Supplementary Material for 'Mix-of-Show'

Yuchao Gu¹, Xintao Wang³, Jay Zhangjie Wu¹, Yujun Shi², Yunpeng Chen², Zihan Fan², Wuyou Xiao², Rui Zhao¹, Shuning Chang¹, Weijia Wu¹, Yixiao Ge³, Ying Shan³, Mike Zheng Shou^{1*}

¹Show Lab, ²National University of Singapore ³ARC Lab, Tencent PCG https://showlab.github.io/Mix-of-Show

Contents

1	Data	aset and Implementation Details	2
	1.1	Dataset Details	2
	1.2	Implementation Details	2
2	Qua	ntitative and Qualitative Evaluation	2
	2.1	Evaluation Setting	2
	2.2	Quantitative Results	3
	2.3	Qualitative Results	3
3	Lim	itation and Future Work	3
	3.1	Limitation	3
	3.2	Future Work	6
	3.3	Potential Negative Society Impact	6

^{*}Corresponding Author.

³⁷th Conference on Neural Information Processing Systems (NeurIPS 2023).

1 Dataset and Implementation Details

1.1 Dataset Details

Previous works, such as Dreambooth [1] and Custom Diffusion [2], have primarily focused on object customization. In contrast, our research encompasses a more comprehensive investigation that involves characters, objects, and scenes. To facilitate our experiments, we curate a dataset comprising 19 different concepts, including 6 real-world characters, 5 anime characters, 6 real-world objects, and 2 real-world scenes. The object part is borrowed from Dreambooth [1] and Custom Diffusion [2].

1.2 Implementation Details

Pretrained Models. Due to the well-known quality issues associated with Stable-Diffusion v1-5 on human faces, we adopt the Chilloutmix² as the pretrained model for real-world concepts. Additionally, we employe the Anything-v4³ as the pretrained model for anime concepts. To ensure fair comparisons with other methods, we run all comparison methods using the same pretrained model.

Single-Client Concept Tuning. In the implementation of ED-LoRA tuning, we incorporate LoRA layers into the linear layers of all attention modules within the text encoder and Unet. Throughout our experiments, we maintain a consistent rank of r = 4 for the LoRA layers for simplicity. To optimize the different components, we utilize the Adam optimizer [3] with specific learning rates: 1e-3 for text embedding, 1e-5 for the text encoder, and 1e-4 for the Unet. We use a 0.01 noise offset for all experiments, which we have found to be crucial for encoding stable identity.

Center-Node Concept Fusion. In the center-node concept fusion, we apply layer-wise optimization to the layer connected with the LoRA layer. Each layer for optimization is initialized with pretrained weights, and we use the LBFGS optimizer [4]. In detail, we optimize the text encoder layer through 500 steps, while the Unet layer requires 50 steps for optimization.

Sample Details. In all the experiments and evaluations conducted in this paper, we utilize the DPM-Solver [5] with 20 sampling steps. To filter out undesired variations in diffusion models, we employ the same negative prompt for both our and the comparison methods during sampling: "longbody, lowres, bad anatomy, bad hands, missing fingers, extra digit, fewer digits, cropped, worst quality, low quality."

Running Times. In single-client concept tuning, the process of tuning each concept takes approximately 10-20 minutes on two Nvidia-A100 GPUs, taking into account variations in data volume. As for the center-node concept fusion, it takes 30 minutes on a single Nvidia-A100 GPU to merge 14 concepts within the pretrained model.

2 Quantitative and Qualitative Evaluation

2.1 Evaluation Setting

Our evaluation focuses on investigating the each concept in the **single-concept tuned model** and **the center-node fused model**. To assess the performance, we employ the evaluation metric, which includes image-alignment and text-alignment, as outlined in Custom Diffusion [2]. Specifically, for text-alignment, we evaluates the text-image similarity of the sampled image with the corresponding sample prompt in the CLIP feature space [6] by CLIP-Score toolkit⁴. For image-alignment, we evaluate the pairwise image similarity between the sampled image and the target concept data in the CLIP Image feature space [6].

