OMG：ORTHOGONAL MIX-OF-SHOW GENERATIVE MODEL

資訊碩一 林榮盛（R13922039）、資訊碩一 鄭妤婕（R13922016）、網媒碩一 張翔（R13944035）、網媒碩一 張凱量（R13944054）

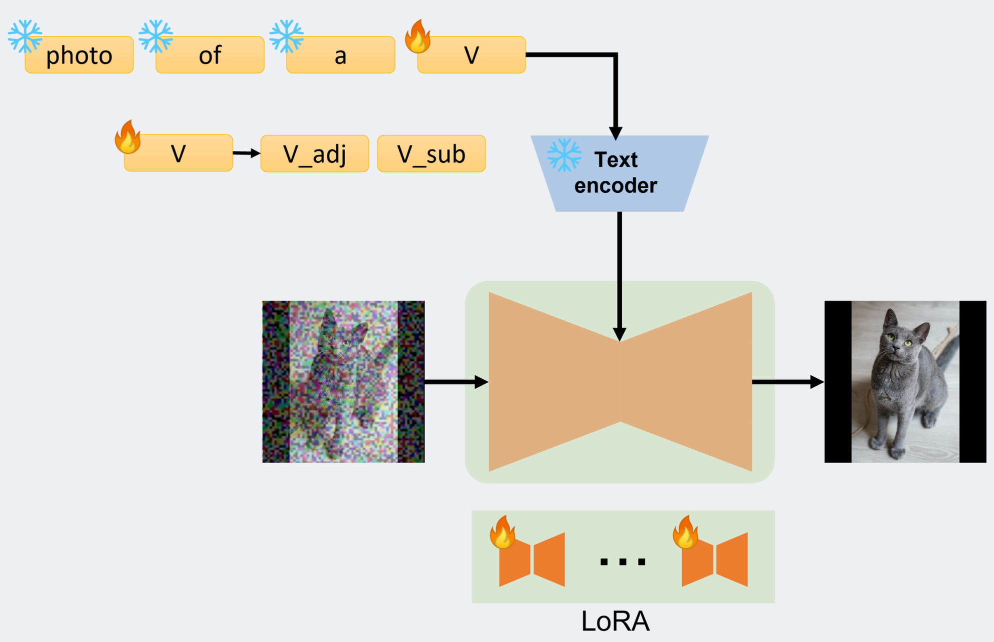
DLCV 2024 Fall - Team DarkMagic

INTRODUCTION

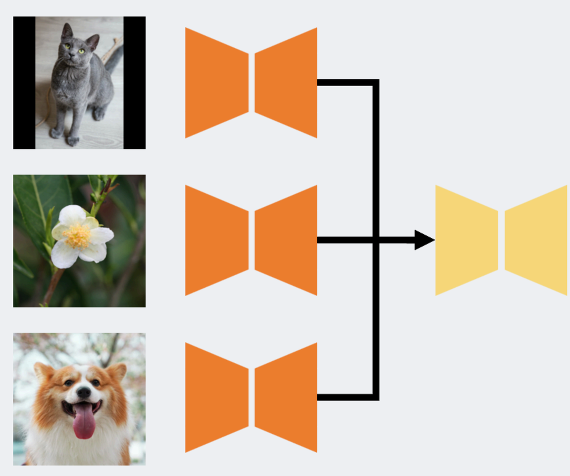
We adopted the Mix-of-Show approach, aiming to achieve comparable results while reducing the training parameters of model and introducing LoRA orthogonality. After merging the LoRA of individually trained concepts, we incorporated region control to manage the generation locations of each concept, preventing them from blending into one another.

