Controllable Continuous Gaze Redirection

Weihao Xia[1] Yujiu Yang[1] Jing-Hao Xue[2] Wensen Feng[3]

[1]Tsinghua Shenzhen International Graduate School, Tsinghua University [2]Department of Statistical Science, University College London [3]Beijing University of Chemical Technology

Problem Definition and Contribution

Goal: Given two gaze images with different attributes, our goal is to redirect the eye gaze of one person into any gaze direction depicted in the reference image or to generate con- tinuous intermediate results.

Motivations:

- We present a novel framework achieving both precise gaze redirection and continuous gaze interpolation. The two different tasks can be readily controlled by altering a control vector.
- We learn a well-disentangled and hierarchically-organized latent space by decoupling the related gaze attributes and equipping with the efficacy of one-shot and diversity.
- We contribute a high-quality gaze dataset, which contains a large range of gaze directions and diversity on eye shapes, glasses, ages and genders.

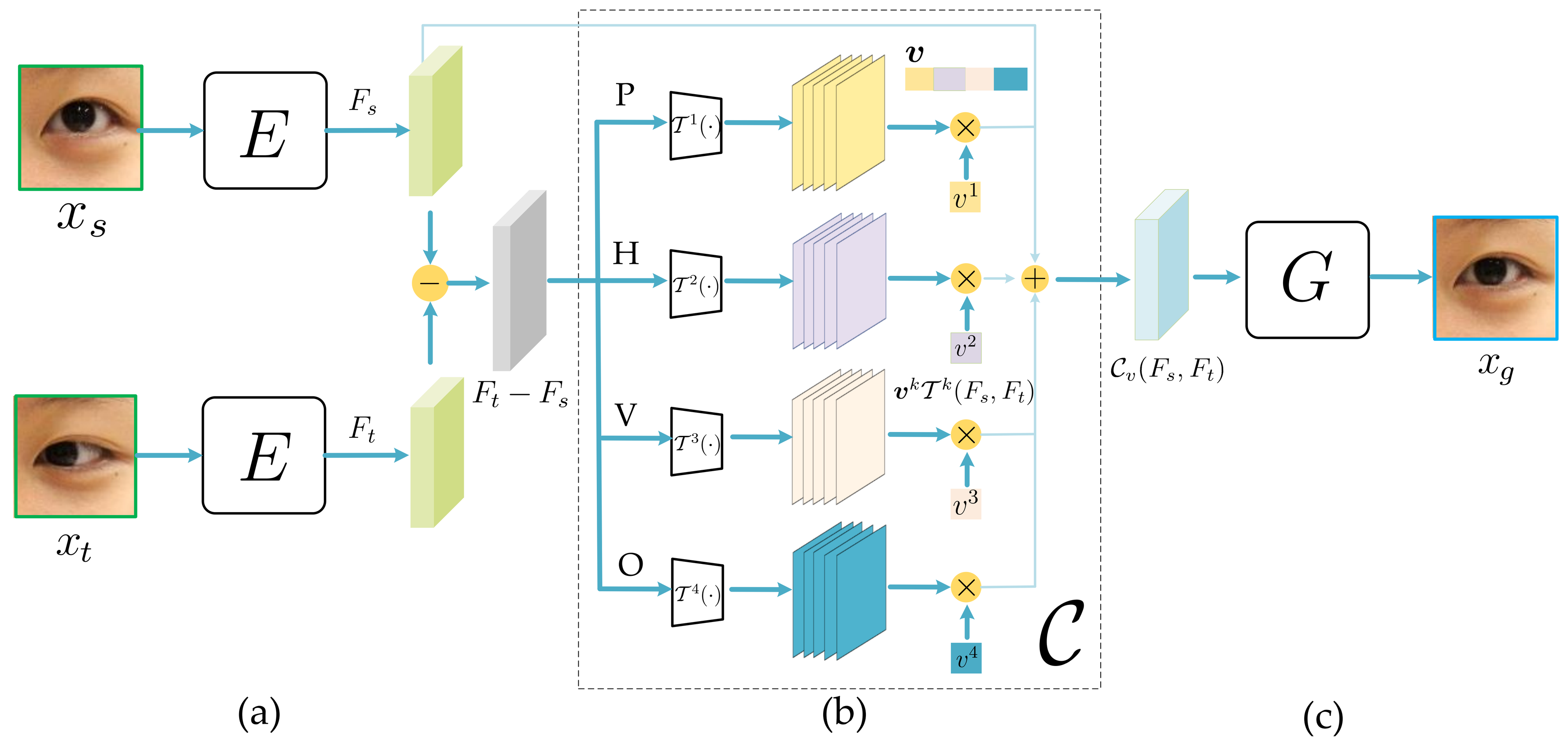
Problem Formulation

Image Formation Model: Our goal is to learn a model which can achieve both **precise redirection** and **continuous interpolation**. For an RGB image of an eye patch $x \in \mathbb{R}^{H\times W\times 3}$, we define three primary attributes, *i.e.