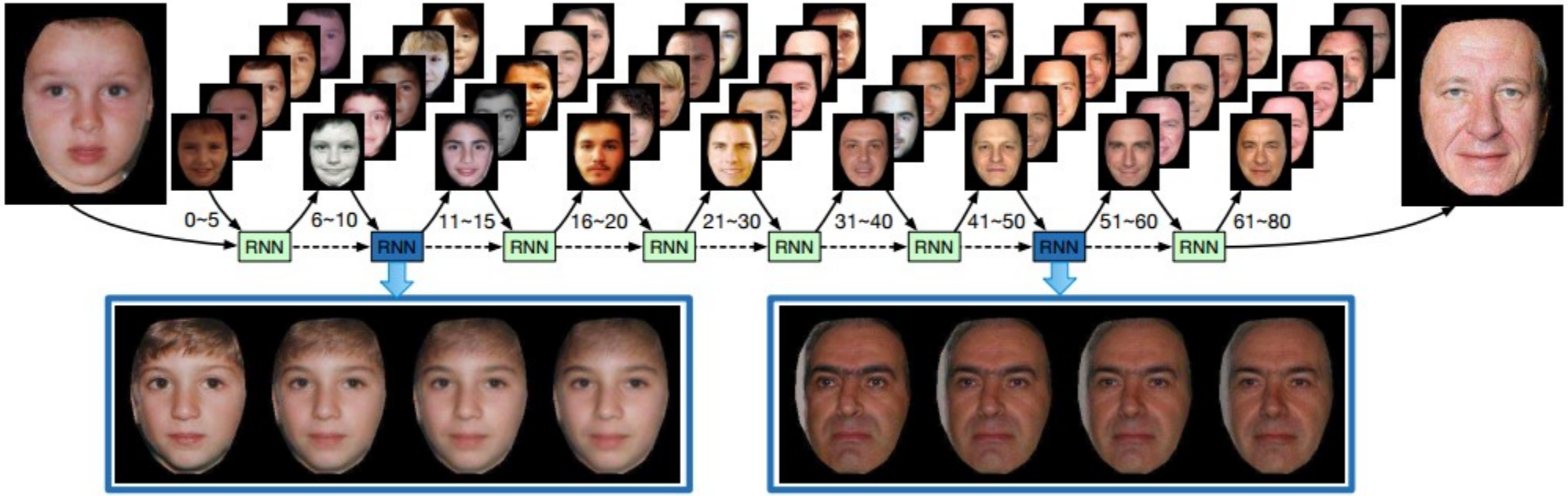
Introduction

Personalized Age Progression with Aging Dictionary [Shu, et al., 2015 (ICCV)]



Introduction

Recurrent Face Aging [Wang, et al., 2016 (CVPR)]

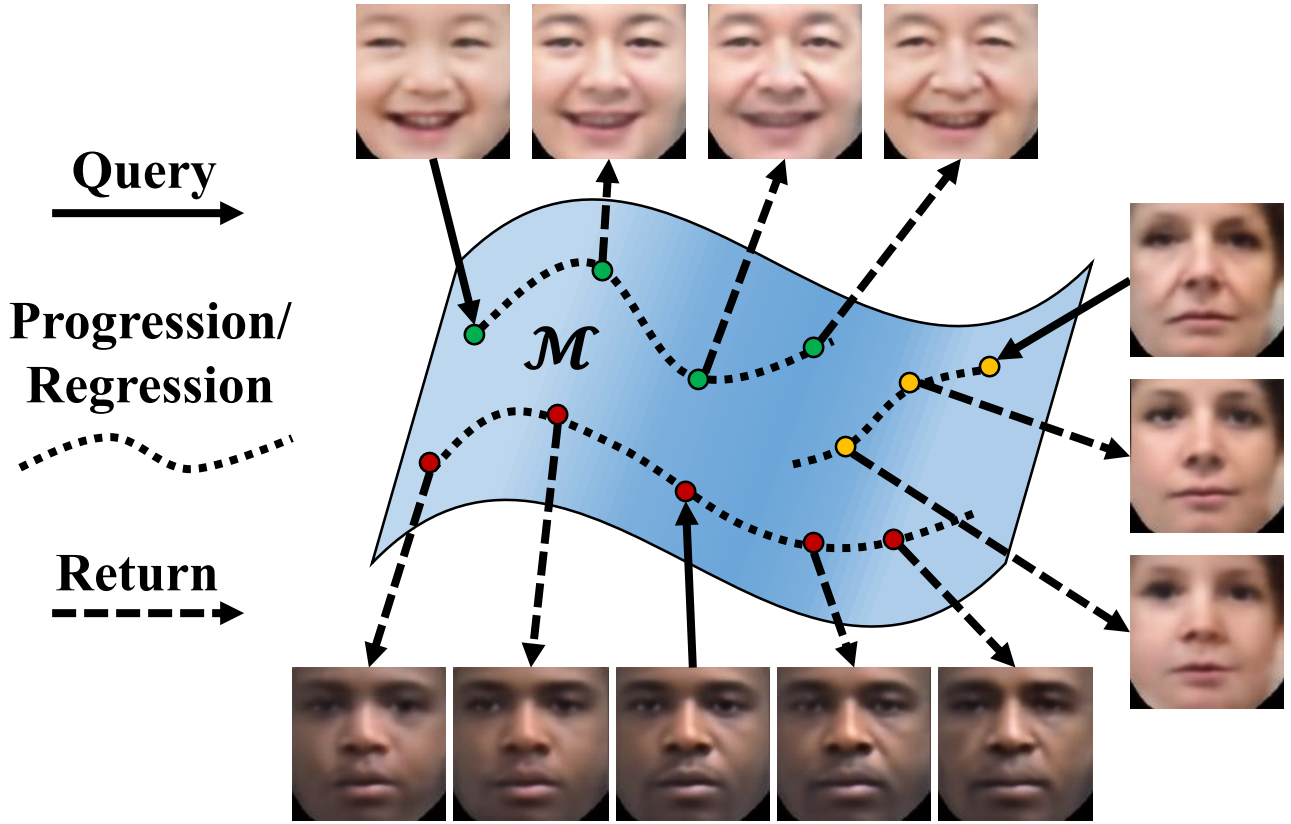


Introduction

Our idea [Zhang, et al., 2017 (CVPR)]

Assumptions:

