



**ICLR**



**武汉大学**  
WUHAN UNIVERSITY



**PRINCETON  
UNIVERSITY**

# Diffusion in Diffusion: Cyclic One-Way Diffusion for Text-Vision Conditioned Generation

Ruoyu Wang<sup>1,\*</sup>, Yongqi Yang<sup>1,\*</sup>, Zhihao Qian<sup>1</sup>, Ye Zhu<sup>2</sup>, Yu Wu<sup>1,†</sup>

<sup>1</sup> School of Computer Science, Wuhan University

<sup>2</sup> Department of Computer Science, Princeton University

\* Equal Contribution

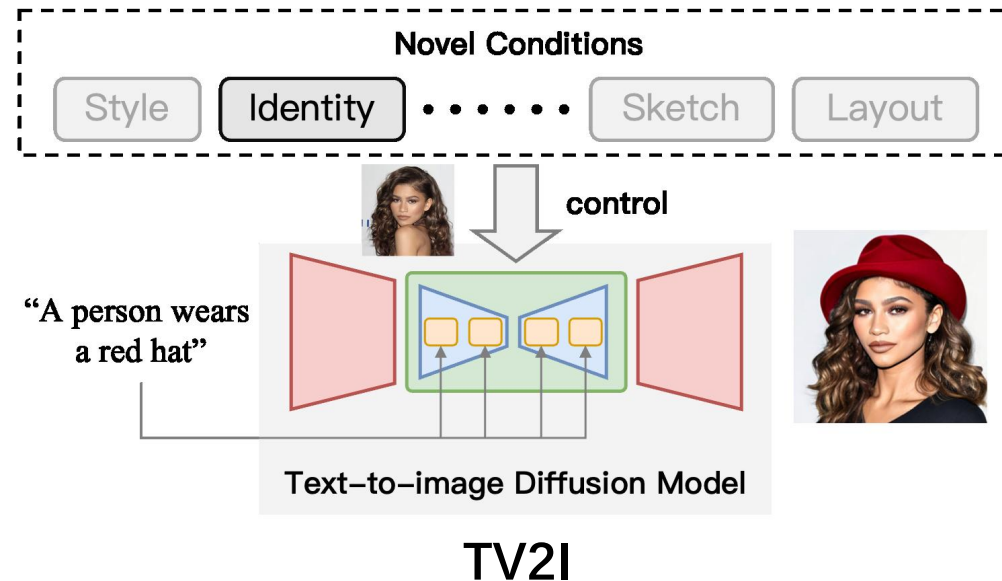
<https://wangruoyu02.github.io/cow.github.io/>

# Text-Vision to Image (TV2I)

“a photograph of an astronaut riding a horse”

