SVGDreamer: Text Guided SVG Generation with Diffusion Model

XiMing Xing[1], HaiTao Zhou[1], Chuang Wang[1], Jing Zhang[1], Qian Yu[1⋆], Dong Xu[2]

[1]School of Software, Beihang University

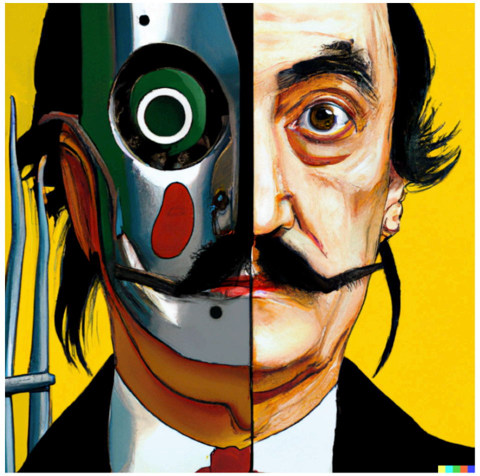
[2]Department of Computer Science, The University of Hong Kong

CVPR

JUNE 17-21, 2024

SEATTLE, WA

Motivation


vibrant portrait painting of Salvador Dalí with a robotic half face


a shiba inu wearing a beret and black turtleneck

Sprouts in the shape of text ‘Imagen’ coming out of a fairytale book.


A photo of a Shiba Inu dog with a backpack riding a bike. It is wearing sunglasses and a beach hat.

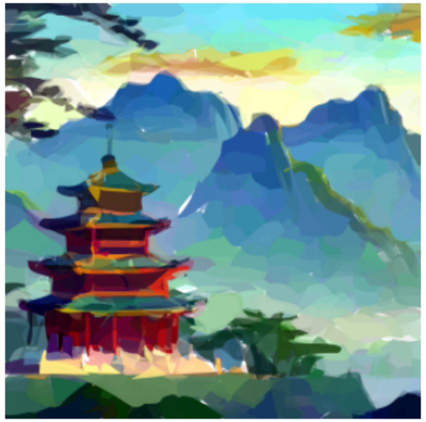

A high contrast portrait of a very happy fuzzy panda dressed as a chef in a high end kitchen making dough. There is a painting of flowers on the wall behind him.

Image generation models, particularly diffusion models, have revolutionized numerous fields by enabling the creation of realistic and diverse data samples. These include various Text-to-Image models, such as ***DALLE, Imagen, Stable Diffusion, SDXL, Deepfloyd IF, Midjourney*** and so on.

Task

"A painting of a Chinese temple with mountains in the background"

Text Prompt



Output
(format: SVG)

- The absence of large datasets like *ImageNet or LAION* within the SVG domain has hindered the advancement of text-to-SVG models.
- Text-to-SVG Synthesis using 2D raster prior.

Ours: SVGDreamer

Text Prompt

"Sydney opera house, oil painting, by Van Gogh"

"A picture of a bald eagle. low-ploy"

"Pikachu, in pastel colors"

"An owl stands on a branch"

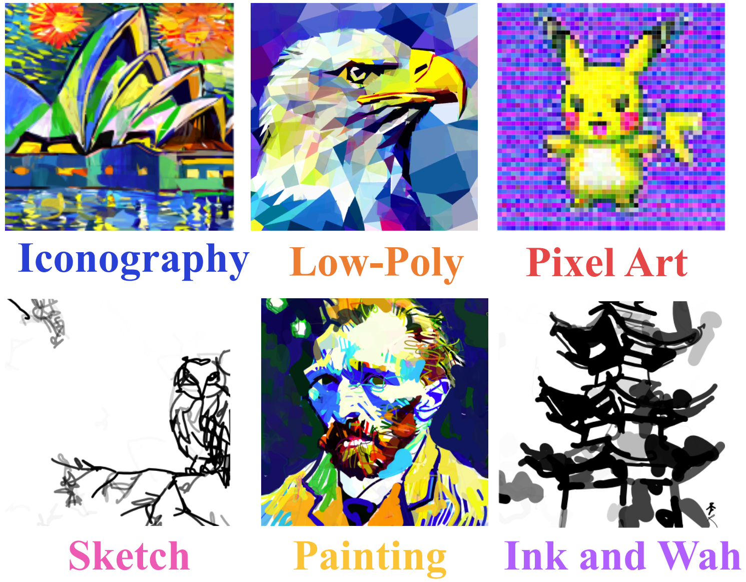
"A self-portrait by Van Gogh"