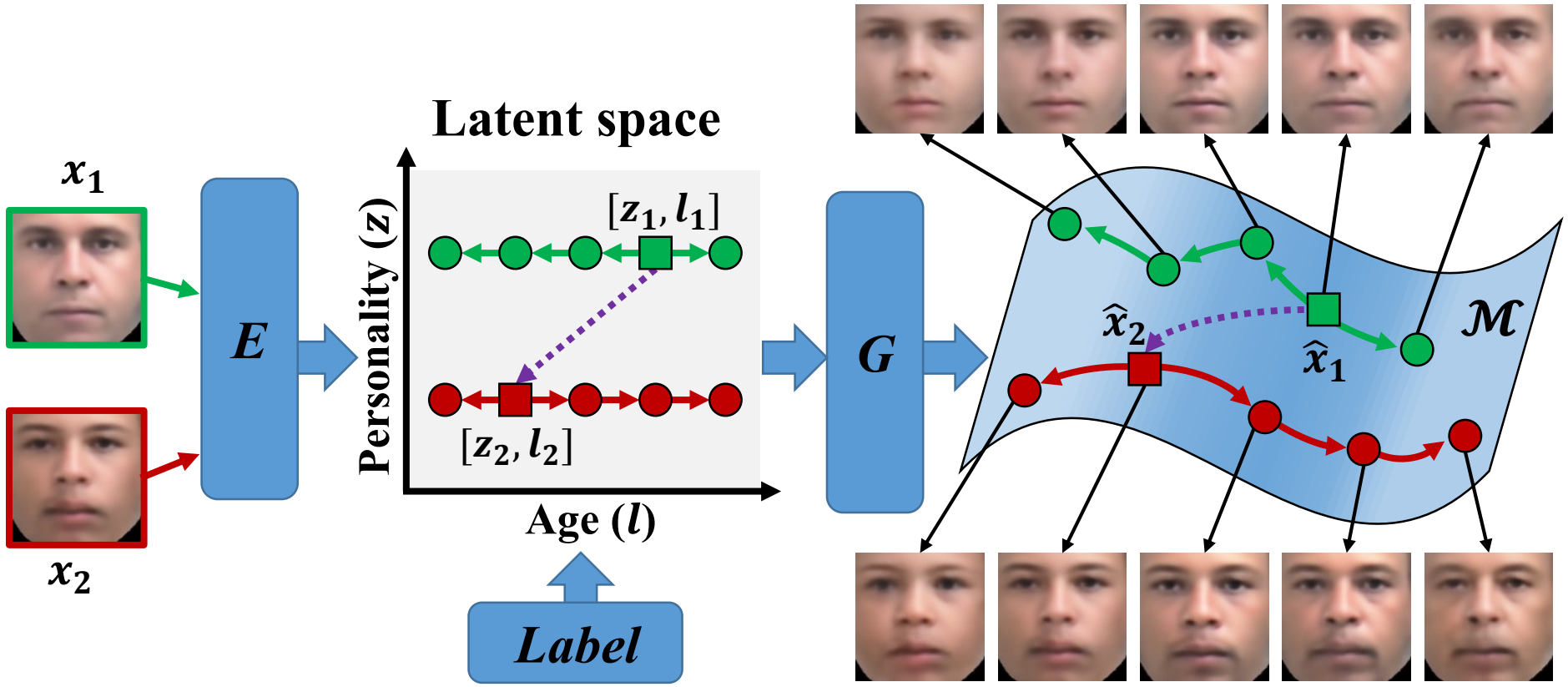
- **The faces lie on a manifold (*M*)**
- **Clustered by ages and personality**
- **Traversing on the manifold corresponds to age/personality transformation.**