MODEL

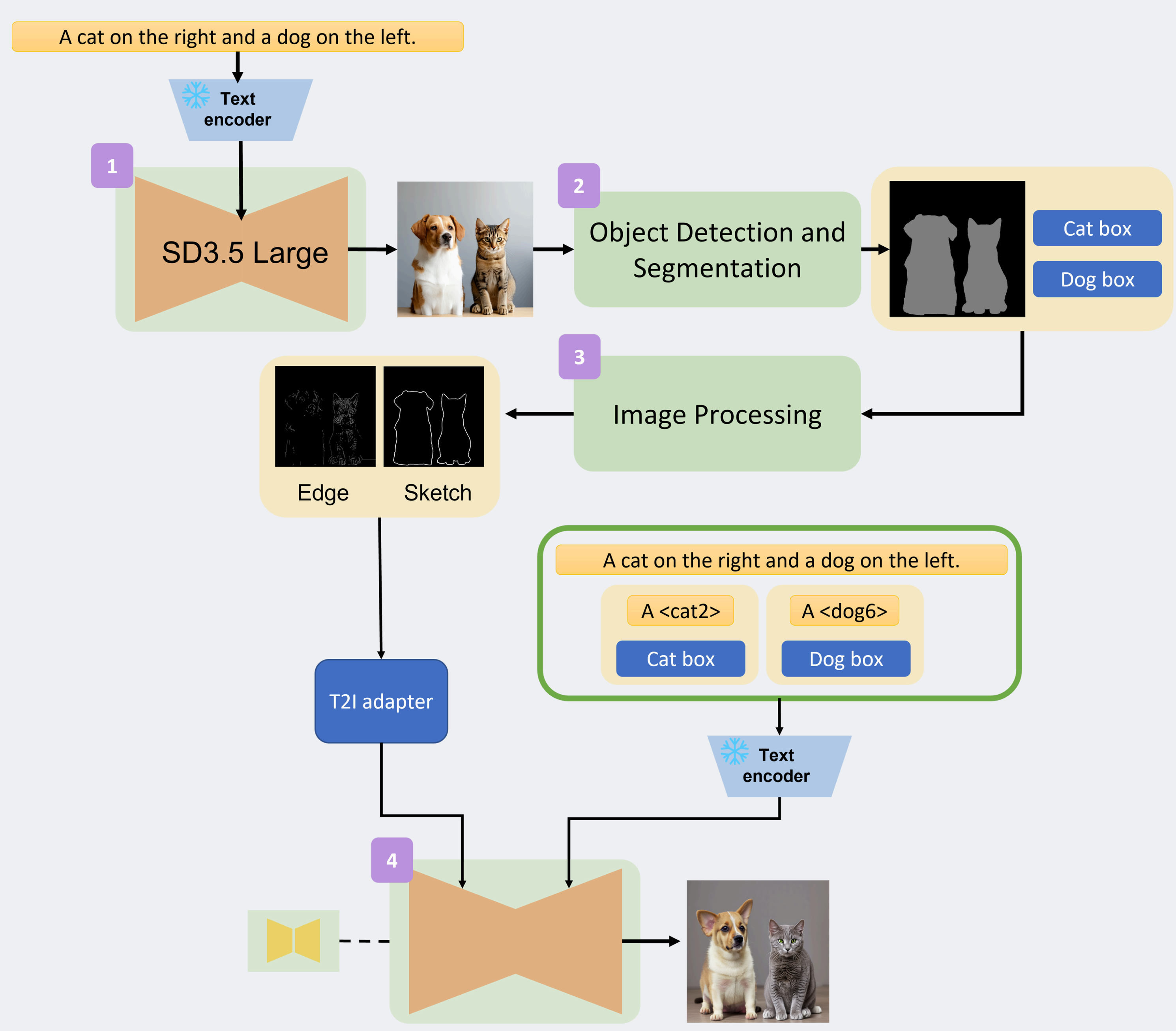
A. Training Stage



B. Fusion Stage



C. Inference Stage



METHOD

In the training phase, we took a different approach from the original paper.