"Big Wild Goose Pagoda"

Rendering

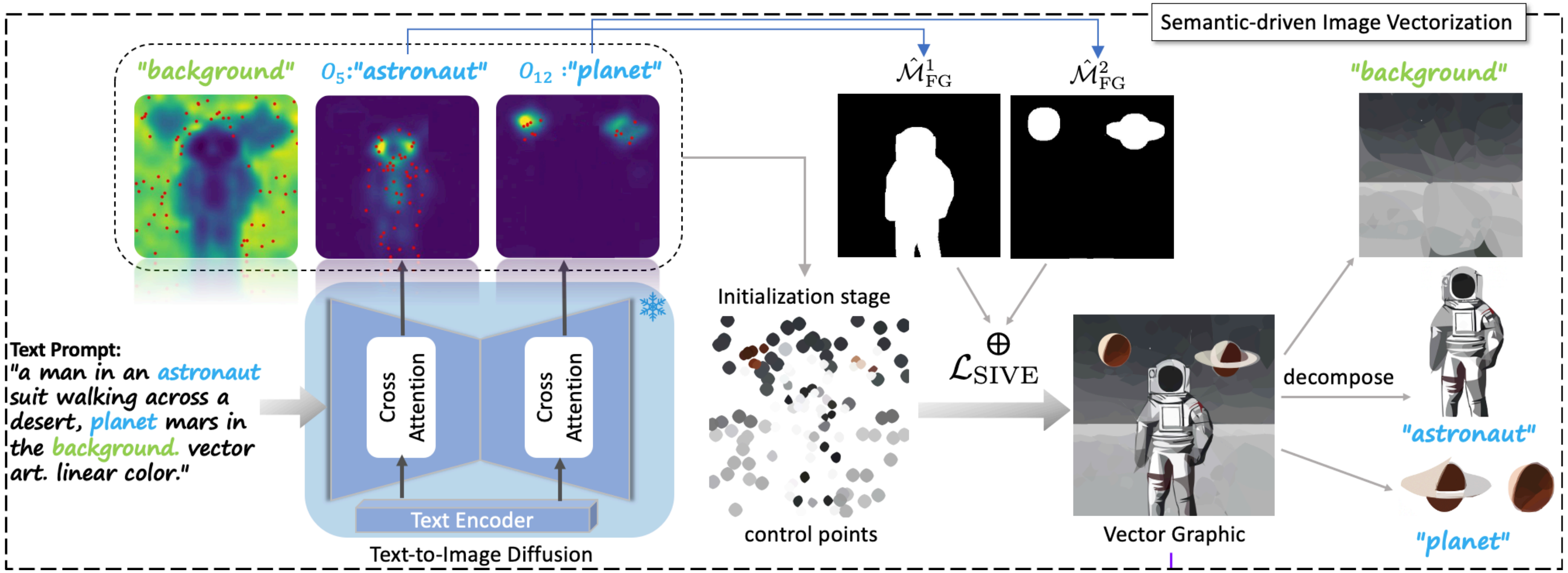
Output

(format: SVG)



- Editability: SVG paths are decoupled at the semantic level
- Visual Quality and Diversity: Overcome the over-smooth and over-saturation by SDS
- Aesthetic Appeal: The object is complete, and the layout is reasonable

SIVE: Semantic-driven Image Vectorization



1. Primitive Initialization

$$\mathcal{M}_{\mathrm{BG}} = 1 - (\sum_{i=1}^{O} \mathcal{M}^i_{\mathrm{FG}});$$

$$\mathcal{M}^i_{\mathrm{FG}} = \mathrm{softmax}(QK_i^T)/\sqrt{d}$$

2. Semantic-aware Optimization

$$\mathcal{L}_{\mathrm{SIVE}} = \sum_i^{O} \left(\hat{\mathcal{M}}_i \odot I - \hat{\mathcal{M}}_i \odot \mathbf{x}\right)^2$$