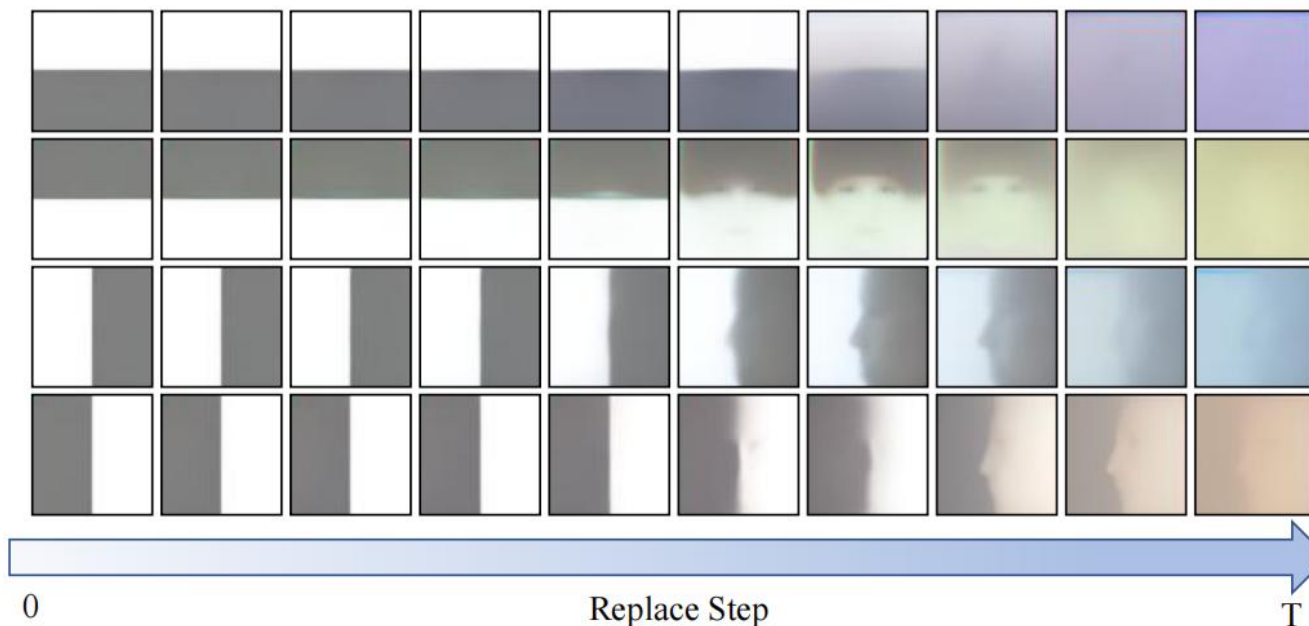
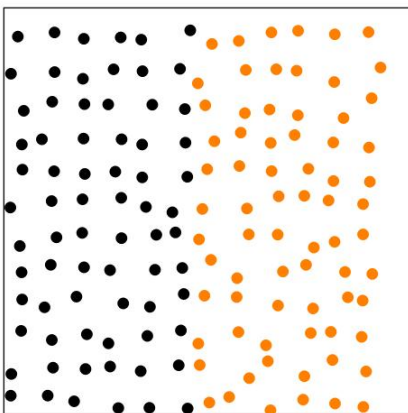


