

Supplementary Material

A. Ablation on module selection

We provide quantitative results of ablation on module to insert mapping layers. Tab. 1 indicates that the decoder-only strategy shows the strongest guidance capability since it does not harm the encoder’s feature space and only performs guidance during generation.

B. Limitation

For plugins to generate personalized concepts, *e.g.*, IP-Adapter [?], our method might not maintain the identity well like shown in Fig. 1. This is because the custom plugins work on the text-encoder other than the feature space concepts that are not directly injected into the upgraded model other than fused as guidance.

C. Qualitative result

We provide more qualitative result on Controlnet [?], T2I-Adapter [?] and LoRA [?] as shown in Fig. 2

| Plugin: ControlNet | FID ↓ | CLIP-score ↑ | Cond. Recon. ↑ |
|---------------------------|--------------|--------------|--------------------|
| Encoder & Decoder | 38.37 | 0.26 | 0.24 ± 0.15 |
| Encoder only | 37.32 | 0.26 | 0.23 ± 0.14 |
| Decoder only | 30.95 | 0.26 | 0.27 ± 0.13 |
| Plugin: LoRA | FID ↓ | CLIP-score ↑ | Style-Sim ↑ |
| Encoder & Decoder | 36.71 | 0.26 | 0.79 |
| Encoder only | 35.54 | 0.26 | 0.80 |
| Decoder only | 29.88 | 0.26 | 0.83 |

Table 1. *Ablation on module to insert mapping layers.* For encoder only and decoder only, we use the same mapping layers to map the frozen base model to the upgraded model’s encoder and decoder separately. For both the encoder and decoder, we use two identical mapping layers and insert them into the encoder and decoder.

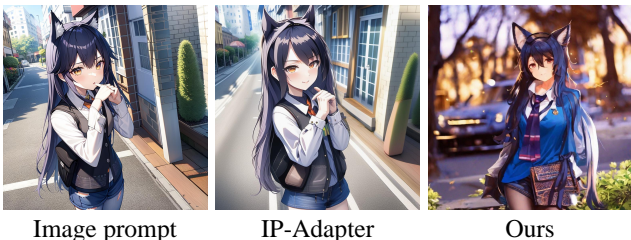


Figure 1. *Limitations.* In IP-Adapter [?], although our method can produce relatively identity-consistent results, the details, *e.g.*, clothes, are still different from the original model.

ControlNet

| | | | |
|---|--|---|--|
| <p><i>X-Adapter & Upgraded Model prompt:</i> A rose</p>  <p>SD2.1 SDXL</p> | | <p><i>X-Adapter & Upgraded Model prompt:</i> Two dears in the wild</p>  <p>SD2.1 SDXL</p> | |
| <p><i>X-Adapter & Upgraded Model prompt:</i> A cute book of spells in mystical cover</p>  <p>SD2.1 SDXL</p> | | <p><i>X-Adapter & Upgraded Model prompt:</i> Cinematic photo hamster in a bauble wearing a christmas hat</p>  <p>SD2.1 SDXL</p> | |

LoRA

shukezouma

X-Adapter Prompt: masterpiece, best quality, ultra detailed, 1girl, solo, smile, looking at viewer, holding flowers. (trigger words)
Upgraded Models prompt: masterpiece, best quality, ultra detailed, 1girl, solo, smile, looking at viewer, holding flowers.



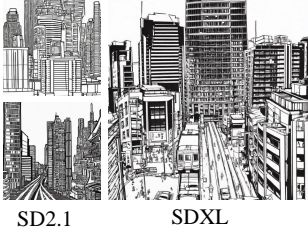
Glass sculpture

X-Adapter Prompt: realistic glasssculpture of a rabbit, translucent, transparent, detailed cityscape background, street, reflections
Upgraded Models prompt: A rabbit, detailed cityscape background, street, reflections



AnimeOutline

X-Adapter Prompt: best quality, a city, (trigger words)
Upgraded Models prompt: best quality, a city



Lego AI

X-Adapter Prompt: Perfect lighting, depth of field, (trigger word), house in the forest
Upgraded Models prompt: Perfect lighting, depth of field, house in the forest



T2I-Adapter

| | | | |
|---|--|--|--|
| <p><i>X-Adapter & Upgraded Model prompt:</i> A cat holding a gun</p>  <p>SD2.1 SDXL</p> | | <p><i>X-Adapter & Upgraded Model prompt:</i> A world of crystal, high quality, extremely detailed</p>  <p>SD2.1 SDXL</p> | |
| <p><i>X-Adapter & Upgraded Model prompt:</i> A cat riding a dolphin</p>  <p>SD2.1 SDXL</p> | | <p><i>X-Adapter & Upgraded Model prompt:</i> An apple made of chocolate, high quality, extremely detailed</p>  <p>SD2.1 SDXL</p> | |

ControlNet-Tile

X-Adapter Prompt: best quality, extremely detailed
Upgraded Models prompt: best quality, extremely detailed

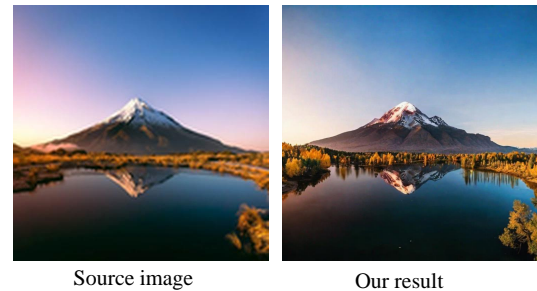


Figure 2. *Qualitative Results on Different Plugins.*