SIVE: Semantic-driven Image Vectorization

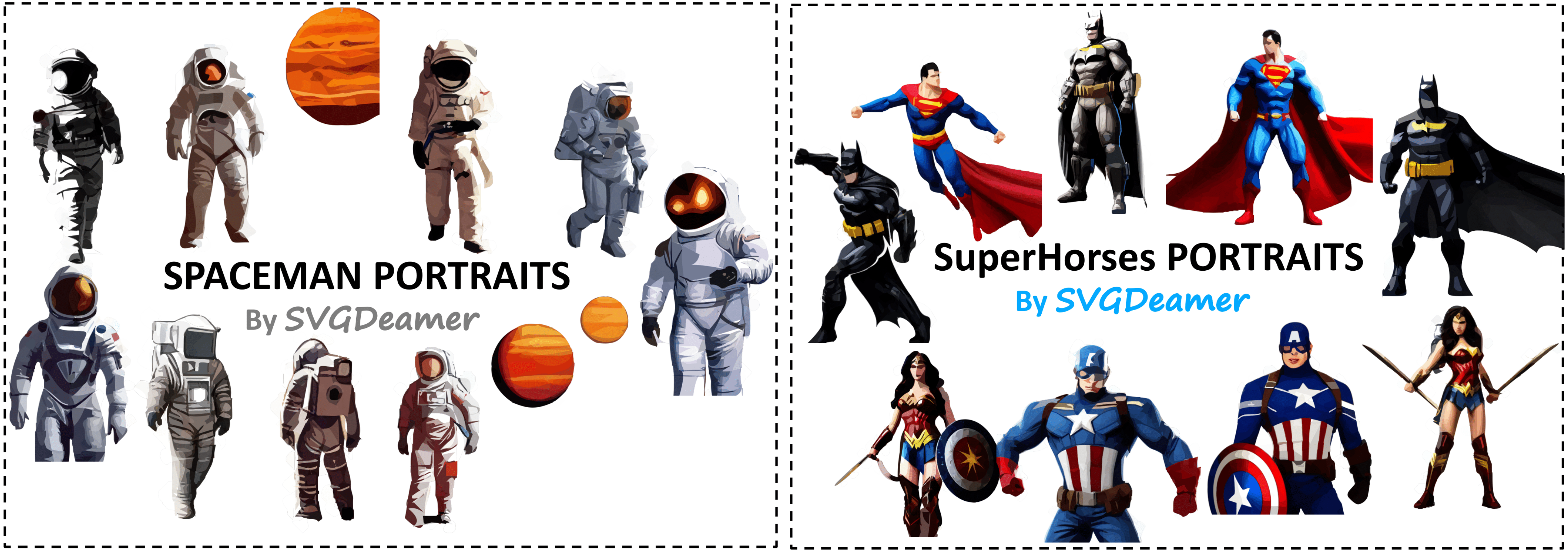


Figure: Examples of vector assets created by SIVE.