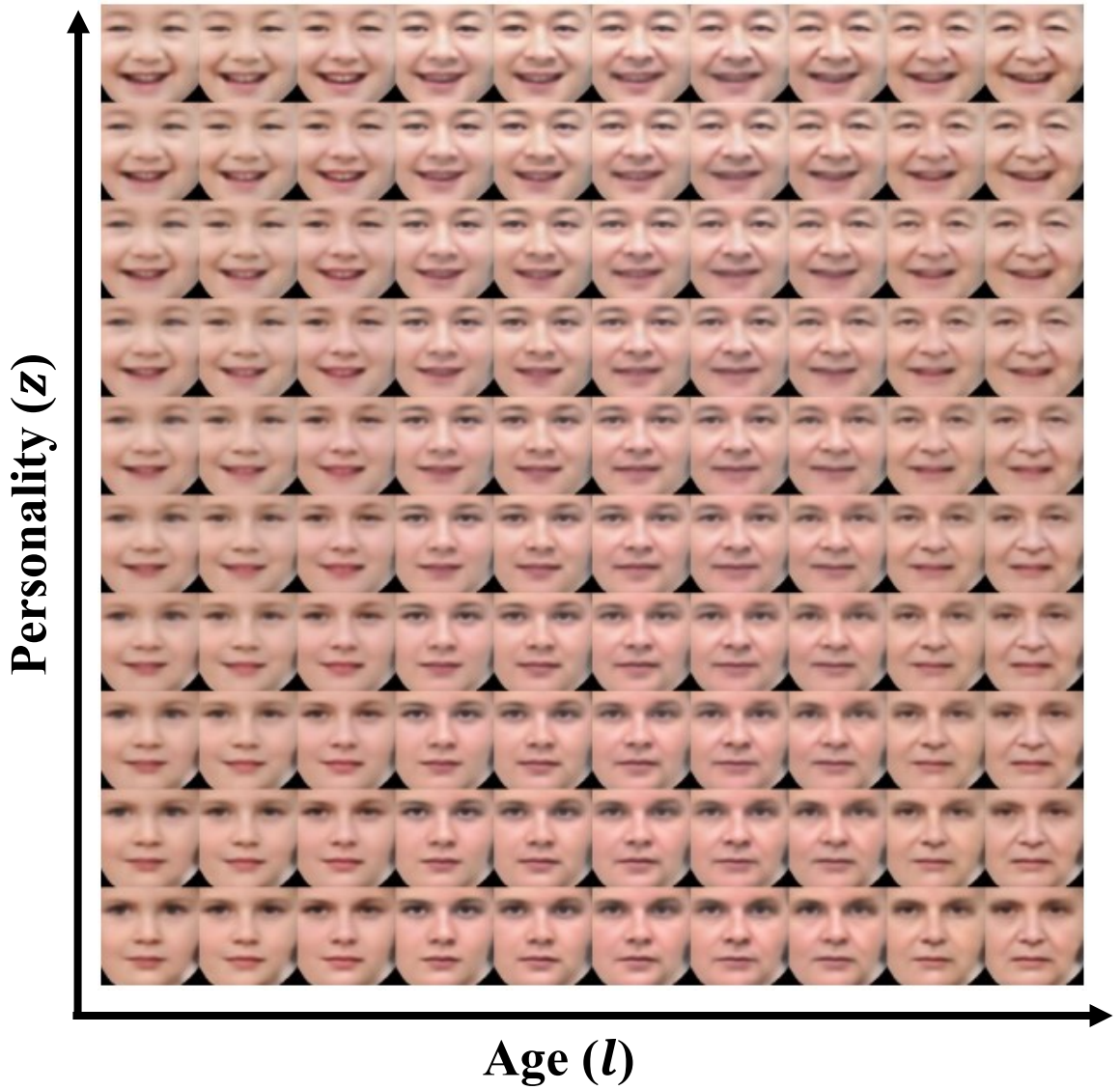
Introduction

	Physical Model	Prototype (Dictionary)	Deep Learning (RNN)	Ours
Reality	High	Low	High	High
Group-wise learning	x	x	x	
Label required during testing	x	x	x	
Require long/short age span	x	x		
Complicate in modeling	x			

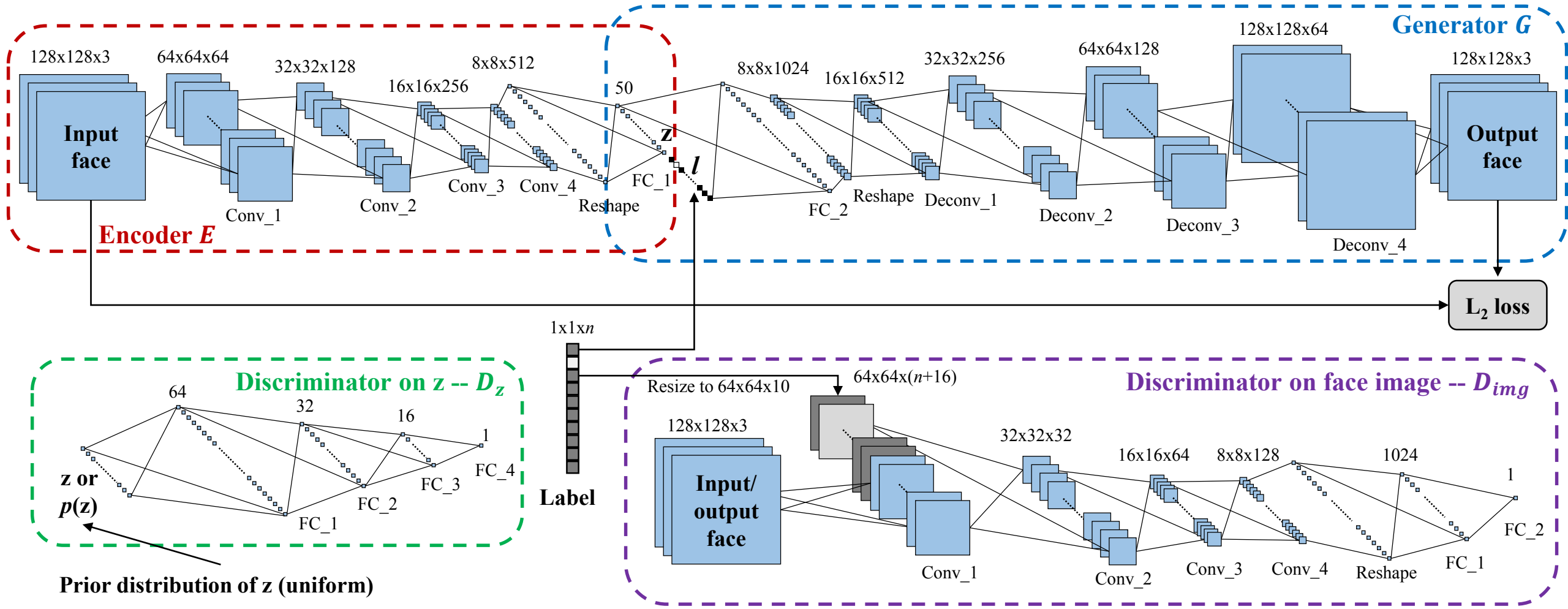
Traversing on the Manifold



Traversing on the Manifold



Approach


Generator G
Encoder E
Input face
128x128x3
64x64x64
32x32x128
16x16x256
8x8x512
50
Conv_1
Conv_2
Conv_3
Conv_4
Reshape
FC_1
z
l
8x8x1024
16x16x512
32x32x256
64x64x128
128x128x64
128x128x3
FC_2
Reshape
Deconv_1
Deconv_2
Deconv_3
Deconv_4
Output face
L2 loss
1x1xn
Label
Discriminator on z -- Dz
64
32
16
1
FC_1
FC_2
FC_3
FC_4
z or p(z)
Prior distribution of z (uniform)
Discriminator on face image -- Dimg
Resize to 64x64x10
64x64x(n+16)
128x128x3
Input/ output face
Conv_1
32x32x32
Conv_2
16x16x64
Conv_3
8x8x128
Conv_4
1024
Reshape
FC_1
1
FC_2

Approach

Reconstruction error

Total variation

$$\min_{E,G} \max_{D_z, D_{img}} \lambda \mathcal{L}\left(x, G(E(x), l)\right) + \gamma TV(G(E(x), l))$$

D on *z*

$$+\mathbb{E}_{z^* \sim p(\mathbf{z})}\left[\log D_z(z^*)\right]$$
$$+\mathbb{E}_{x \sim p_{data}(\mathbf{x})}\left[\log(1 - D_z(E(x)))\right] \quad (5)$$

