

# RectifID: Personalizing Rectified Flow with Anchored Classifier Guidance

Zhicheng Sun<sup>1</sup>, Zhenhao Yang<sup>3</sup>, Yang Jin<sup>1</sup>, Haozhe Chi<sup>1</sup>, Kun Xu<sup>2</sup>, Kun Xu<sup>2</sup>,  
Liwei Chen<sup>2</sup>, Hao Jiang<sup>1</sup>, Yang Song, Kun Gai<sup>2</sup>, Yadong Mu<sup>1</sup>

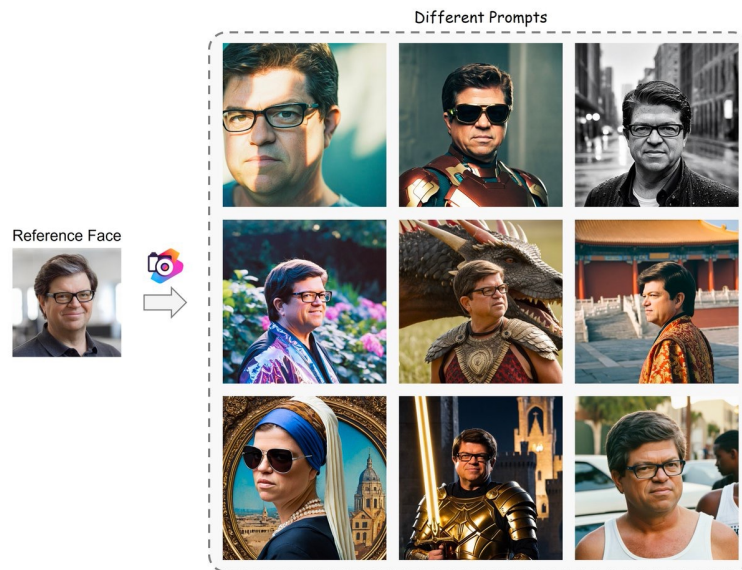
<sup>1</sup>Peking University, <sup>2</sup>Kuaishou Technology,

<sup>3</sup>University of Electronic Science and Technology of China

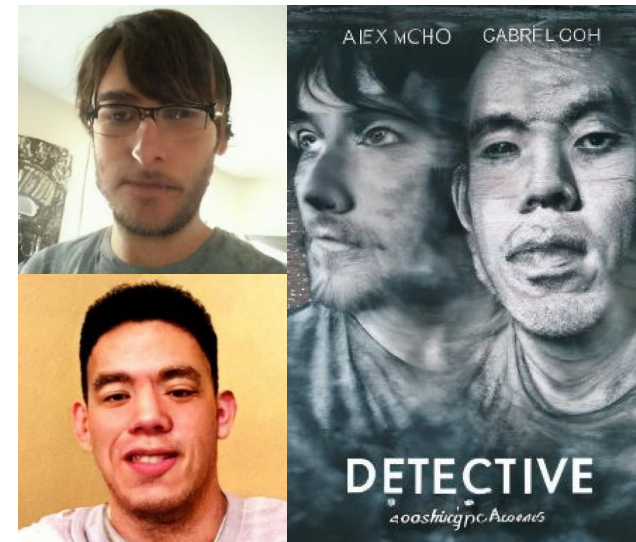
# Introduction

## Problem Description

- Generate identity-preserving images from user-provided reference image.
- Challenges: 1) controllability: the denoising objective is too weak to control subject identity. 2) flexibility: existing models except GPT-4o cannot handle more general scenarios.