SIVE: Semantic-driven Image Vectorization

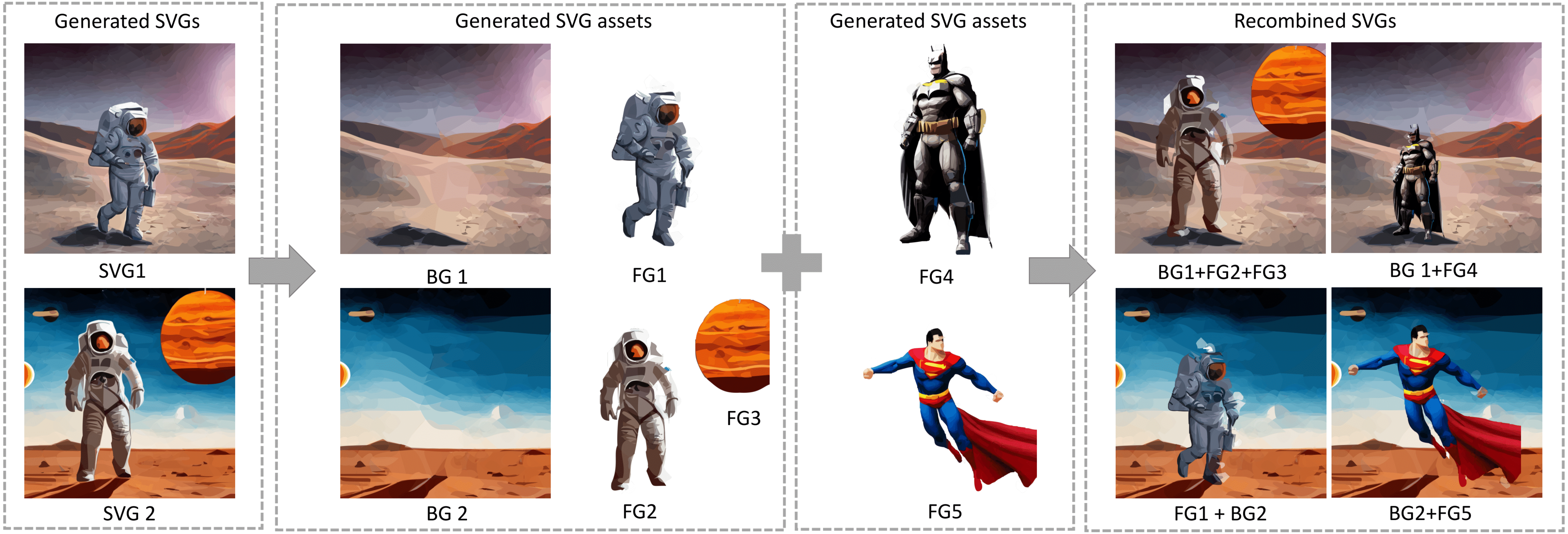
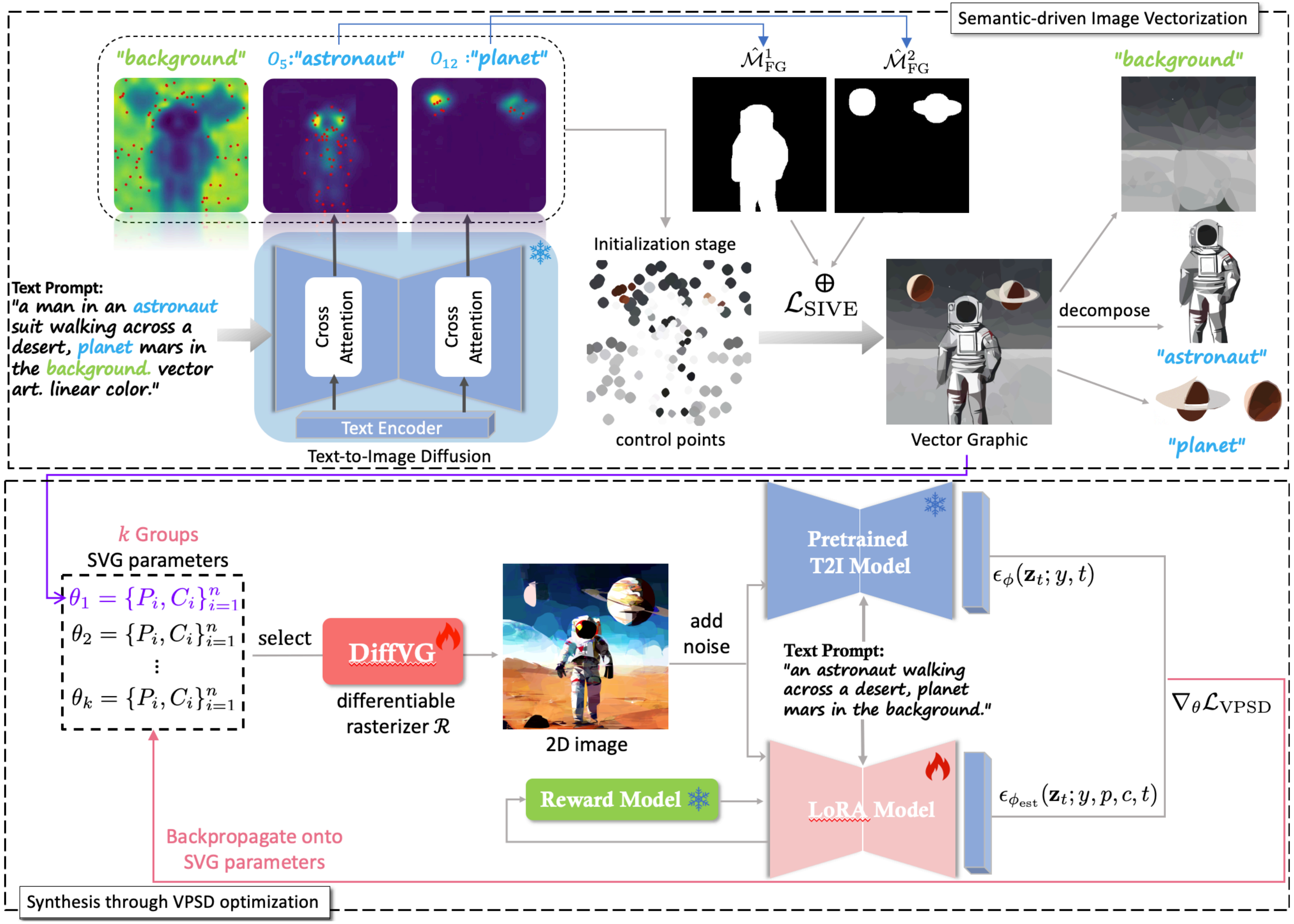


Figure: Examples showcasing the editability of the results generated by our SVGDreamer.