A. Orthogonality

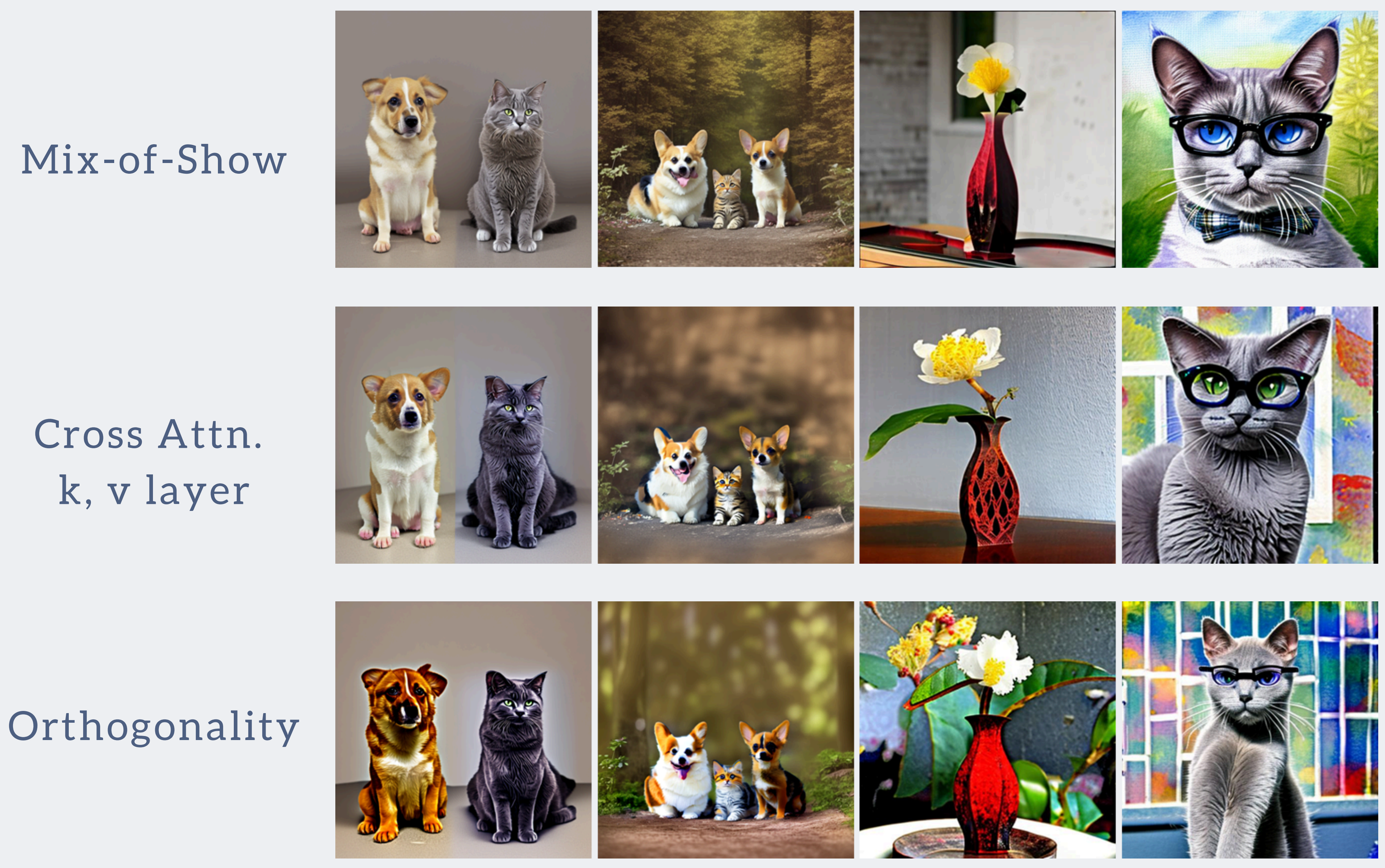
we made the LoRA components in the Unet orthogonal for each concept and directly added the LoRA weights during the fusion stage, bypassing gradient fusion.

B. Cross Attn. k, v layer

We focused on training only the k and v layers of the cross-attention in the Unet to reduce the number of trainable parameters.

These two strategies were designed to maintain performance while minimizing computational resource usage.

VISUALIZATION (EXPERIMENTS)



We tried two alternatives to Mix-of-Show: orthogonality to save gradient fusion time, and training only cross attention k and v layers, reducing parameters by 75%, while preserving image quality. We found that for similar concepts, applying region control is necessary to achieve better results.

CONCLUSION

By adopting the Mix-of-Show approach, we reduced training parameters by training only k and v layers and utilizing LoRA orthogonality, while maintaining image quality. Region control was crucial for separating similar concepts, ensuring clarity and preventing blending. Our results demonstrate a balance between efficiency and performance, validating the effectiveness of our methods.

REFERENCE

[1] Mix-of-Show: Decentralized Low-Rank Adaptation for Multi-Concept Customization of Diffusion Models

[2] Multi-Concept Customization of Text-to-Image Diffusion

[3] Orthogonal Adaptation for Modular Customization of Diffusion Models