*, gaze direction $\boldsymbol{d}_g$ and head pose $\boldsymbol{d}_h$, where $\boldsymbol{d}_g = [\phi_g, \theta_g]$, $\phi_g \in \mathbb{R}$ and $\theta_g \in \mathbb{R}$ denote the target yaw and pitch angles, respectively. Given two gaze images x_s and x_t with different attributes, the task is to redirect the eye gaze of one person x_s into any gaze direction x_g depicted in the reference image of another person x_t, or to generate continuous intermediate results x_g between x_s and x_t of the same person.

Main Idea: we design a model including an encoder $\boldsymbol{E}$, a controller $\boldsymbol{\mathcal{C}}$ and a decoder $\boldsymbol{G}$. The encoder $\boldsymbol{E}$ maps images x_s and x_t into feature space $F_s = \boldsymbol{E}(x_s)$ and $F_t = \boldsymbol{E}(x_t)$. The controller $\boldsymbol{\mathcal{C}}$ produces morphing results of two samples $\mathcal{C}(F_s, F_t)$. The decoder $\boldsymbol{G}$ maps the desired features back to the image space.

Method

Network Architecture:



Loss Function for Encoder and Decoder:

$$\mathcal{L}_{\boldsymbol{G}} = \lambda_p \mathcal{L}_p + \lambda_{\boldsymbol{G}} \mathcal{L}_{recon},$$

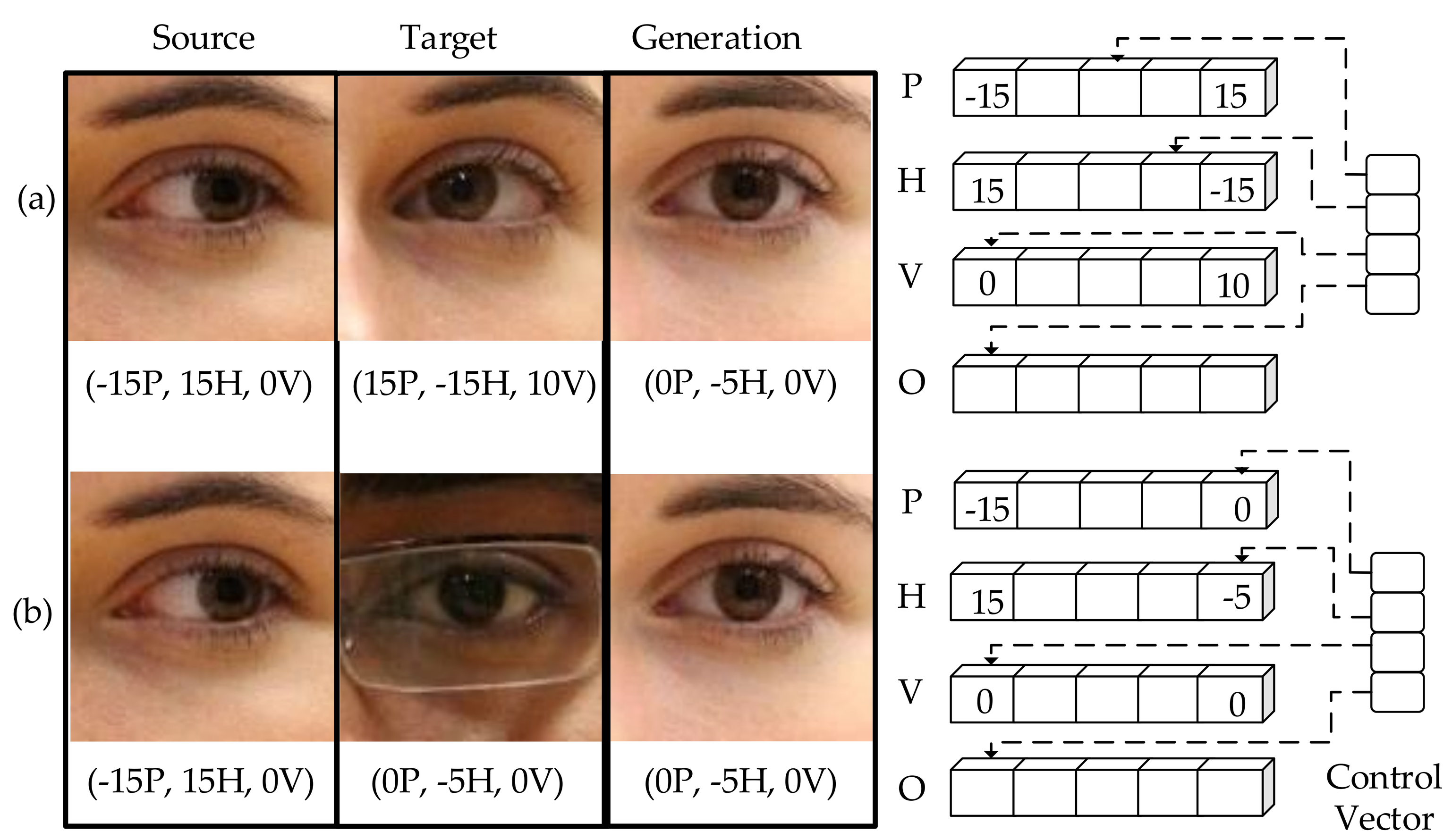
$$\mathcal{L}_{\boldsymbol{\mathcal{D}}} = \lambda_{GAN_{\boldsymbol{\mathcal{D}}}} \mathcal{L}_{GAN_{\boldsymbol{\mathcal{D}}}},$$