PhotoMaker



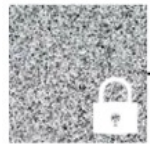
GPT-4o

# Introduction

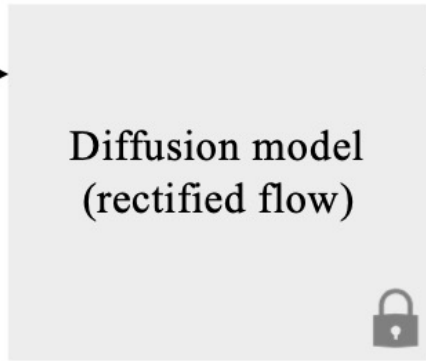
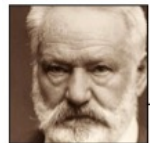
## Motivation

- We advocate a training-free approach that utilizes the guidance of a pre-trained classifier without extra training of the generative model.
- It can be flexibly applied to various tasks by plug-and-play combination with classifiers.

Training-free



ref img

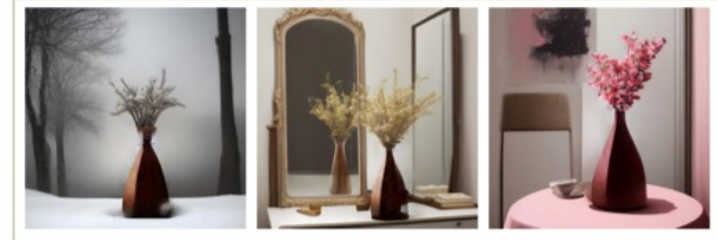


Off-the-shelf classifier

face discriminator



object discriminator



# Method

## Preliminary: Classifier Guidance

- Consider an ODE-based diffusion model that learns a velocity field from noise  $\mathbf{z}_0$  to data  $\mathbf{z}_1$ :

$$d\mathbf{z}_t = \mathbf{v}(\mathbf{z}_t, t)dt.$$

- It can be controlled at test-time by adding a classifier gradient term:

$$\hat{\mathbf{v}}(\mathbf{z}_t, t) = \mathbf{v}(\mathbf{z}_t, t) + s \cdot \nabla_{\mathbf{z}_t} \log p(c|\mathbf{z}_t).$$

## Limitations

- It relies on a special classifier trained on noisy images  $\mathbf{z}_t$ , and therefore can't reuse existing classifiers trained on clean images  $\mathbf{z}_1$ .
- Existing approximation techniques require a high number of iterations (100+).

# Method

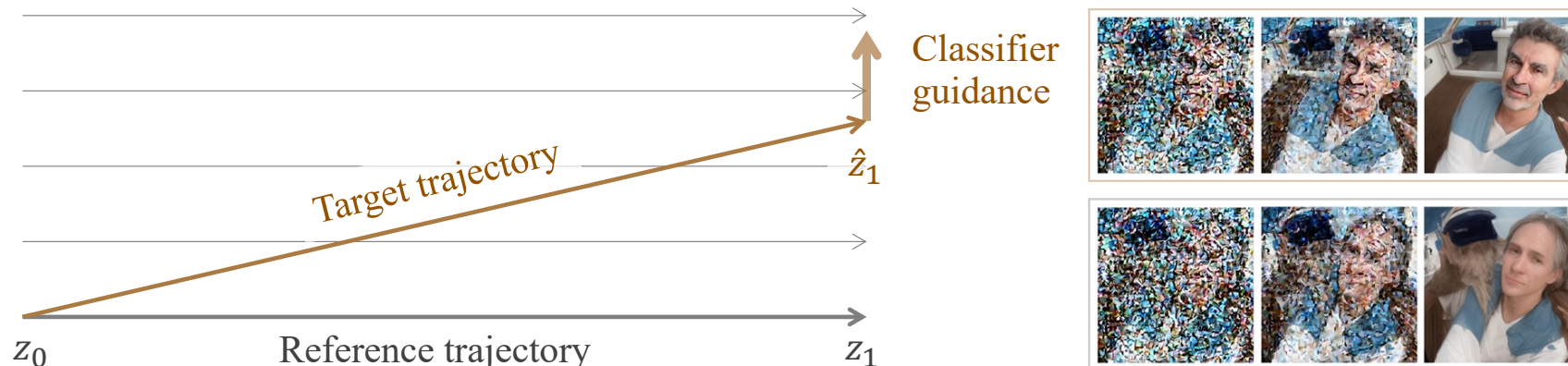
## Classifier Guidance for Rectified Flow

