

# Story generation

## ONE-PROMPT-ONE-STORY: FREE-LUNCH CONSISTENT TEXT-TO-IMAGE GENERATION USING A SINGLE PROMPT



<https://github.com/byliutao/storydiffusion>

# Existing Methods

“A hyper-realistic digital painting of a fairy<sup>0</sup>, dressed in a cloak of spider silk<sup>1</sup>, wearing a garland of fireflies<sup>2</sup>.”



SDXL

Juggernaut-X-v10

Texture-Inversion

Ip-Adapter

ConsiStory

NPR

**Ours**

Base model

Training method

Training-free method

[1] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt

# Setup

## ■ Multi Prompt

- 1 “A photo of a fox<sup>0</sup>, wearing a scarf in a meadow<sup>1</sup>.”
- 2 “A photo of a fox<sup>0</sup>, playing in the snow<sup>2</sup>.”
- 3 “A photo of a fox<sup>0</sup>, at the edge of a river<sup>3</sup>.”

## ■ Single Prompt

- 1 “A photo of a fox<sup>0</sup>, wearing a scarf in a meadow<sup>1</sup>,  
playing in the snow<sup>2</sup>, at the edge of a river<sup>3</sup>.”

# Motivation

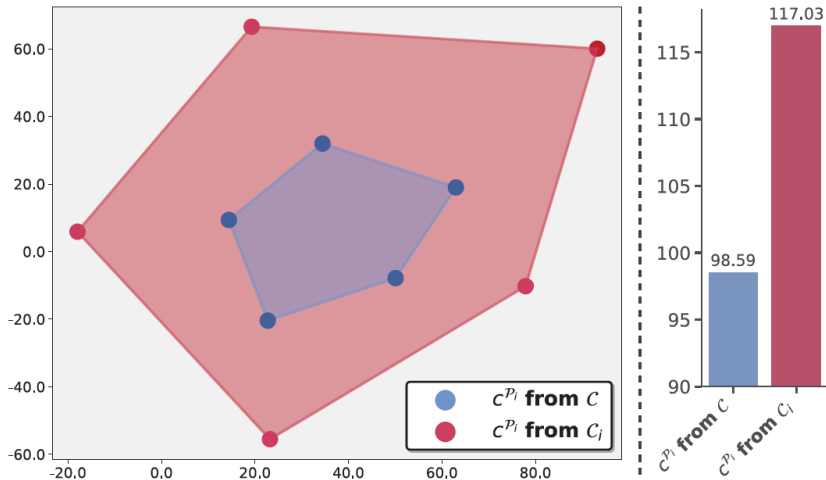
- Consolidates all desired prompts into a single longer sentence
- Reweighting the consolidated prompt embeddings for each frame

“A watercolor illustration of a cute kitten<sup>0</sup>, in a garden<sup>1</sup>, dressed in a super hero cape<sup>2</sup>, wearing a collar with a bell<sup>3</sup>, sitting in a basket<sup>4</sup>, dressed in a cute sweater<sup>5</sup>.”



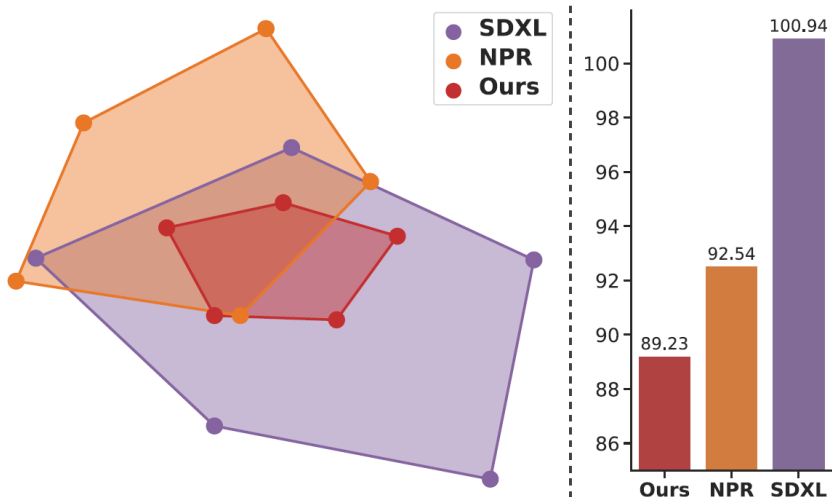
[1] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt

# Context Consistency In Text Embedding





[1] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt

# Context Consistency In Image Generation



[1] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt

# Improve Naive Prompt Reweighting (NPR)

- Lack of prompt-image alignment, have similar background  SVR
- Identity are inconsistent among all frames  IPCA

NPR



# Overall Pipeline

**Identity Prompt:** “A robust painting of a sturdy dwarf<sup>0</sup>.”

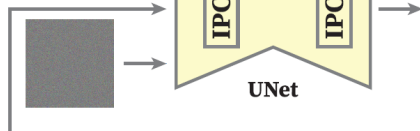
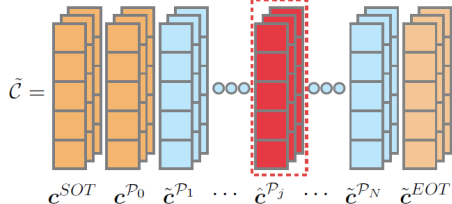
**Frame Prompt:** “runes into stone tablets<sup>1</sup>, ....., **gath-  
-ering herbs in a forest<sup>j</sup>**, ....., mining for gems in a  
glittering cave<sup>N</sup>.”

■ express ■ suppress

**Text Encoder**

$$\mathcal{C} = [\mathbf{c}^{SOT}, \mathbf{c}^{\mathcal{P}_0}, \mathbf{c}^{\mathcal{P}_1}, \dots, \mathbf{c}^{\mathcal{P}_j}, \dots, \mathbf{c}^{\mathcal{P}_N}, \mathbf{c}^{EOT}]$$