$$\mathcal{L}_{\boldsymbol{E}} = \lambda_{GAN_{\boldsymbol{E}}} \mathcal{L}_{GAN_{\boldsymbol{E},\boldsymbol{\mathcal{C}}}} + \lambda_{\boldsymbol{E}} \mathcal{L}_{recon} + \lambda_{distill} \mathcal{L}_{distill},$$

Loss Function for Controller:

$$\mathcal{L}_{\boldsymbol{\mathcal{C}}} = \lambda_{GAN_{\boldsymbol{\mathcal{C}}}} \mathcal{L}_{GAN_{\boldsymbol{E},\boldsymbol{\mathcal{C}}}} + \lambda_{\boldsymbol{\mathcal{C}}_{isp}} \mathcal{L}_{\boldsymbol{\mathcal{C}}_{isp}} + \lambda_{\boldsymbol{\mathcal{C}}_t} \mathcal{L}_{\boldsymbol{\mathcal{C}}_t}$$

Control Mechanism



- For gaze redirection, the inputs are a gaze patch of a certain person as the source image and that of another person with the desired directions as the target. If we want to redirect the same source image to $(0\text{P}, 0\text{V}, -5\text{H})$ defined in the reference image, we can set $\boldsymbol{v}_2$ as $[1.0, 1.0, 0, 0]$, which alters the target attribute while keeping others almost intact.
- For gaze interpolation, the inputs are two gaze patches of the same person. Given $(-15\text{P}, 0\text{V}, 15\text{H})$ as the source and $(15\text{P}, 0\text{V}, -15\text{H})$ as the target, we can set the control vector $\boldsymbol{v}_1$ as $[0.5, 0.667, 0, 0]$ to generate a specific intermediate result $(0\text{P}, 0\text{V}, -5\text{H})$. Similarly, we can generate predictable interpolation sequences in different **orders** using different control vectors.
- The appearance of the reference image is allowed to be dramatically different from the source.