T2I

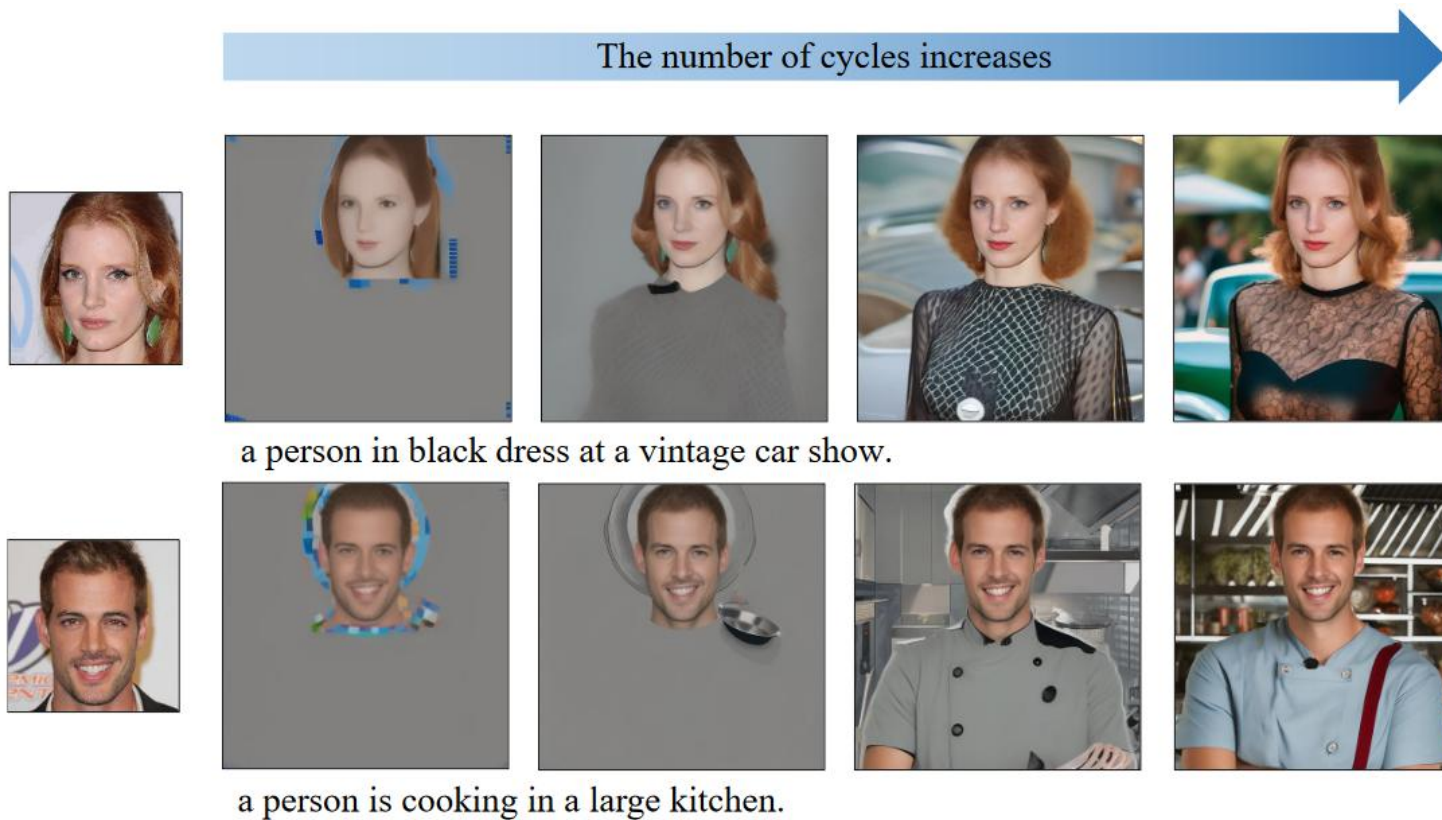


# Diffusion in Diffusion

- The diffusion models inherit the characteristics of stochastic random walk in the data space along the denoising trajectory