For each concept, we utilize 20 evaluation prompts, which can be roughly categorized into four types: Recontextualization, Restylization, Interaction, and Property Modification. In Recontextualization, we assess the concept's performance by changing its context to different settings, such as the Eiffel Tower or Mount Fuji. In Restylization, we explore the concept's ability to adapt to various artistic styles. In Interaction, we investigate the concept's capability to interact with other objects, such as

²https://civitai.com/models/6424/chilloutmix

³https://huggingface.co/andite/anything-v4.0/tree/main

⁴https://github.com/jmhessel/clipscore

	prompts for characters	prompts for pets	prompts for table
	A photo of <tok> on the beach, small waves, detailed</tok>	A <tok>, in the swimming pool</tok>	A <tok>, in the swimming pool</tok>
	symmetric face, beautiful composition	A <tok>, in front of Eiffel tower</tok>	A <tok>, in front of Eiffel tower</tok>
	A <tok>, in front of Eiffel tower</tok>	A <tok>, near the mount fuji</tok>	A <tok>, near the mount fuji</tok>
Recontextualization	A <tok>, near the mount fuji</tok>		A <tok>, in the forest</tok>
	A <tok>, in the forest</tok>	A <tok>, walking on the street</tok>	A <tok>, walking on the street</tok>
	A <tok>, walking on the street</tok>	-	-
		A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine</tok>	
Deatedleastic a	A watercolor painting of a <tok></tok>		A watercolor painting of a <tok></tok>
Restylization	A painting of a <tok> in the style of Vincent Van Gogh</tok>		A painting of a <tok> in the style of Vincent Van Gogh</tok>
	A painting of a <tok> in the style of Claude Monet</tok>		A painting of a <tok> in the style of Claude Monet</tok>
	A <tok> in the style of Pixel Art</tok>	A <tok> in the style of Pixel Art</tok>	A <tok> in the style of Pixel Art</tok>
	A <tok> sit on the chair</tok>		A <tok> sit on the chair</tok>
Interaction	A <tok> ride a horse</tok>		A <tok> on the boat</tok>
interaction	A <10K> ride a horse A <tok>, wearing a headphone</tok>		A <tok>, wearing a headphone</tok>
	A <tok>, wearing a sunglass</tok>		A <tok>, wearing a sunglass</tok>
	A <tok>, wearing a Santa hat</tok>	A <tok> playing with a ball</tok>	A <tok> playing with a ball</tok>
	A smiling <tok></tok>		A sad <tok></tok>
	An angry <tok></tok>		An angry <tok></tok>
Property Change			A running <tok></tok>
	A jumping <tok></tok>		A jumping <tok></tok>
	A <tok> is lying down</tok>	A <tok> is lying down</tok>	A <tok> is lying down</tok>
	prompts for chair	prompts for vase	prompts for scene
	prompts for chair A <tok>, in the snow</tok>	A <tok>, in the snow</tok>	A <tok>, in the snow</tok>
	A <tok>, in the snow A <tok>, at night</tok></tok>	A <tok>, in the snow A <tok>, at night</tok></tok>	A <tok>, in the snow A <tok>, at night</tok></tok>
Recontextualization	A <tok>, in the snow A <tok>, at night</tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn</tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn</tok></tok></tok>
Recontextualization	A <tok>, in the snow A <tok>, at night</tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day</tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day</tok></tok></tok></tok>
Recontextualization	A <tok>, in the snow A <tok>, at night A <tok>, in autumn</tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn</tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn</tok></tok></tok>
Recontextualization	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning</tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning</tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning</tok></tok></tok></tok></tok>
Recontextualization	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine</tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine</tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine</tok></tok></tok></tok></tok></tok>
	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, a autumn A <tok>, in a sumny day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok></tok></tok></tok></tok></tok></tok></tok></tok>
	A <tok>, in the snow A <tok, at="" night<br="">A <tok, autumn<br="" in="">A <toks, a="" day<br="" in="" sunny="">A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh</tok></tok></tok></tok></tok></toks,></tok,></tok,></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumny day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok gogh<="" in="" of="" style="" td="" the="" van="" vincent=""><td>A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumn day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh</tok></tok></tok></tok></tok></tok></tok></tok></tok></td></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumn day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh</tok></tok></tok></tok></tok></tok></tok></tok></tok>
	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet</tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet</tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet</tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