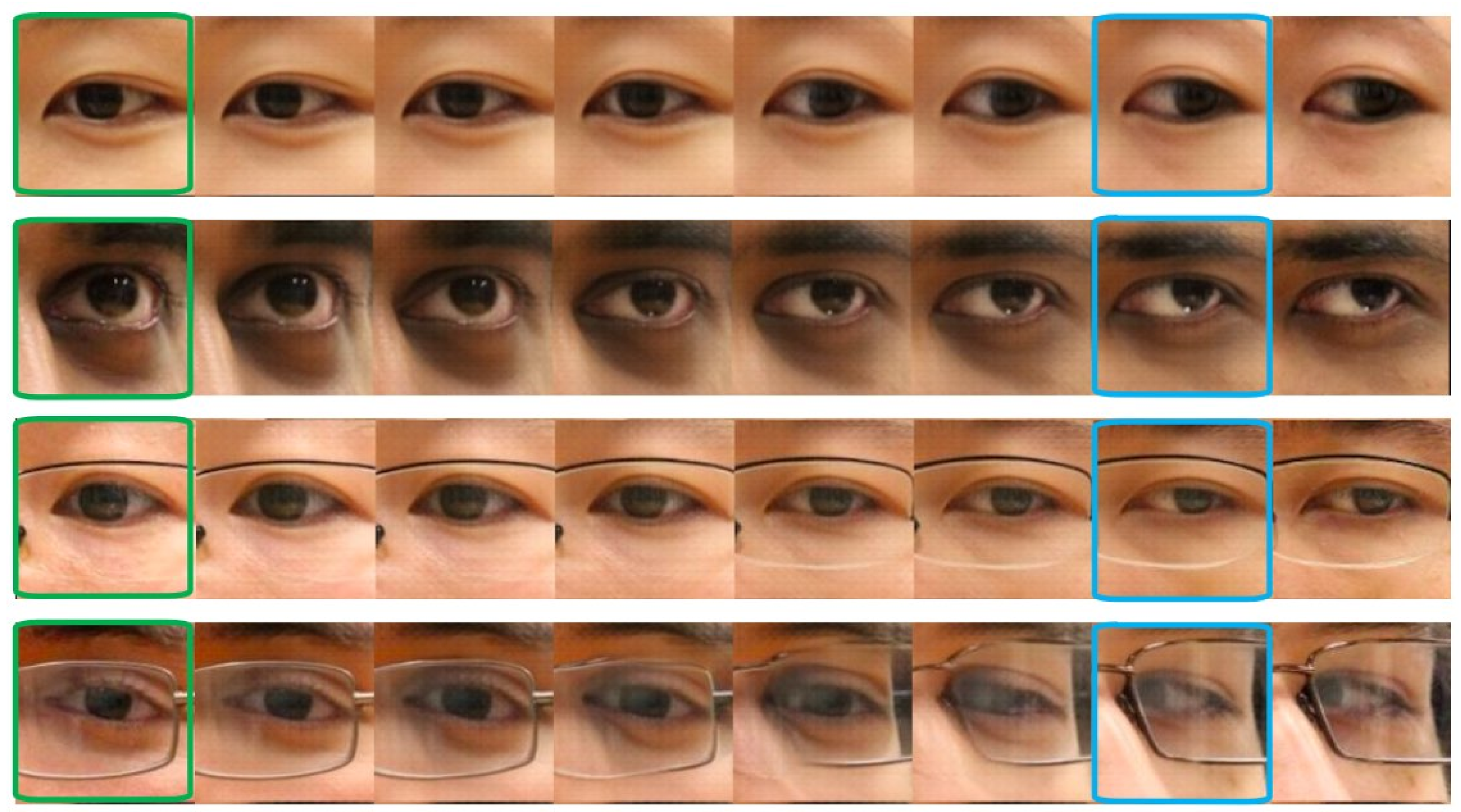
Experiments & Results

Data Collection and Comparison:

- To facilitate covering the full space of gaze directions, we introduce a high-quality gaze image dataset with a large range of directions, which also benefits researchers in related areas. We collect images in the same way as Columbia Gaze dataset, which is a high-resolution, publicly available human gaze dataset collected from 56 subjects in a constrained environment.

Dataset	Real	High Res	Constrained	Annotation Type	Num. Image	Head Pose	Gaze Range
CMU Multi-Pie	✓	✗	✓	Facial landmarks	755,370	✓	Small
Gaze Capture	✓	✗	✗	2D position on screen	>2.5M	✗	Small
SynthesEyes	✗	✓	✗	Gaze vector	11,382	✓	Full
UnityEyes	✗	✓	✗	Gaze vector	1,000,000	✓	Full
MPII Gaze	✓	✗	✗	Gaze vector	213,659	✓	Small
UT Multi-View	✓	✗	✓	Gaze vector	1,152,000	✓	Large
Columbia	✓	✓	✓	Gaze vector	5,880	✓	Medium
Ours	✓	✓	✓	Gaze vector	29,250	✓	Large

Results of Gaze Interpolation and Extrapolation



Results of Gaze Redirection