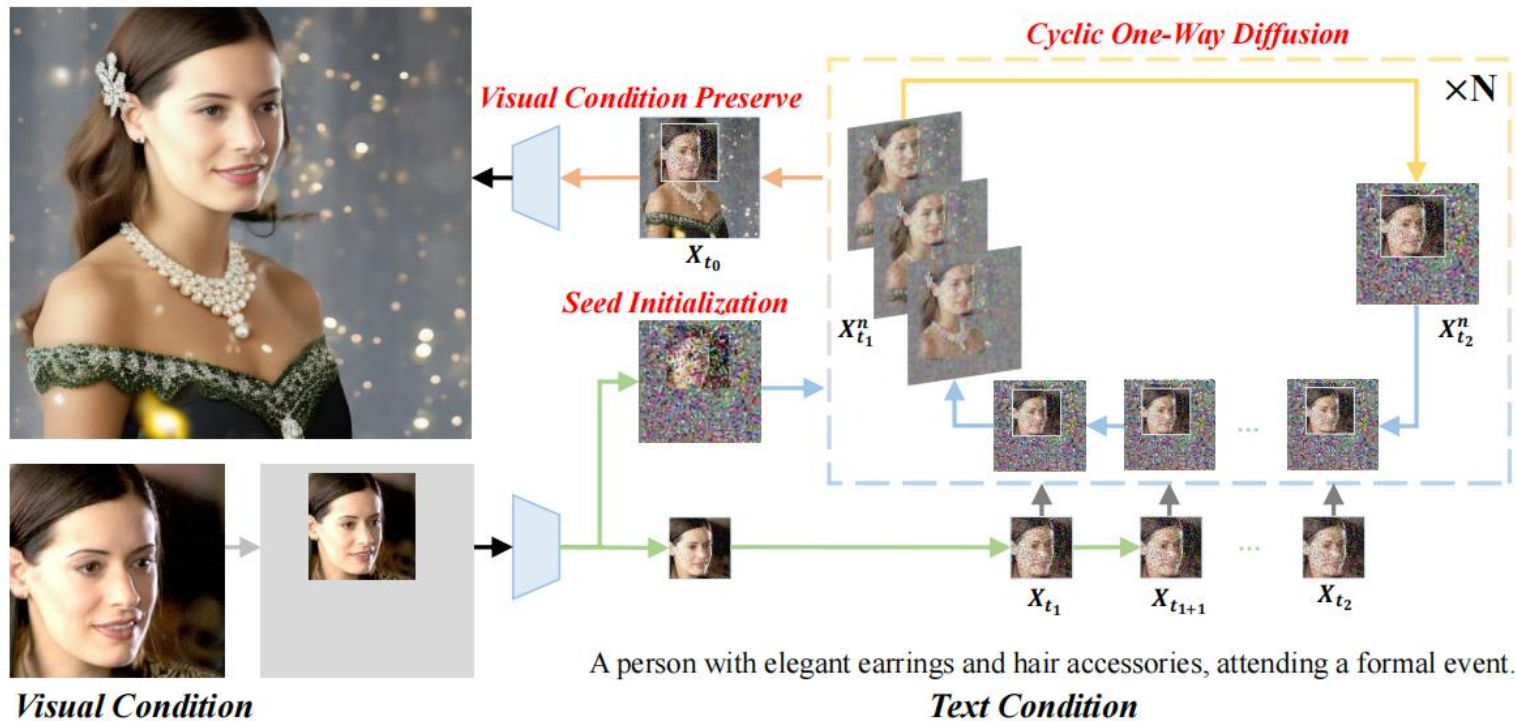


# Motivation



# Cyclic One-Way Diffusion

- Re-inject the semantics in a cyclic way to maximize the information flow



→ Inject noise   
 → ODE inverse   
 → DDIM(eta=0)   
 → DDIM(eta=1)   
 → Replace

# Tradeoff between T-V Conditions

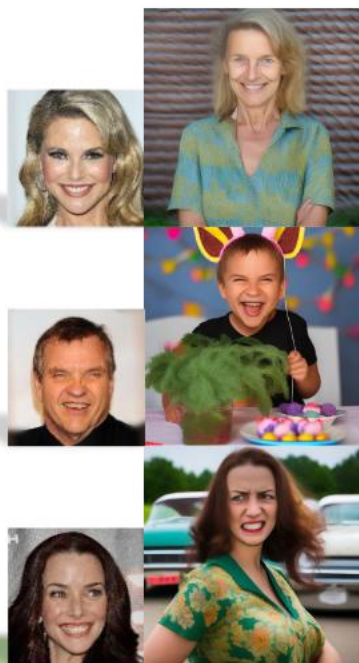
*Almost unchanged*



*Add accessories*



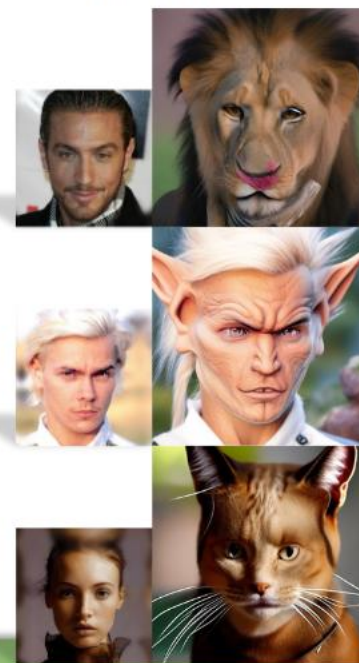
*Attribute edit*



*Style transfer*



*Cross-domain transformation*



# Tradeoff between T-V Conditions



# Comparison with SOTA

## Inpainting



"in a black suit"



## Attribute Editing



"an old person"



## Style Transfer



"Impressionist style"



Stable Diffusion  
(Inpainting)

ControlNet  
(Generation)

TI  
(Generation)

DreamBooth  
(Generation)

Ours  
(Generation)



# Comparison with SOTA

Table 1: Quantitative and qualitative comparison between COW and SOTA methods.