	A <tok>, in the snow A <tok, at="" night<br="">A <tok, autumn<br="" in="">A <toks, a="" day<br="" in="" sunny="">A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh</tok></tok></tok></tok></tok></toks,></tok,></tok,></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumny day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok gogh<="" in="" of="" style="" th="" the="" van="" vincent=""><th>A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumn day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh</tok></tok></tok></tok></tok></tok></tok></tok></tok></th></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumn day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh</tok></tok></tok></tok></tok></tok></tok></tok></tok>
	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok art<br="" fuel="" in="" of="" style="" the="">A gjirl near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at hight A <tok>, in autumn A <tok>, in a sumn day A <tok>, in a sumn day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor pushting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> In the style of Pixel Art A girl near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumny day A <tok>, in a sumny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A cTOK> In the style of Pixel Art A girl near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
Restylization	A <tok>, in the snow A <tok>, at night A <tok>, a nught A <tok>, in autumn A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Pixel Art A CJK> in the style of Pixel Art A girl near the <tok> A bov near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A boy near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in a sumny day A <tok>, in a sumny day A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok>, in the style of Pixel Art A girl near the <tok> A bay near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
Restylization	A <tok>, in the snow A <tok>, at night A <tok>, in atutum A <tok>, in atutum A <tok>, in atutum A <tok>, in a sunny day A <tok>, in hunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A boy near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A dog near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <toks, in="" snow<br="" the="">A <toks, at="" night<br="">A <toks, autumn<br="" in="">A <toks, a="" day<br="" in="" sumny="">A <toks, a="" day<br="" in="" sumny="">A <toks, and="" in="" lightning<br="" thunder="">A <toks, 2077,="" 3d="" 4k,="" cyberpunk="" engine<br="" in="" render="" unreal="">A watercolor painting of a <toks A painting of a <toks gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <toks gogh<br="" in="" of="" style="" the="" van="" vincent="">A <toks art<br="" in="" of="" pixel="" style="" the="">A girl near the <toks A dog near the <toks< td=""></toks<></toks </toks></toks></toks></toks </toks,></toks,></toks,></toks,></toks,></toks,></toks,>
Restylization	A <tok>, in the snow A <tok>, at night A <tok>, a summ A <tok>, in a summ day A <tok>, in a summ day A <tok>, in a summ day A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A cTOK> in the style of Pixel Art A style near the <tok> A boy near the <tok> A dag near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A cTOK in the style of Pixel Art A <tok> in the stYle of Vince A cta near the <tok> A cta the art he <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at hight A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A dog near the <tok> A cat near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
Restylization	A <tok>, in the snow A <tok>, at night A <tok>, in atutum A <tok>, in atutum A <tok>, in atutum A <tok>, in a sunny day A <tok>, in hunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A boy near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A dog near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <toks, in="" snow<br="" the="">A <toks, at="" night<br="">A <toks, autumn<br="" in="">A <toks, a="" day<br="" in="" sumny="">A <toks, a="" day<br="" in="" sumny="">A <toks, and="" in="" lightning<br="" thunder="">A <toks, 2077,="" 3d="" 4k,="" cyberpunk="" engine<br="" in="" render="" unreal="">A watercolor painting of a <toks A painting of a <toks gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <toks gogh<br="" in="" of="" style="" the="" van="" vincent="">A <toks art<br="" in="" of="" pixel="" style="" the="">A girl near the <toks A bog near the <toks< th=""></toks<></toks </toks></toks></toks></toks </toks,></toks,></toks,></toks,></toks,></toks,></toks,>