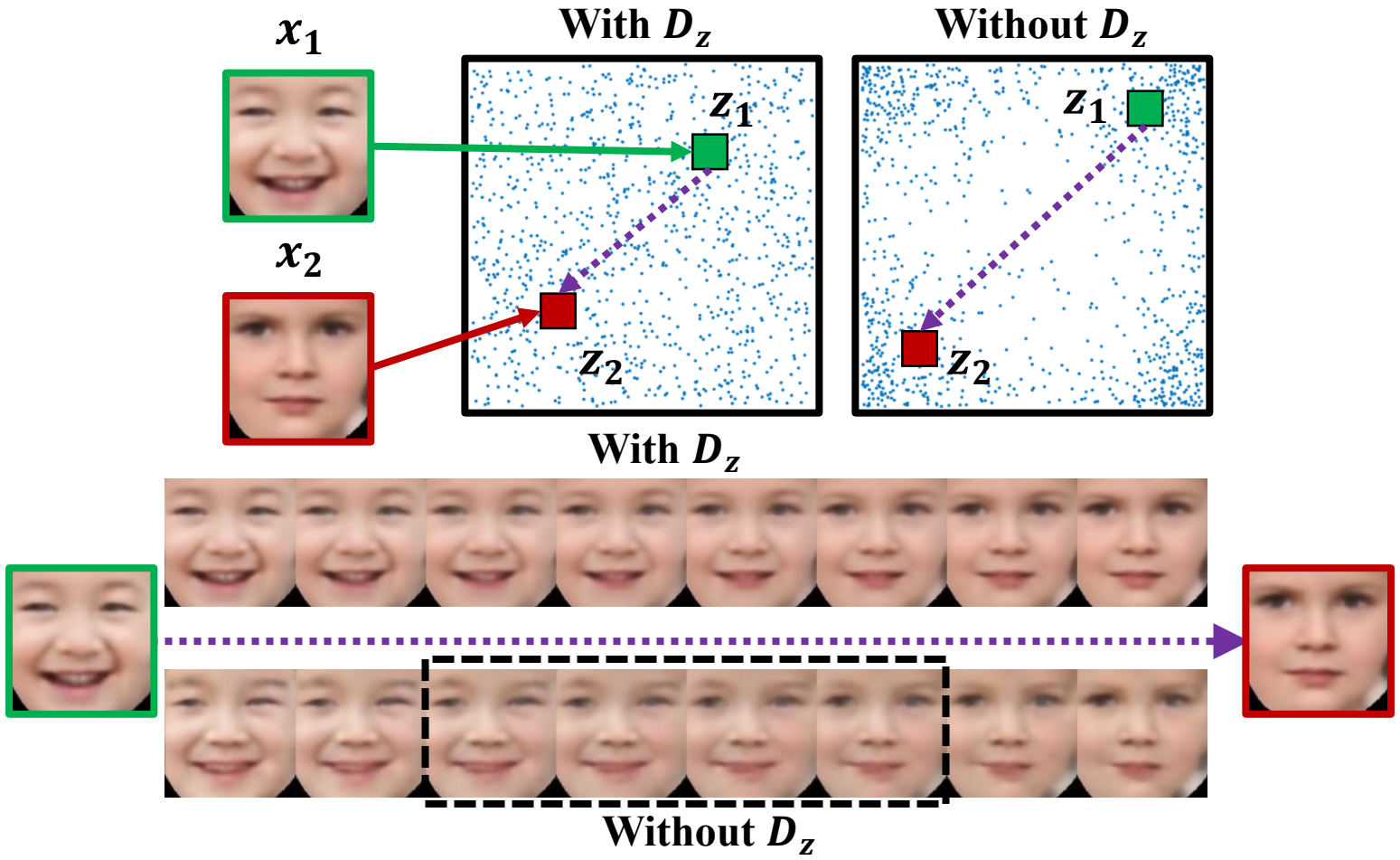
D on image

$$+\mathbb{E}_{x,l \sim p_{data}(\mathbf{x},\mathbf{l})}\left[\log D_{img}(x, l)\right]$$
$$+\mathbb{E}_{x,l \sim p_{data}(\mathbf{x},\mathbf{l})}\left[\log(1 - D_{img}(G(E(x), l)))\right],$$

where $TV(\cdot)$ denotes the total variation which is effective in removing the ghosting artifacts. The coefficients λ and γ balance the smoothness and high resolution.

Approach

Effect of the Discriminator on *z*

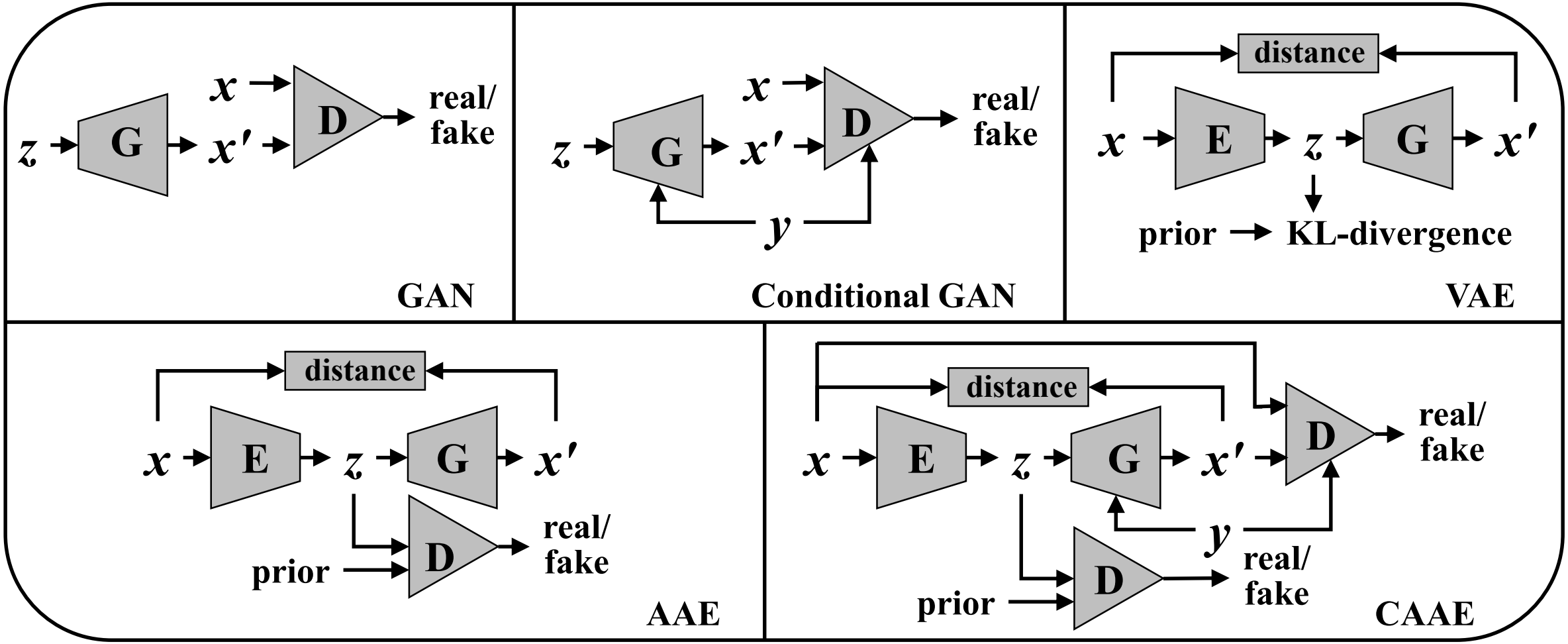


Approach

Effect of the Discriminator on images

Approach

Comparison to Related Structures



Experimental Evaluation

Data Collection:

