

RefDrop:

Controllable Consistency in Image or Video Generation via Reference Feature Guidance

Jiaojiao Fan, Haotian Xue, Qinsheng Zhang, Yongxin Chen

Georgia Tech, NVIDIA

Multi-Subject Consistency



“A happy girl with brown eyes is ...”

coding in front of a computer with her dog bungee jumping with her dog riding bicycles in the park with her dog cooking a meal in the kitchen with her dog reading a book under a tree with her dog



Ours

SDXL

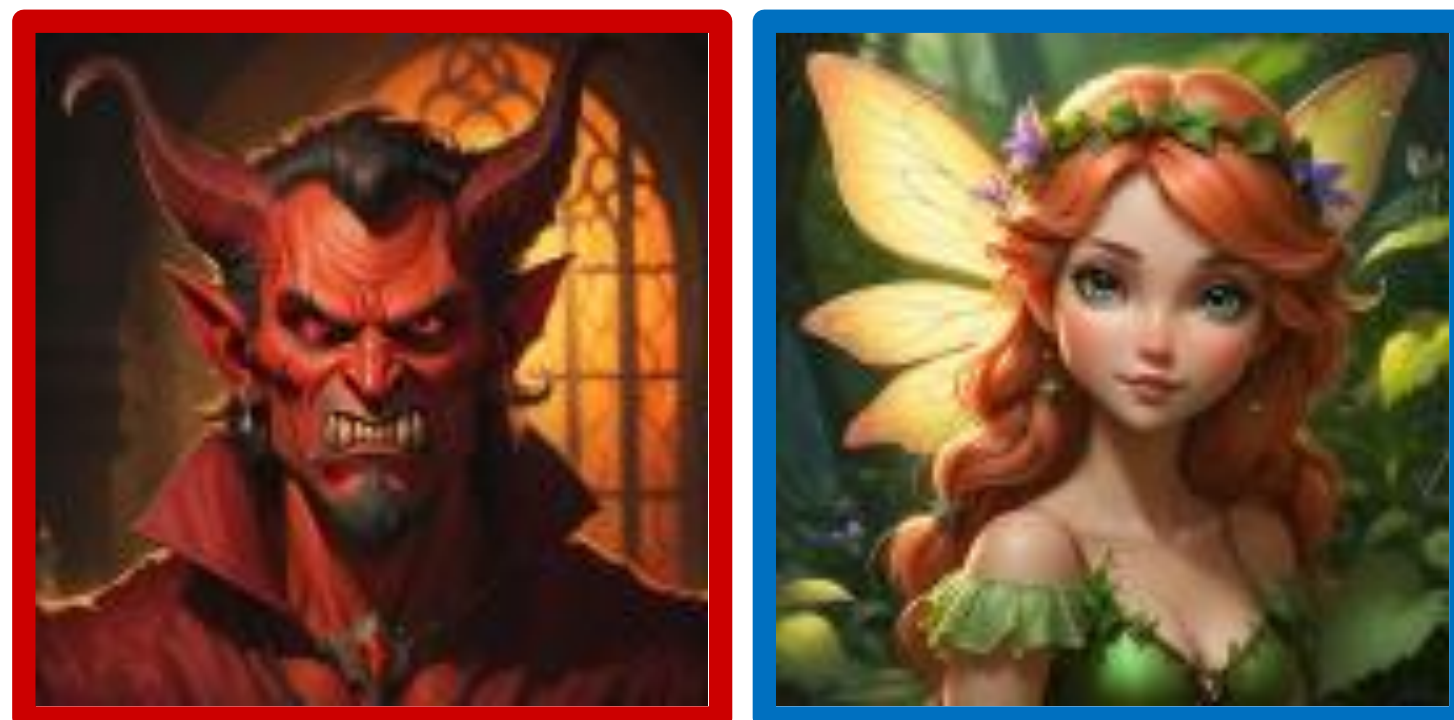
RefDrop:

Controllable Consistency in Image or Video Generation via Reference Feature Guidance

Jiaojiao Fan, Haotian Xue, Qinsheng Zhang, Yongxin Chen

Georgia Tech, NVIDIA

Blend feature



“A half **devil** and half **fairy** woman is ...”