**SVR**

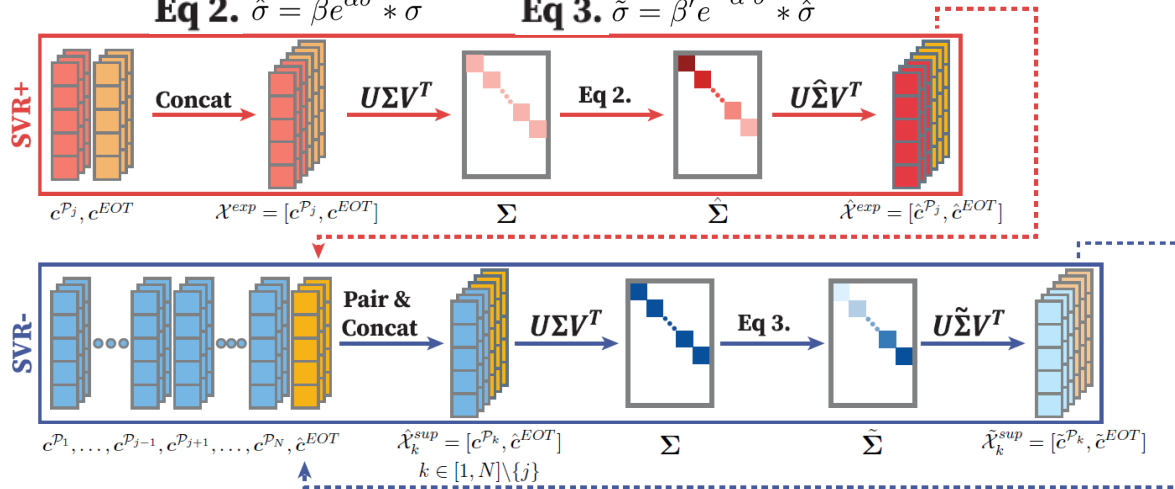




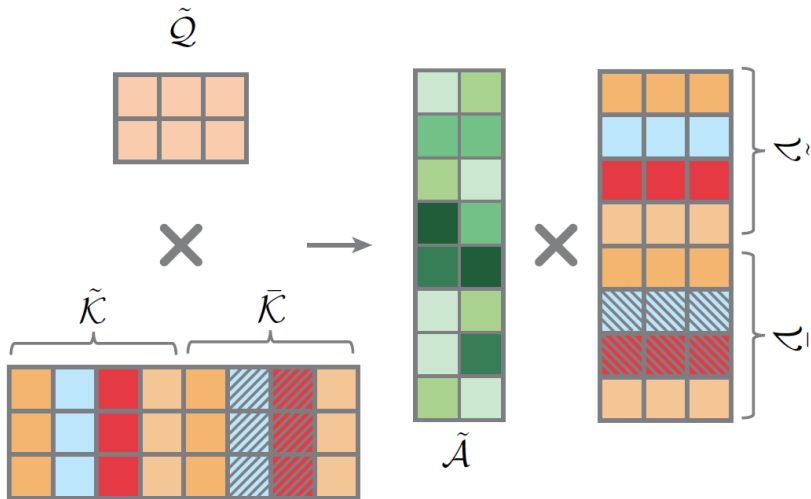
# Singular-Value Reweighting

**Eq 2.**  $\hat{\sigma} = \beta e^{\alpha\sigma} * \sigma$

**Eq 3.**  $\tilde{\sigma} = \beta' e^{-\alpha'\hat{\sigma}} * \hat{\sigma}$



# Identity-Preserving Cross-Attention (IPCA)

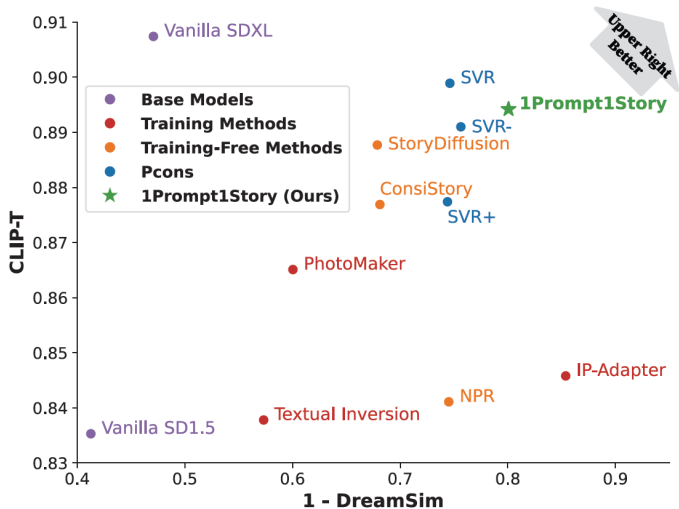


# Qualitative Comparison



[1] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt

# Quantitative Comparison



[1] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt

# Ablation Study

“A photo of a dog<sup>0</sup>, chasing a frisbee in a colorful park<sup>1</sup>, dancing to music at a vibrant street festival<sup>2</sup>, jumping through a hoop at a circus performance<sup>3</sup>, posing for a photoshoot in a modern art gallery<sup>4</sup>.”



[1] One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt

# Additional Applications



"A photo of an elderly gentleman<sup>0</sup>, cross-legged on a mountain, surrounded by ancient scrolls<sup>1</sup>, sitting on a bench under cherry blossoms<sup>2</sup>, sitting in a sunny garden, playing with a puppy<sup>3</sup>."



"a photo of a woman, enjoying a cup of coffee at a cozy outdoor café<sup>1</sup>, painting a beautiful landscape on a canvas<sup>2</sup>, talking on her phone<sup>3</sup>."

# Thanks!