- **MORPH dataset. 55,000 faces of 13,000 subjects from 16 to 77 years old**
- **CACD dataset. 163,446 faces of 2,000 subjects from 16 to 62 years old**
- **Search the keywords: baby, boy, teenager, 15 years old, etc. and coll**

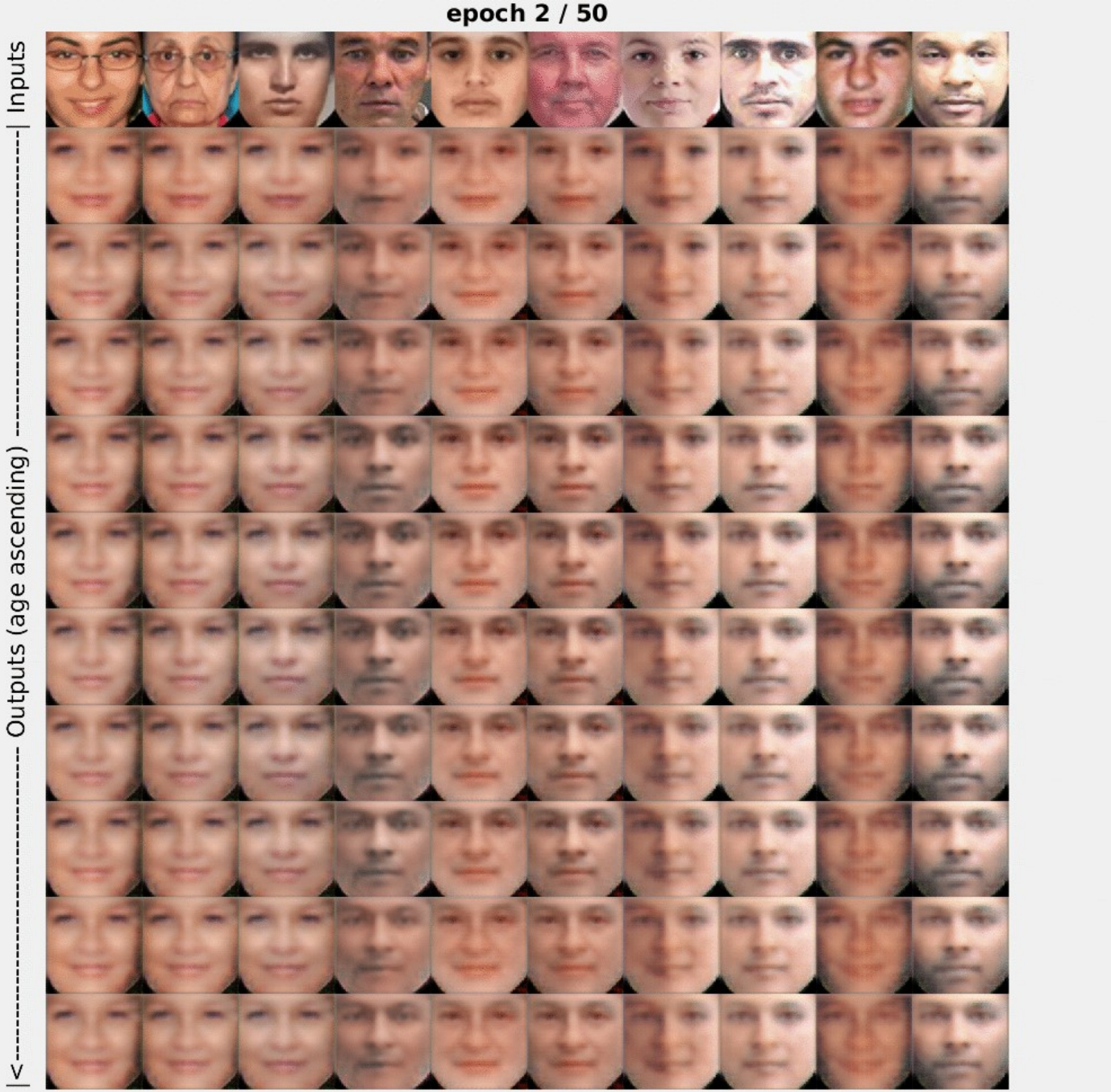
Experimental Evaluation

Raw Images

Alignment

Experimental Evaluation

Training process



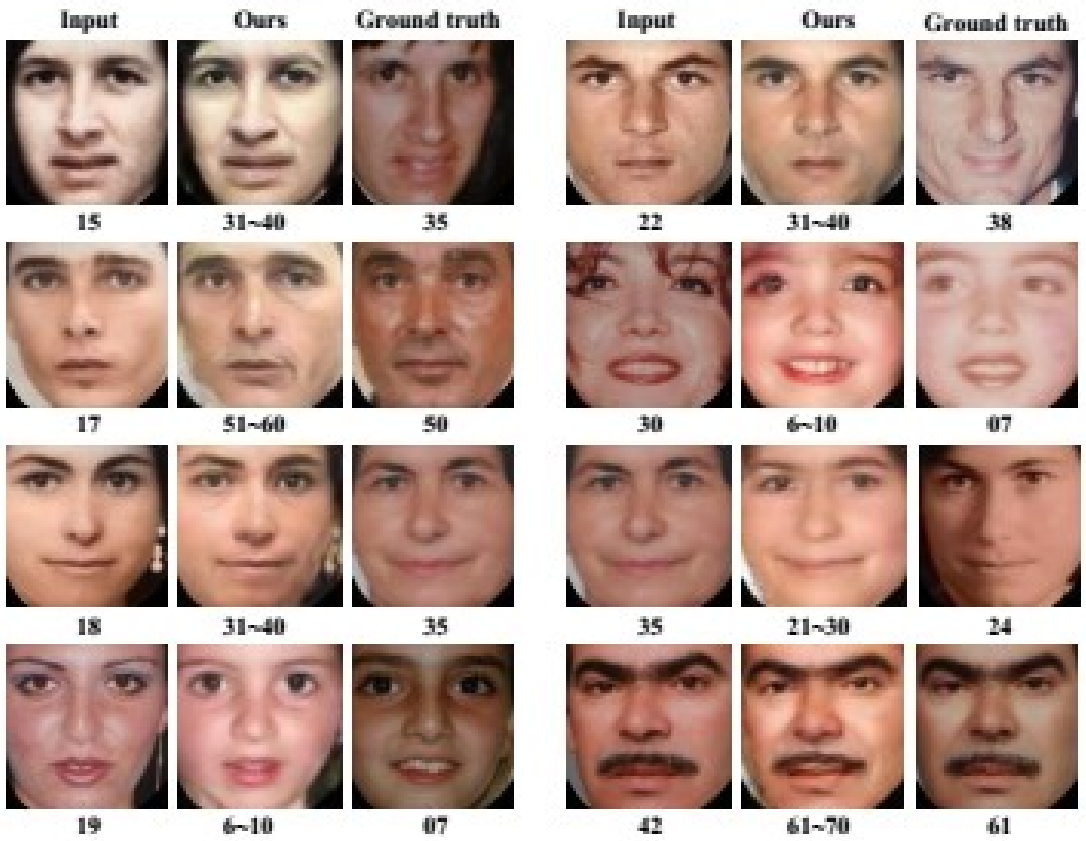
Experimental Evaluation

Qualitative Comparison

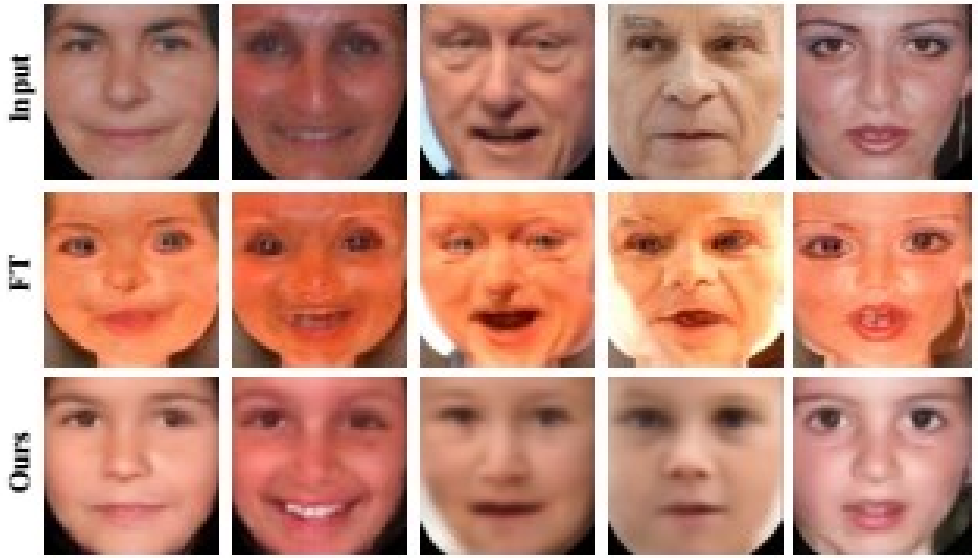
Prior: The BEST result achieved by existing works

Experimental Evaluation

Qualitative Comparison