- Using rectified flow (which is trained to be straight), training-free guidance can be cast to a fixed-point problem:

$$\begin{aligned} \mathbf{z}_1 &= \mathbf{z}_0 + \hat{\mathbf{v}}(\mathbf{z}_1, 1) \\ &= \mathbf{z}_0 + \mathbf{v}(\mathbf{z}_1, 1) + s \cdot \nabla_{\mathbf{z}_1} \log p(c|\mathbf{z}_1). \end{aligned}$$

- We further introduce a reference trajectory to improve the stability of its solving process:

$$\hat{\mathbf{z}}_1 = \mathbf{z}_1 + s \cdot [\nabla_{\mathbf{z}_0} \mathbf{z}_1] \nabla_{\hat{\mathbf{z}}_1} \log p(c|\hat{\mathbf{z}}_1).$$



# Method

## Implementation

- The derived method is implemented for a practical class of rectified flow assumed to be piecewise straight.

---

**Algorithm 1** Anchored Classifier Guidance

---

**Input:** rectified flow  $v$ , classifier  $p(c|\cdot)$ , sampling steps  $K$ , iterations  $N$ .

Initialize reference trajectory  $\mathbf{z}_{t_k}^{[0]}$  from  $v$ . ▷ Eq. (13)

Initialize target trajectory  $\hat{\mathbf{z}}_{t_k}^{[0]} \leftarrow \mathbf{z}_{t_k}^{[0]}$ .

**for**  $i \leftarrow 0$  to  $N - 1$  **do**

    Update reference trajectory with predicted starting points  $\mathbf{z}_{t_k}^{[i+1]}$ . ▷ Eq. (15)

    Update target trajectory  $\hat{\mathbf{z}}_{t_k}^{[i+1]}$  with classifier output  $p(c|\hat{\mathbf{z}}_1^{[i]})$ . ▷ Eq. (16)

**Output:** target trajectory  $\hat{\mathbf{z}}_{t_k}^{[N]}$  subject to condition  $c$ .

---

- Combined with face (ArcFace) or object discriminators (DINOv2), it supports personalization for human faces and more general objects,

$$p(c|\hat{\mathbf{z}}_1^{[i]}) = \text{sim} \left( f \circ g(\hat{\mathbf{z}}_1^{[i]}), f \circ g(\mathbf{z}_{\text{ref}}) \right).$$

# Personalized Generation Results

RectifID achieves better results than training-based approaches in a reasonable inference time.

Method	Base model	Training	Identity $\uparrow$	Prompt $\uparrow$	Time $\downarrow$
Textual Inversion (Gal et al., 2023)	SD 2.1	-	0.2115	0.2498	6331
DreamBooth (Ruiz et al., 2023a)	SD 2.1	-	0.2053	0.3015	623
NeTI (Alaluf et al., 2023)	SD 1.4	-	0.3789	0.2325	1527
Celeb Basis (Yuan et al., 2023)	SD 1.4	-	0.2070	0.2683	140
Cross Initialization (Pang et al., 2024)	SD 2.1	-	0.2517	0.2859	346
IP-Adapter (Ye et al., 2023)	SD 1.5	10M	0.4778	0.2627	<b>2</b>
PhotoMaker (Li et al., 2024)	SDXL	112K	0.2271	<u>0.3079</u>	<u>4</u>
InstantID (Wang et al., 2024)	SDXL	60M	<u>0.5806</u>	0.3071	6
RectifID (20 iterations)	SD 1.5	-	0.4860	0.2995	9
RectifID (100 iterations)	SD 1.5	-	<b>0.5930</b>	0.2933	46
RectifID (20 iterations)	SD 2.1	-	0.5034	<b>0.3151</b>	20