RefDrop:

Controllable Consistency in Image or Video Generation via Reference Feature Guidance

Jiaojiao Fan, Haotian Xue, Qinsheng Zhang, Yongxin Chen

Georgia Tech, NVIDIA

**Personalized
Consistent Video**



Subject Image



Stable Video Diffusion

Ours

Ours

RefDrop

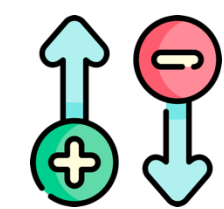
Controllable Consistency in Image or Video Generation via Reference Feature Guidance

RefDrop is a **training-free, plug-and-play** method that provides flexible control over the consistency in image and video generation

Major benefits



Training-free



Allow **negative** consistency



Allow **single** reference image



Handle **both image and video** consistency

RefDrop

Controllable Consistency in Image or Video Generation via Reference Feature Guidance

Reference strength =

-0.4

-0.2

0

0.2

0.4

Reference image



A cyberpunk style of a young woman with blonde hair riding a neon-lit hoverbike in a park



A cylindrical red night cream bottle with silver cap nestled in a forest clearing



Negative similarity

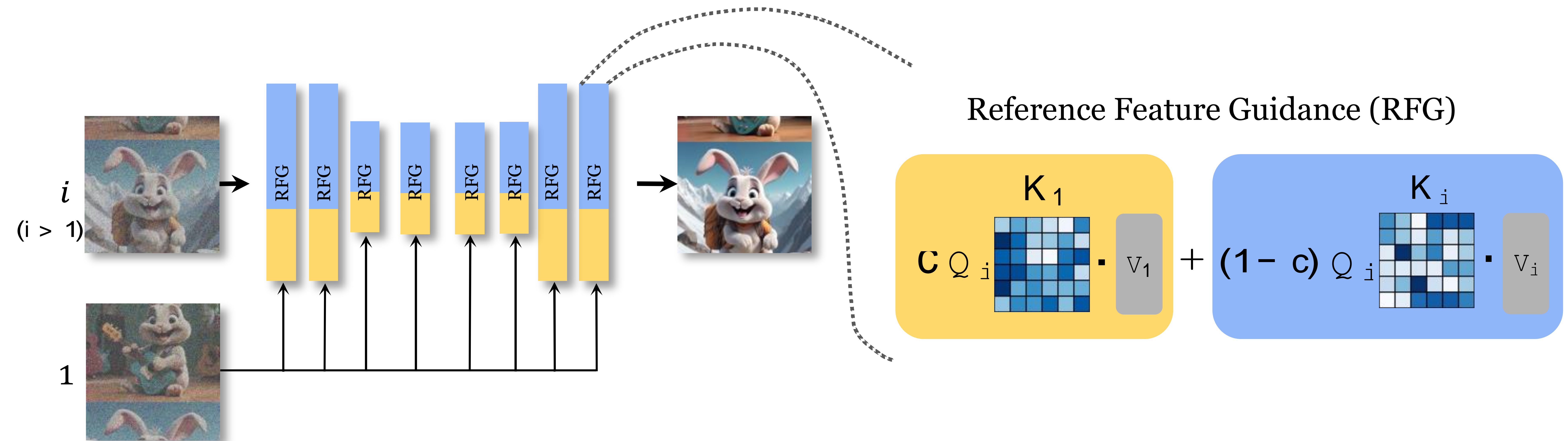
SDXL

Positive similarity

RefDrop

Controllable Consistency in Image or Video Generation via Reference Feature Guidance

Method



RefDrop

Controllable Consistency in Image or Video Generation via Reference Feature Guidance

Evaluation