No existing work reported regression/rejuvenation results



Experimental Evaluation

Quantitative Comparison

We struggled with this comparison because there is still no a good metric to measure reality of images.

1. Survey for comparison with prior work

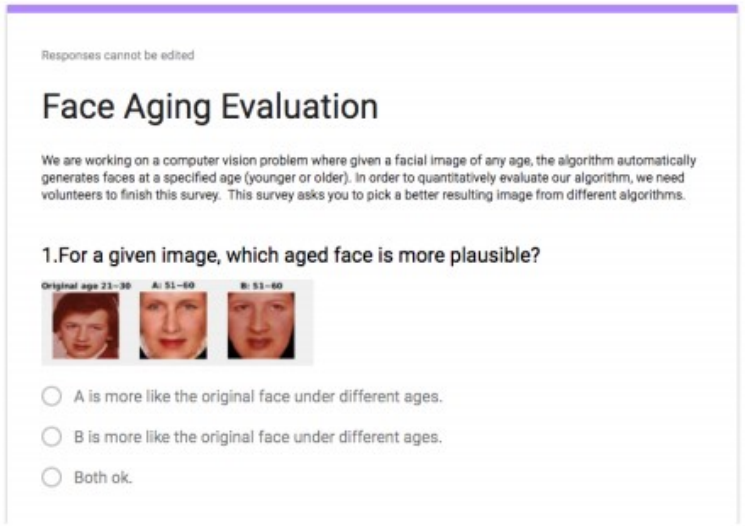


Figure 1. Our Google survey form to compare with ground truth. We show the input image (left), our result (middle), and ground truth (right). Note that in this survey, we didn't indicate which one is ground truth and which one is generated image. To avoid similarity bias, we compare with generated image under the same age group of the ground truth.