# Personalized Generation Results



Input



cave mural



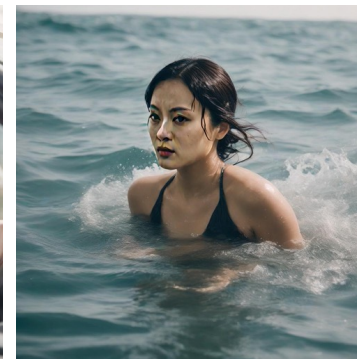
piloting a  
fighter jet



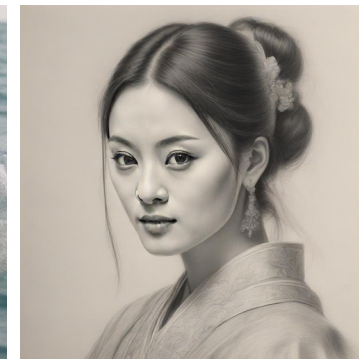
wear a magician  
hat and a blue  
coat in a garden



driving a car



swimming



pencil  
drawing



Input



as an amazon  
warrior



wear a magician  
hat and a blue  
coat in a garden



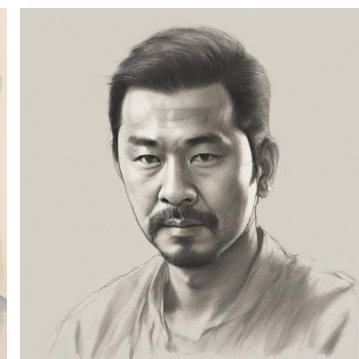
in a cowboy  
hat



wearing a hat



watercolor  
painting



pencil  
drawing



# Personalized Generation Results



Input



graduating after finishing PhD



wear a magician hat and a blue coat in a garden



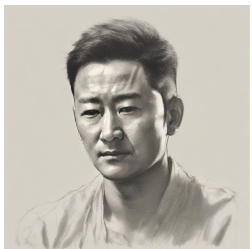
as White Queen



wearing headphones



concert poster



pencil drawing



Input



in a police outfit



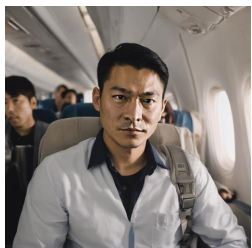
holding roses in front of the Eiffel Tower



wear a magician hat and a blue coat in a garden



reading on the train



buckled in his seat on a plane



sipping coffee at a cafe



Input



stained glass window



painting in Van Gogh style



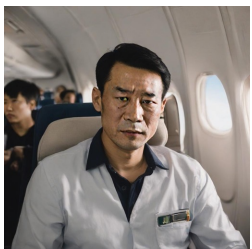
watercolor painting



driving a car



colorful mural on a street wall



buckled in his seat on a plane

# More Customization Results

 <p>Input</p>	 <p>on pink fabric</p>	 <p>in a firefighter outfit</p>	 <p>in a purple wizard outfit</p>	 <p>wearing a rainbow scarf</p>	 <p>in the jungle</p>	 <p>in the snow</p>
 <p>Input</p>	 <p>wearing a santa hat</p>	 <p>in a firefighter outfit</p>	 <p>in a purple wizard outfit</p>	 <p>in a police outfit</p>	 <p>shiny</p>	 <p>in front of a mountain</p>
 <p>Input</p>	 <p>wearing a santa hat</p>	 <p>in a firefighter outfit</p>	 <p>in a purple wizard outfit</p>	 <p>in a police outfit</p>	 <p>wearing a yellow shirt</p>	 <p>in the snow</p>
 <p>Input</p>	 <p>wearing a santa hat</p>	 <p>in a firefighter outfit</p>	 <p>in a purple wizard outfit</p>	 <p>in a police outfit</p>	 <p>wearing a rainbow scarf</p>	 <p>wet</p>

# Generalization



(a) Segmentation map



(b) Style transfer

# Thanks for listening

Code is available at <https://github.com/feifeiobama/RectifID>