Restylization	A <tok>, in the snow A <tok>, at night A <tok>, a summ A <tok>, in a summ day A <tok>, in a summ day A <tok>, in a summ day A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A cTOK> in the style of Pixel Art A style near the <tok> A boy near the <tok> A dag near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A cTOK in the style of Pixel Art A <tok> in the stYle of Vince A cta near the <tok> A cta the art he <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at hight A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A boy near the <tok> A cat near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
Restylization	A <tok>, in the snow A <tok>, at night A <tok>, a summ A <tok>, in a summ day A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A <tok art<br="" in="" of="" pixel="" style="" the="">A <tok art<br="" in="" of="" pixel="" style="" the="">A sign near the <tok> A dog near the <tok> A cat near the <tok> Many people near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, at night A <tok>, in a summ A <tok>, in a summ A <tok>, in a summ A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A dog near the <tok> A dog near the <tok> Many people near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Loude Monet A <tok> in the style of Pixel Art A dog near the <tok> A dog near the <tok> A cat near the <tok> Many people near the <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
Restylization Interaction	A <tok>, in the snow A <tok>, at night A <tok>, a stutumn A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, tyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A totK> in the style of Pixel Art A dig near the <tok> A dag near the <tok> A <tok> in the Style A <tok> in rainbow colors A <tok> in rainbow colors</tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Claude Monet A <tok> in the style of Pixel Art A girl near the <tok> A dog near the <tok> A dog near the <tok> A cort near the <tok> A <tok> in rainbow colors</tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, at night A <tok>, in a summ day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok =="" funcent="" gogh<br="" in="" of="" style="" the="" van="">A painting of a <tok =="" funcent="" gogh<br="" in="" of="" style="" the="" van="">A day in the style of Pixel Art A cTOK> in the style of Pixel Art A dog near the <tok> A cot near the <tok> A <tok> in rainbow colors</tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
Restylization	A <tok>, in the snow A <tok>, at night A <tok>, at night A <tok>, autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, in the style of line and lightning A autor of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A context of the style of Pixel Art A corr b <tok> A boy near the <tok> A dog near the <tok> A cot near the <tok> A <tok> in rainbow colors A <tok> made of metal A <tose <tok="" of="" view=""></tose></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at hight A <tok>, at hight A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Laude Monet A <tok> in the style of Pixel Art A girl near the <tok> A dog near the <tok> A dag near the <tok> A <tok> in the <tok> A <tok> in the <tok> A <tok> in the dof metal</tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in a sunny day A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Laude Monet A <tok> in the style of Pixel Art A girl near the <tok> A dog near the <tok> A can the art the <tok> Many people near the <tok> A <tok> in tabobw colors A <tok had="" metal<="" of="" td=""></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>
Restylization Interaction	A <tok>, in the snow A <tok>, at night A <tok>, a stutumn A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, in a sunny day A <tok>, tyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A painting of a <tok> in the style of Vincent Van Gogh A totK> in the style of Pixel Art A dig near the <tok> A dag near the <tok> A <tok> in the Style A <tok> in rainbow colors A <tok> in rainbow colors</tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in sunny day A <tok>, in thunder and lightning A autor and a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A contrast of the style of Pixel Art A <tok art<br="" in="" of="" pixel="" style="" the="">A dig rin ear the <tok> A dog near the <tok> A dat near the <tok> Many people near the <tok> A <tok> in rainbow colors A <tok> in ainbow colors A <tok <tok="" a="" in=""> A close view of <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>	A <tok>, in the snow A <tok>, at night A <tok>, at night A <tok>, in autumn A <tok>, in autumn A <tok>, in autumn A <tok>, in thunder and lightning A <tok>, cyberpunk 2077, 4K, 3d render in unreal engine A watercolor painting of a <tok> A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A painting of a <tok gogh<br="" in="" of="" style="" the="" van="" vincent="">A contrast of the style of Pixel Art A <tok> in the style of Pixel Art A dig near the <tok> A dog near the <tok> Many people near the <tok> A <tok> in rainbow colors A <tok> in anibow colors A <tok made="" metal<br="" of="">A close view of <tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok></tok>