2. Survey for comparison with ground truth

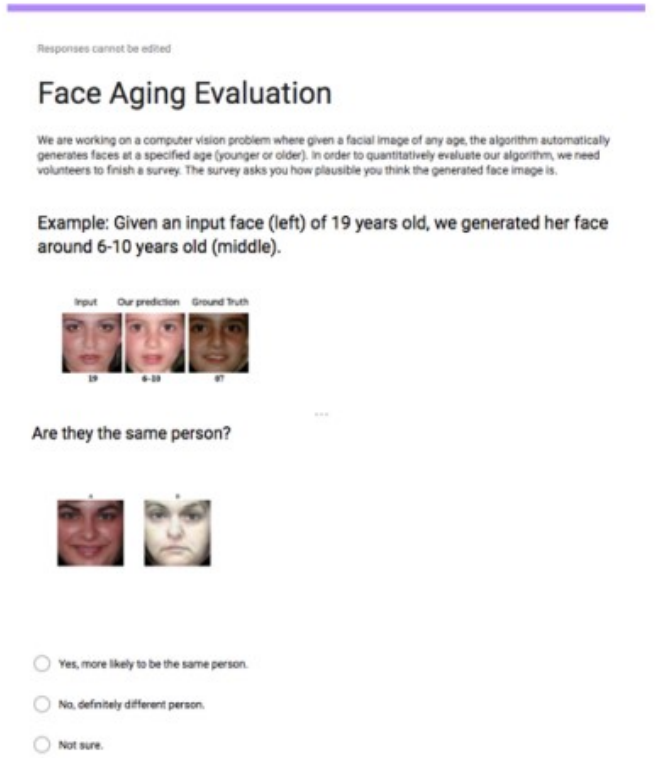


Figure 2. Our Google survey form to compare with prior work. We show the our result and prior work; Note that we didn't specific which one is ours and which is prior work. We randomly order ours and prior work.

Experimental Evaluation

- 47 volunteers
- Each is randomly assigned 45 out of 235 results
- 1,508 voles in total

1. Survey for comparison with prior work

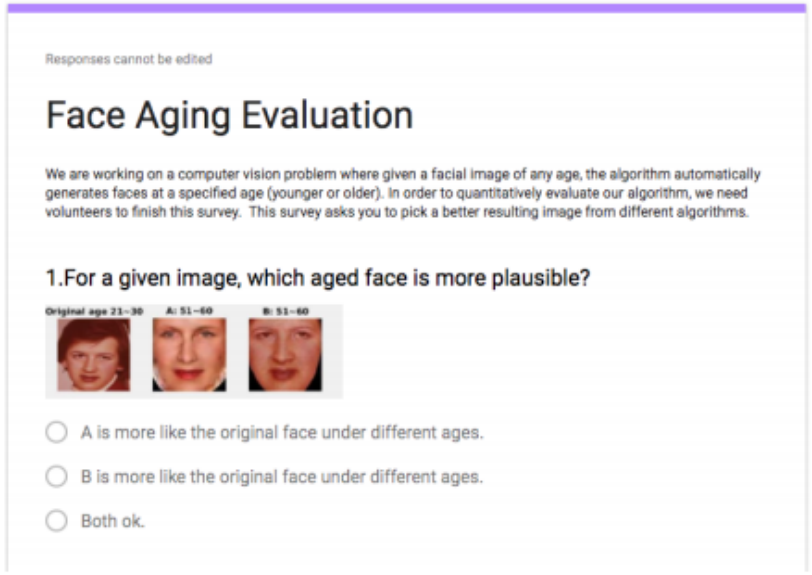
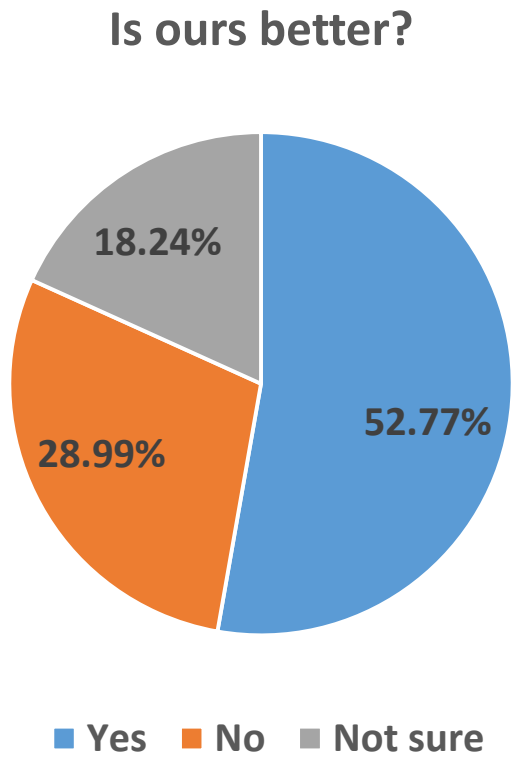


Figure 1. Our Google survey form to compare with ground truth. We show the input image (left), our result (middle), and ground truth (right). Note that in this survey, we didn't indicate which one is ground truth and which one is generated image. To avoid similarity bias, we compare with generated image under the same age group of the ground truth.



Experimental Evaluation

- 63 volunteers
- Each is randomly assigned 45 out of 865 results
- 3,208 voles in total