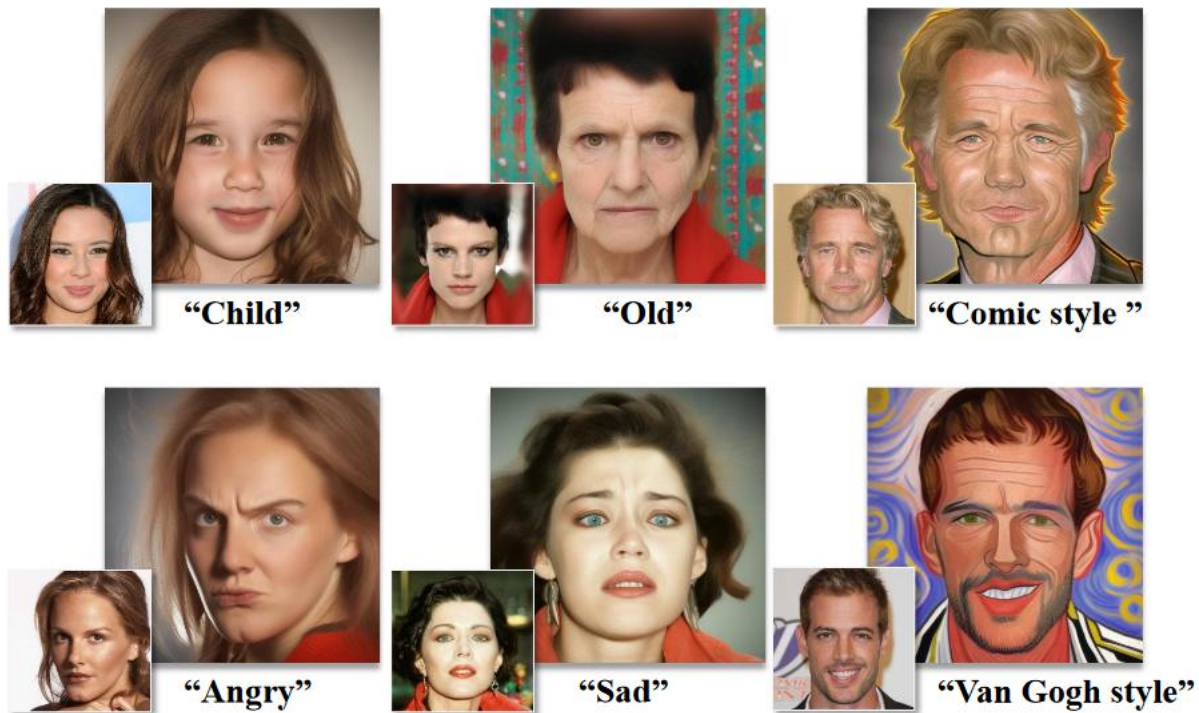
| Methodology   | Clip-T $\uparrow$ | ID-Distance $\downarrow$ | Face Detection Rate $\uparrow$ | Time $\downarrow$ | Condition Consistency $\uparrow$ | General Fidelity $\uparrow$ |
|---------------|-------------------|--------------------------|--------------------------------|-------------------|----------------------------------|-----------------------------|
| TI            | 0.253             | 1.186                    | 70.66%                         | 3025s             | 02.73%                           | 12.60%                      |
| DreamBooth    | 0.329             | 1.361                    | 70.50%                         | 732s              | 09.60%                           | 28.73%                      |
| ControlNet    | 0.305             | 1.194                    | 45.66%                         | 4s                | 11.60%                           | 04.73%                      |
| SD inpainting | 0.300             | 0.408                    | 100.00%                        | 5s                | 06.33%                           | 02.07%                      |
| COW (ours)    | 0.306             | 0.901                    | 100.00%                        | 6s                | <b>69.73%</b>                    | <b>51.87%</b>               |

| Methodology   | Condition Consistency |                |                   | General Fidelity |                |                   |
|---------------|-----------------------|----------------|-------------------|------------------|----------------|-------------------|
|               | inpainting            | style transfer | attribute editing | inpainting       | style transfer | attribute editing |
| TI            | 2.00%                 | 3.60%          | 2.60%             | 10.60%           | 15.40%         | 11.80%            |
| DB            | 2.80%                 | 13.40%         | 12.80%            | 21.40%           | 26.80%         | 37.80%            |
| ControlNet    | 5.40%                 | 24.20%         | 5.40%             | 3.00%            | 7.20%          | 4.00%             |
| SD inpainting | 5.20%                 | 10.60%         | 3.20%             | 3.00%            | 7.20%          | 4.00%             |
| COW (Ours)    | <b>84.60%</b>         | <b>48.20%</b>  | <b>76.00%</b>     | <b>63.00%</b>    | <b>47.60%</b>  | <b>45.20%</b>     |

# More Applications



A cat next to ball of yarn on couch   A cat in front of the Eiffel Tower   A cat in a cute hat on a roof



“Child”

“Old”

“Comic style”

“Angry”

“Sad”

“Van Gogh style”

# Check out our paper!



<https://arxiv.org/abs/2306.08247>



<https://github.com/wangruoyu02/COW>



[wangruoyu@whu.edu.cn](mailto:wangruoyu@whu.edu.cn)  
[yongqiyang@whu.edu.cn](mailto:yongqiyang@whu.edu.cn)