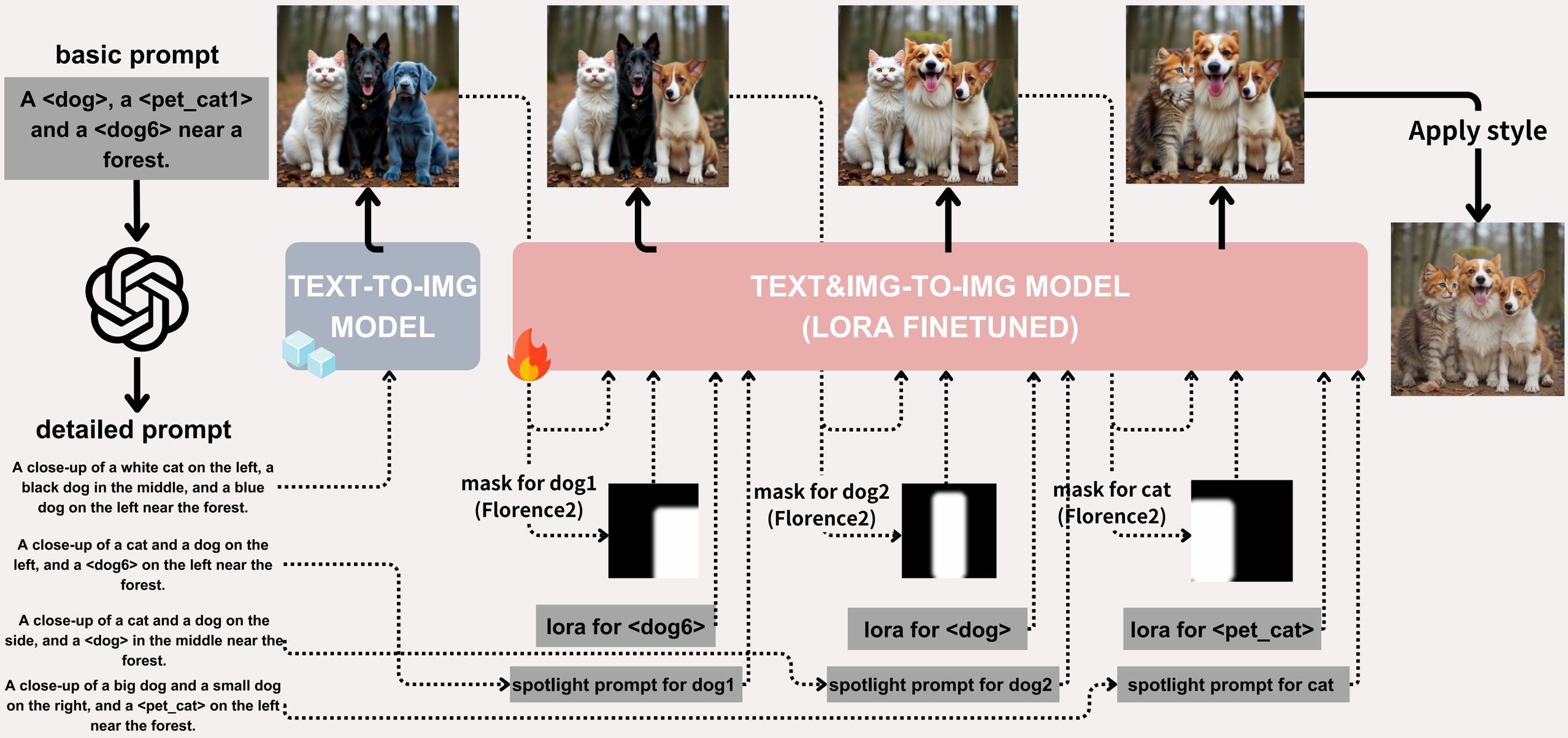
StageLight: Sequentially Targeted and Guided Editing for Layered Inpainting and Generative Hierarchical Transitions

B10901184 錡亭勳　B10901058 傅渝翔　B10901163 張顥譽　B10901176 蔡弘祥

ABSTRACT

Decomposing the prompt for multi-layer inpainting using LLM can significantly improve the model's capability, including creativity, precision, and customization similarity.

ARCHITECTURE



EXPERIMENT

StageLight(Ours)





CLIP-I : 75.20
CLIP-T : 33.18

OmniGen





CLIP-I : 74.77
CLIP-T : 31.97

Custom-Diffusion





CLIP-I : 71.51
CLIP-T : 29.72

Textual Inversion





CLIP-I : 70.80
CLIP-T : 29.53

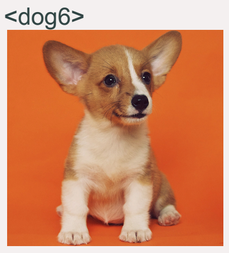
Reference

A <cat2> on the right and a <dog6> on the left.

<cat2>

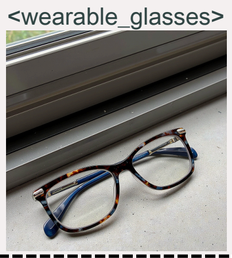


<dog6>

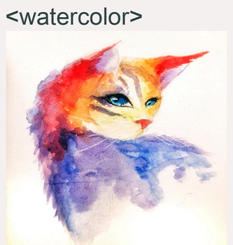


A <cat2> wearing <wearable_glasses> in a <watercolor> style.

<wearable_glasses>



<watercolor>



Baselines

- Textual Inversion
  - Find a special token to represent the reference images.
  - Designed for **single concept**.
- Custom Diffusion
  - Find a special token to represent the reference images.
  - In addition, fine-tune the cross-attention layer of the stable diffusion.
  - Designed for **multiple concepts**.
- OmniGen
  - **Zero-shot** generates images with special concepts.
  - Designed for **multiple concepts**.

Interaction



Interaction



Interaction



Camera



Camera / Depth



Style



Features

- Interaction
- Camera
- Depth
- Style

ABLATION STUDY

LD+SP(Ours)



W/o LD



SP on <dog>

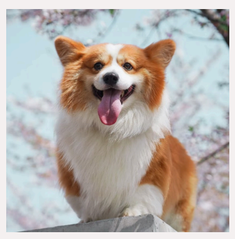


W/o SP on <dog>



LD : Layer Decomposition
SP : Spotlight Prompt

Reference



<dog>



<pet_cat>



<dog6>

Prompt: A <dog>, a <pet_cat1> and a <dog6> near a forest.