Framework Overview



SIVE synthesizes vector graphics with decoupled semantic hierarchy based on textual prompts.
It consists of two parts:
1. Primitive Initialization;
2. Semantic-aware Optimization.

VPSD synthesizes high-quality, diverse, and aesthetically appealing vector graphics through score distillation from pre-trained diffusion models using vector examples.

Vectorized Particle-based Score Distillation

VPSD Loss :

$$\nabla_\theta \mathcal{L}_{\text{VPSD}}(\phi, \phi_{\text{est}}, \mathbf{x} = \mathcal{R}(\theta)) \triangleq \mathbb{E}_{t,\epsilon,p,c}\left[w(t)(\epsilon_\phi(\mathbf{z}_t; y, t) - \epsilon_{\phi_{\text{est}}}(\mathbf{z}_t; y, p, c, t))\frac{\partial \mathbf{z}}{\partial \theta}\right]$$

LoRA Loss :

$$\mathcal{L}_{\text{lora}} = \mathbb{E}_{t,\epsilon,p,c}\left\|\epsilon_{\phi_{\text{est}}}(\mathbf{z}_t; y, p, c, t) - \epsilon\right\|_2^2$$

Reward Loss :

$$\mathcal{L}_{\text{reward}} = \lambda \mathbb{E}_y\left[\psi(r(y, g_{\phi_{\text{est}}}(y)))\right]$$

Total Loss :

$$\min_\theta \nabla_\theta \mathcal{L}_{\text{VPSD}} + \mathcal{L}_{\text{lora}} + \lambda_{\text{r}} \mathcal{L}_{\text{reward}}$$