Figure 1: Summarization of our evaluation prompts for each concept.

associations or actions like sitting on a chair. In Property Modification, we modify the internal state of the concept, including expressions or states like running or jumping. Each type consists of 5 prompts, some of which are borrowed from previous work, resulting in a total of 20 prompts per concept. We sample 50 images for each prompt, ensuring reproducibility by fixing the random seed within the range of [1, 50]. This yields a total of 1000 images for each concept. The evaluation prompts for each concept are presented in Fig. 1.

2.2 Quantitative Results

According to the evaluation setting described in Sec. 2.1, we have compiled the complete evaluation results for each concept, which are summarized in Table. 1. The summarized results of different categories can be found in Table. 1 of the main paper.

2.3 Qualitative Results

The qualitative comparison of Mix-of-Show and other methods on the single-client tuned model and center-node fused model is presented in Fig. 2. From the results, it is evident that the LoRA experiences the most significant loss of concept identity after concept fusion. Additionally, due to the limited tunability of positions in the P+ and Custom Diffusion, they exhibit oversaturated results or semantic collapse in some examples. Conversely, Mix-of-Show consistently achieves the best concept identity and quality across various examples, while also minimizing the loss of identity after the center-node concept fusion.

3 Limitation and Future Work

3.1 Limitation

The first limitation is related to regionally controllable sampling, as depicted in Fig. 3(a), where attributes from one region may influence another due to the encoding of some attributes in the global

Single-Concept Mode	el Methods	Cat (5)	DogA (5)	Chair (5)	Table (4)	DogB (5)	Vase (6)	Mear
	Upper Bound	0.832	0.821	0.795	0.792	0.821	0.807	0.811
Text-alignment	P+ [7]	0.828	0.801	0.726	0.696	0.799	0.776	0.771
rext-angiment	Custom Diffusion [2]	0.784	0.744	0.698	0.689	0.786	0.768	0.745
	LoRA [8]	0.761	0.673	0.655	0.670	0.779	0.779	0.720
	Mix-of-Show (Ours)	0.771	0.703	0.666	0.671	0.772	0.759	0.724
	Lower Bound	0.753	0.755	0.674	0.682	0.769	0.692	0.721
Image-alignment	P+ [7]	0.783	0.753	0.809	0.810	0.825	0.761	0.790
image-angiment	Custom Diffusion [2]	0.869	0.830	0.825	0.888	0.848	0.794	0.842
	LoRA [8]	0.859	0.871	0.872	0.917	0.887	0.778	0.864
	Mix-of-Show (Ours)	0.874	0.864	0.890	0.879	0.889	0.811	0.868
Fused Model	Methods	Cat (5)	DogA (5)	Chair (5)	Table (4)	DogB (5)	Vase (6)	Mear
	Upper Bound	0.832	0.821	0.795	0.792	0.821	0.807	0.81
- Text-alignment	P+ [7]	0.828	0.801	0.726	0.696	0.799	0.776	0.771
Text-anginnent	Custom Diffusion [2]	0.756	0.735	0.726	0.724	0.782	0.758	0.747
	LoRA [8]	0.827	0.803	0.757	0.766	0.814	0.801	0.795
-	Mix-of-Show (Ours)	0.801	0.738	0.673	0.709	0.786	0.761	0.745
	Lower Bound	0.753	0.755	0.674	0.682	0.769	0.692	0.721
- Image-alignment	P+ [7]	0.783	0.753	0.809	0.810	0.825	0.761	0.790
	C . D'CC . [0]	0.871	0.813	0.774	0.794	0.823	0.773	0.808
88	Custom Diffusion [2]	0.071						
88	LoRA [8]	0.805	0.800	0.761	0.778	0.808	0.715	0.778

(a) Quantitative results from single-concept model and center-node fused model on *real-world objects*.