2. Survey for comparison with ground truth

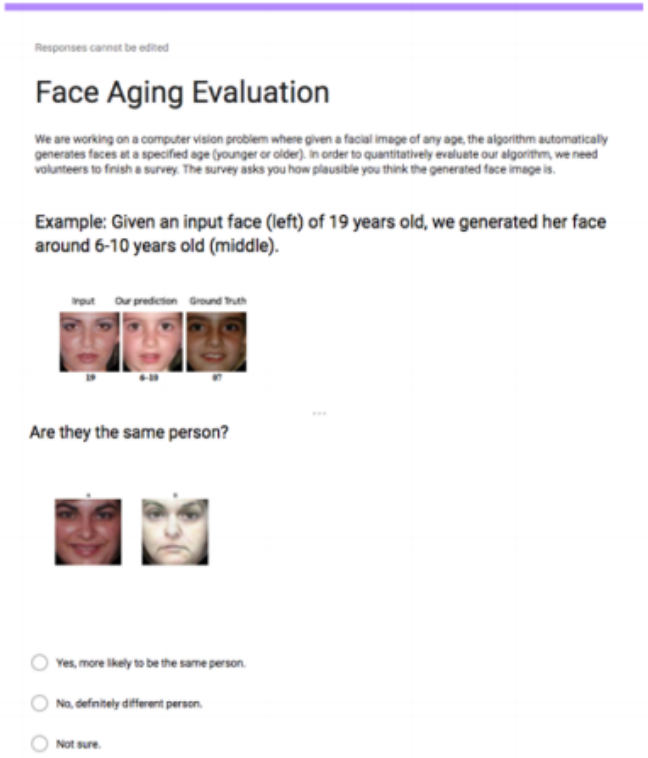
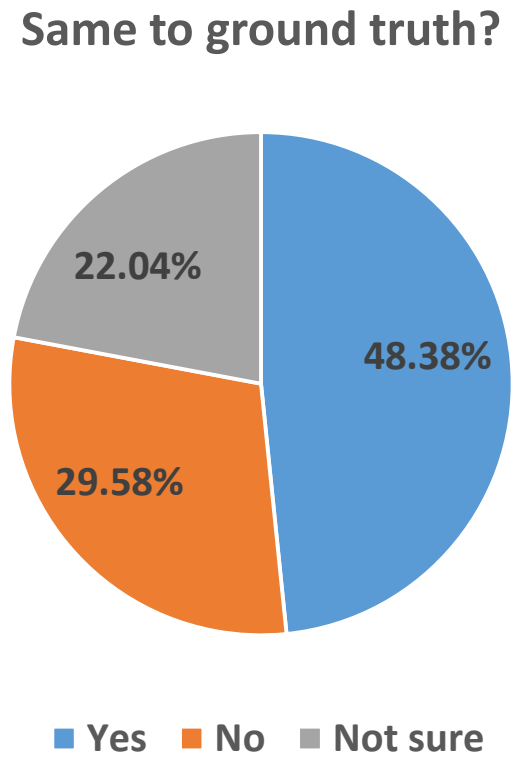
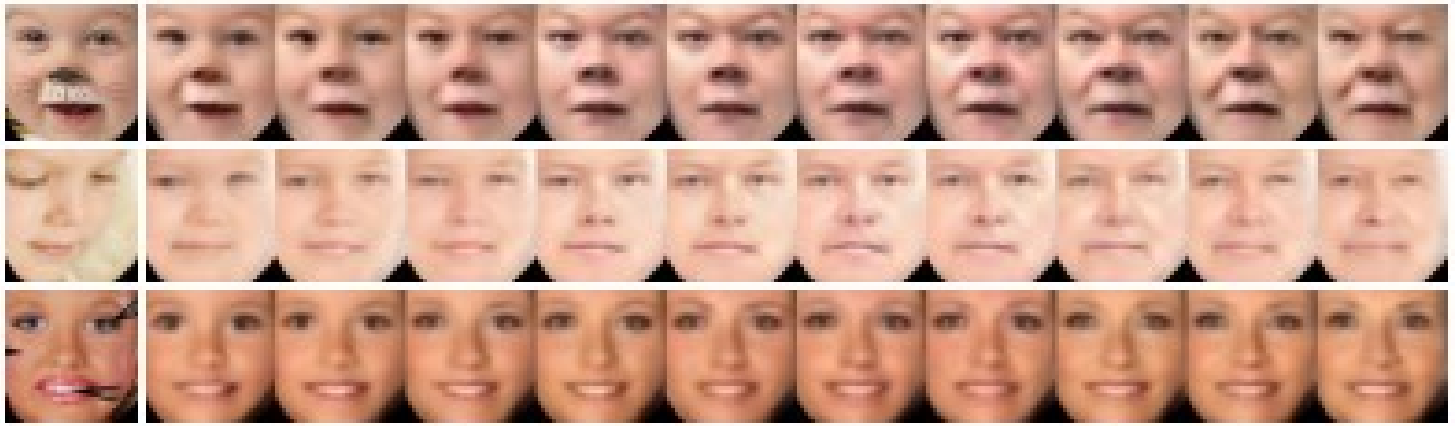


Figure 2. Our Google survey form to compare with prior work. We show the our result and prior work; Note that we didn't specific which one is ours and which is prior work. We randomly order ours and prior work.



Experimental Evaluation

Tolerance to Pose, Expression, and Occlusion

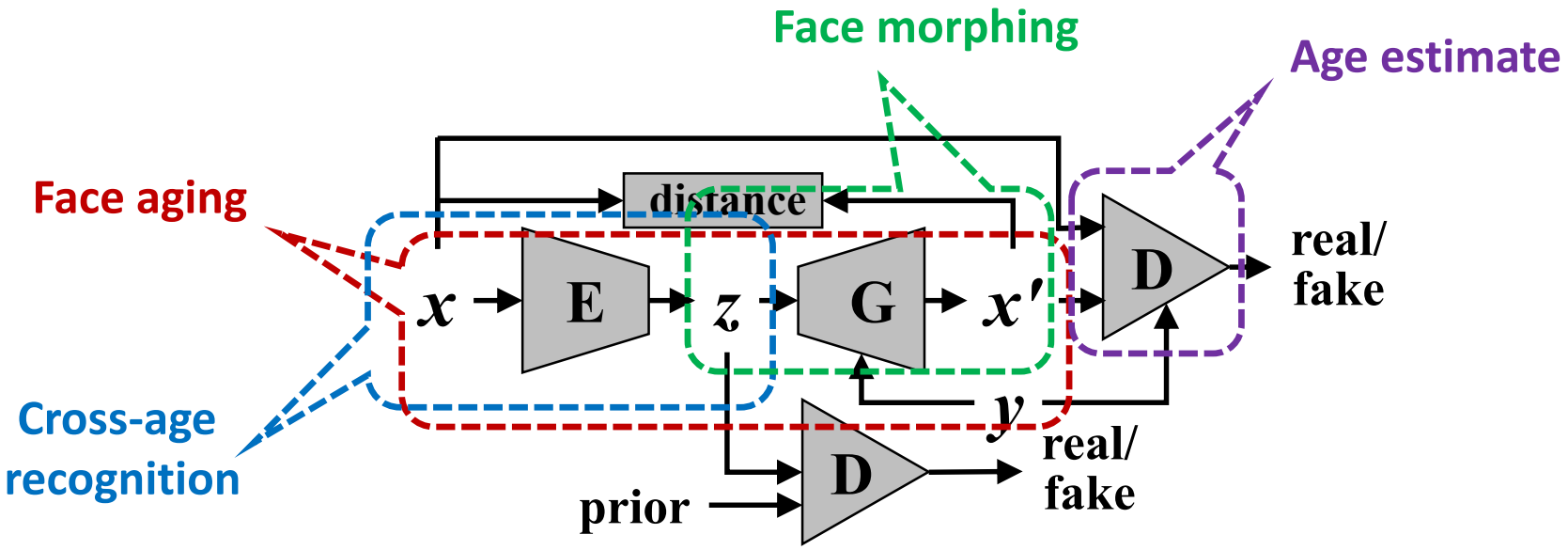


TensorFlow Code is available at:

- Bitbucket: https://bitbucket.org/aicip/face-aging-caae
- Github: https://zzutk.github.io/Face-Aging-CAAE

Conclusion

Potential Framework for face-age related applications




thank
you