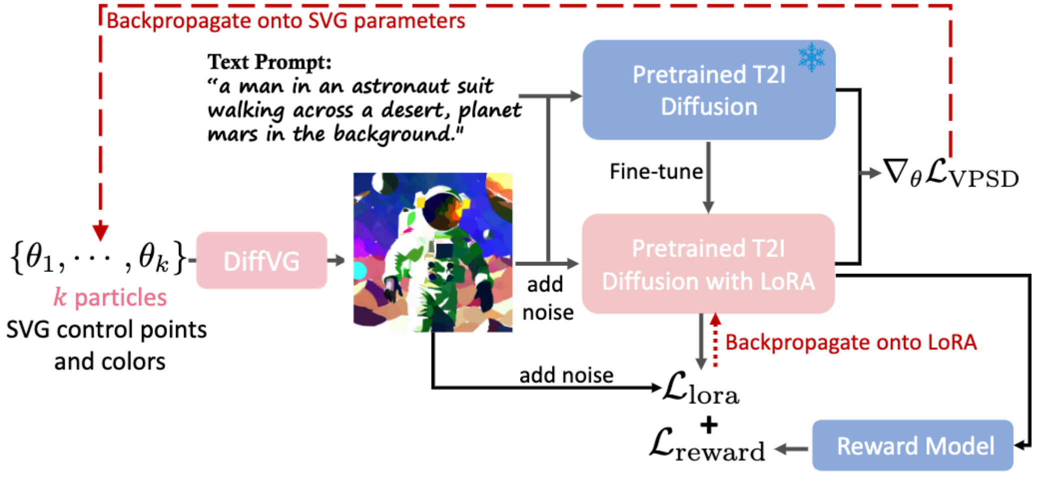
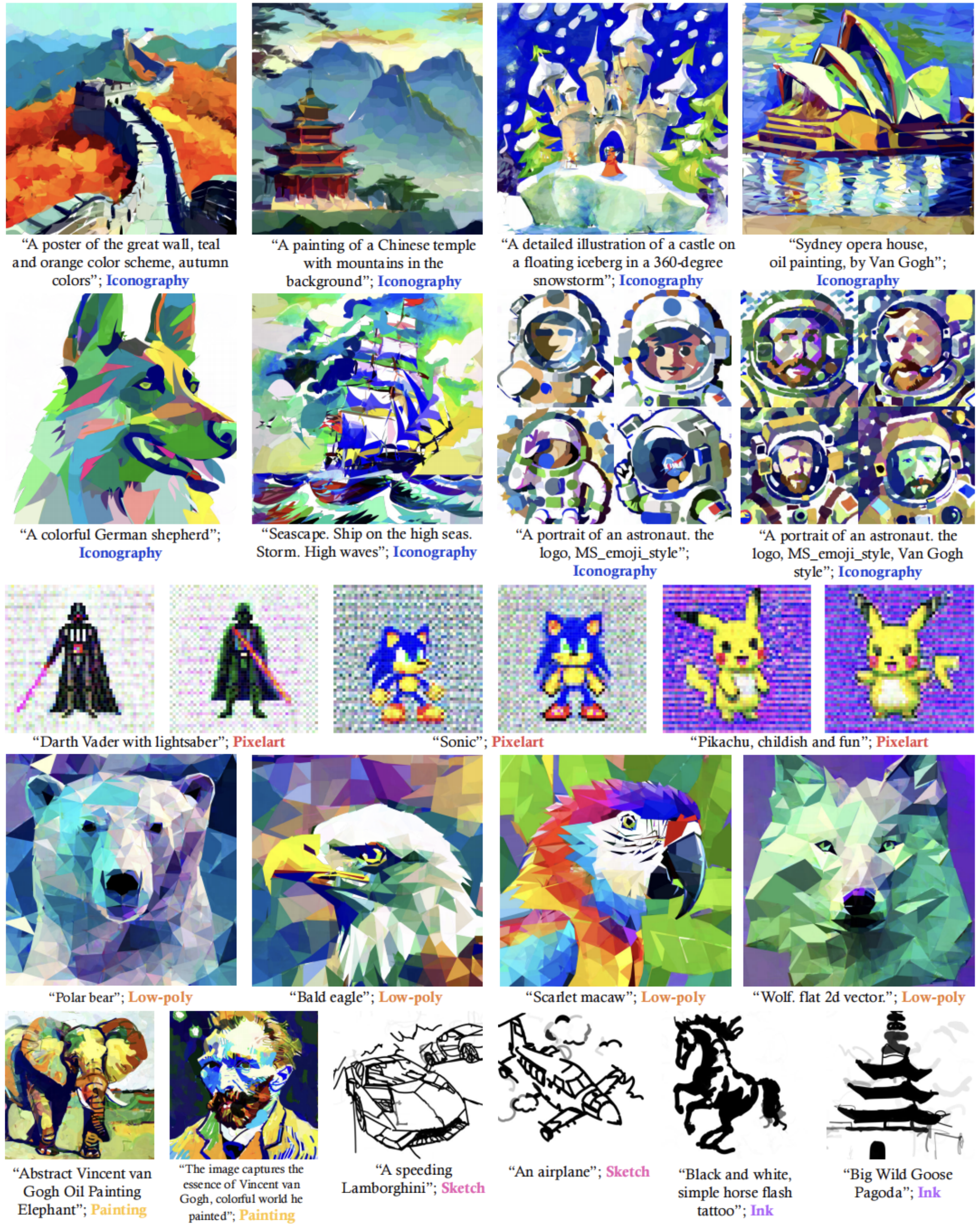


Figure. The process of VPSD

Visualization results

- SVGDreamer supports **six** styles of SVG results: Iconography, Pixel-Art, Ink and Wash, Low-poly, Sketch and Painting.
- Different color suffixes represent different SVG style types, and these style types **do not** need to be given in prompt, just by controlling the vector primitives.



Qualitative Evaluation

Figure: Qualitative comparison of different methods.

More applications

Stable Diffusion | DeepFloyd IF | GlyphControl | TextDiffuser | SVGDreamer (Ours)

Text Prompt:
"a man in an astronaut suit walking across a desert, planet mars in the background."

Glyph:
SPACE
ADVENTURE
Journey To Mars

Text Prompt:
"an astronaut walking across a jungle, cold color palette, muted colors ."

Glyph:
SPACE
ADVENTURE
Journey To Mars

Text Prompt:
"a beautiful girl in a business suit walking across a street, street view in the background."

Glyph:
SHE
POWER

Figure: Comparison of synthetic posters generated by different methods. The input text prompts and glyphs to be added to the posters are displayed on the left side.

Thank you for your attention!

https://ximinng.github.io/SVGDreamer-project/

Project

https://github.com/ximinng/SVGDreamer

Code