Single-Concept Mod	lel Methods	Potter (14)	Hermione (15)	Thanos (15)	Hinton (14)	Lecun (17)	Bengio (15)	Mean
	Upper Bound	0.765	0.776	0.765	0.765	0.765	0.765	0.767
T	P+[7]	0.640	0.744	0.622	0.708	0.693	0.711	0.686
Text-alignment	Custom Diffusion [2]	0.654	0.720	0.594	0.717	0.683	0.674	0.674
	LoRA [8]	0.580	0.696	0.568	0.716	0.681	0.684	0.654
	Mix-of-Show (Ours)	0.575	0.650	0.562	0.665	0.680	0.662	0.632
	Lower Bound	0.485	0.458	0.510	0.422	0.510	0.441	0.471
.	P+[7]	0.778	0.608	0.809	0.582	0.614	0.629	0.670
Image-alignment	Custom Diffusion [2]	0.737	0.663	0.852	0.627	0.694	0.710	0.714
	LoRA [8]	0.866	0.679	0.917	0.683	0.716	0.705	0.761
	Mix-of-Show (Ours)	0.869	0.785	0.921	0.731	0.723	0.782	0.802
Fused Model	Methods Po	otter (14)	Hermione (15)	Thanos (15)	Hinton (14)	Lecun (17)	Bengio (15)	Mean
Fused Model	Methods Pe Upper Bound	otter (14) I 0.765	Hermione (15) 0.776	Thanos (15) 0.765	Hinton (14) 0.765	Lecun (17) 0.765	Bengio (15) 0.765	Mean 0.767
-		. ,			. ,		0 ()	
Fused Model Text-alignment	Upper Bound	0.765	0.776	0.765	0.765	0.765	0.765	0.767
-	Upper Bound P+ [7]	0.765 0.640	0.776	0.765	0.765	0.765	0.765	0.767
-	Upper Bound P+ [7] Custom Diffusion [2]	0.765 0.640 0.604	0.776 0.744 0.678	0.765 0.622 0.624	0.765 0.708 0.699	0.765 0.693 0.651	0.765 0.711 0.641	0.767 0.686 0.650
-	Upper Bound P+ [7] Custom Diffusion [2] LoRA [8]	0.765 0.640 0.604 0.693	0.776 0.744 0.678 0.717	0.765 0.622 0.624 0.656	0.765 0.708 0.699 0.694	0.765 0.693 0.651 0.714	0.765 0.711 0.641 0.725	0.767 0.686 0.650 0.700
Text-alignment	Upper Bound P+ [7] Custom Diffusion [2] LoRA [8] Mix-of-Show (Ours)	0.765 0.640 0.604 0.693 0.632	0.776 0.744 0.678 0.717 0.677	0.765 0.622 0.624 0.656 0.611	0.765 0.708 0.699 0.694 0.673	0.765 0.693 0.651 0.714 0.678	0.765 0.711 0.641 0.725 0.701	0.767 0.686 0.650 0.700 0.662
-	Upper Bound P+ [7] Custom Diffusion [2] LoRA [8] Mix-of-Show (Ours) Lower Bound	0.765 0.640 0.604 0.693 0.632 0.485	0.776 0.744 0.678 0.717 0.677 0.458	0.765 0.622 0.624 0.656 0.611 0.510	0.765 0.708 0.699 0.694 0.673 0.422	0.765 0.693 0.651 0.714 0.678 0.510	0.765 0.711 0.641 0.725 0.701 0.441	0.767 0.686 0.650 0.700 0.662 0.471
Text-alignment	Upper Bound P+ [7] Custom Diffusion [2] LoRA [8] Mix-of-Show (Ours) Lower Bound P+ [7]	0.765 0.640 0.604 0.693 0.632 0.485 0.778	0.776 0.744 0.678 0.717 0.677 0.458 0.608	0.765 0.622 0.624 0.656 0.611 0.510 0.809	0.765 0.708 0.699 0.694 0.673 0.422 0.582	0.765 0.693 0.651 0.714 0.678 0.510 0.614	0.765 0.711 0.641 0.725 0.701 0.441 0.629	0.767 0.686 0.650 0.700 0.662 0.471 0.670

(b) Quantitative results from single-concept model and center-node fused model on *real-world characters*.

Single-Concept Model	Methods	Rock (20)	Pyramid (20)	Mean	Fused Model	Methods	Rock (20)	Pyramid (20)	Mean
	Upper Bound	0.869	0.798	0.834	Text-alignment	Upper Bound	0.869	0.798	0.834
Text-alignment	P+ [7] Custom Diffusion [2] LoRA [8]	0.801 0.809 0.737	0.716 0.686 0.697	0.759 0.748 0.717		P+ [7] Custom Diffusion [2] LoRA [8]	0.801 0.793 0.786	0.716 0.682 0.734	0.759 0.738 0.760
	Mix-of-Show (Ours)	0.754	0.677	0.716		Mix-of-Show (Ours)	0.770	0.702	0.736
	Lower Bound	0.672	0.517	0.595	Image-alignment	Lower Bound	0.672	0.517	0.595
Image-alignment	P+ [7] Custom Diffusion [2] LoRA [8]	0.821 0.808 0.863	0.770 0.800 0.784	0.796 0.804 0.824		P+ [7] Custom Diffusion [2] LoRA [8]	0.821 0.742 0.810	0.770 0.757 0.728	0.796 0.750 0.769
	Mix-of-Show (Ours)	0.859	0.857	0.858		Mix-of-Show (Ours)	0.832	0.844	0.838

(c) Quantitative results from single-concept model and center-node fused model on *real-world scenes*.

Table 1: *Text-alignment* and *image-alignment* of the single-client tuned model and the center-node fused model. The upper bound of text-alignment and the lower bound of image-alignment are computed by replacing the concept's token with its class token and sampling using the pretrained model. For instance, to assess the upper-bound text alignment and lower-bound image alignment for the concept " V^{dogA} ," we substitute " V^{dogA} " in the sample prompts with the class token "dog" and sample it using the pretrained model. (N) means each concept has N images for tuning.



Center-Node Fused Model

Figure 2: Qualitative comparison of Mix-of-Show *vs.* P+ [7], Custom Diffusion [2], and LoRA [8]. Our Mix-of-Show demonstrates superior concept identity and quality in single-concept tuned model and achieves the least identity loss after concept fusion.

^{atman}. V^{thanos} and V^{ironman} standing

V^{hermione} is walking near the mount fuji.



(a) Limitation of Attribute Leakage

(b) Limitation in Small Face

Figure 3: Limitation of Mix-of-Show. (a) Attribute leakage in regionally controllable sampling. (b) Limitation in small face generation.

embedding. This issue can be partially alleviated by specifying undesired attributes using negative prompts for each region.

The second limitation concerns the center-node concept fusion, which requires a relatively lengthy time to merge concepts. The primary bottleneck in this process arises from the presence of large spatial features in the Unet layer during layer-wise optimization.

The final limitation relates to the generation of small faces. In the case of Stable Diffusion, the information loss in the VAE can affect the generation of high-quality full-body characters, especially in the small face region, resulting in a loss of facial details, as shown in Fig. 3(b). To mitigate this limitation, increasing the sample size can be a potential solution.

3.2 Future Work

Our Mix-of-Show framework enables the reusability and scalability of tuned concepts, facilitating the creation of complex multi-concept compositions. In future work, it would be interesting to explore how Mix-of-Show can enhance storybook generation by generating character and object interactions across various plots. Furthermore, as Mix-of-Show supports stable identity encoding, it has the potential to assist in concept customization for video or 3D scenarios.

3.3 Potential Negative Society Impact

This project aims to provide the community with an effective tool for decentralized creation of high-quality customized concept models and the ability to reuse and combine different concepts to compose complex images. However, a risk exists wherein malicious entities could exploit this framework to create deceptive interactions with real-world figures, potentially misleading the public. This concern is not unique to our approach but rather a shared consideration in other multi-concept customization methodologies. One potential solution to mitigate such risks involves adopting methods similar to anti-dreambooth [9], which introduce subtle noise perturbations to the published images to mislead the customization process. Additionally, applying unseen watermarking to the generated images could deter misuse and prevent them from being used without proper recognition.

References

[1] Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. arXiv preprint arXiv:2208.12242, 2022. 2

- [2] Nupur Kumari, Bingliang Zhang, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Multiconcept customization of text-to-image diffusion. arXiv preprint arXiv:2212.04488, 2022. 2, 4, 5
- [3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 2
- [4] Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. Mathematical programming, 45(1-3):503–528, 1989.
- [5] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. arXiv preprint arXiv:2206.00927, 2022. 2
- [6] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763. PMLR, 2021. 2
- [7] Andrey Voynov, Qinghao Chu, Daniel Cohen-Or, and Kfir Aberman. *p*+: Extended textual conditioning in text-to-image generation. *arXiv preprint arXiv:2303.09522*, 2023. 4, 5
- [8] Simo Ryu. Low-rank adaptation for fast text-to-image diffusion fine-tuning. https://github. com/cloneofsimo/lora. 4, 5
- [9] Thanh Van Le, Hao Phung, Thuan Hoang Nguyen, Quan Dao, Ngoc N Tran, and Anh Tran. Anti-dreambooth: Protecting users from personalized text-to-image synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2116–2127, 2023. 6