Supplementary Materials

This appendix includes our supplementary materials as follows:

- Evaluation setting in Sec. A.
- More qualitative results in Sec. B.
- Combination with ControlNet in Sec. C.
- Combination with style LoRAs in Sec. D.
- More results of layout preservation in Sec. E.
- Comparison of different base models for OMG combined with InstantID in Sec. F.
- Comparison of different base models for OMG combined with LoRA in Sec. G.
- Compare with more single image based methods in Sec. H.
- Model compression in Sec. I.
- Limitation and future work in Sec. J.

A Evaluation Setting

Our dataset for image customization comprises 10 characters and 5 objects. For each model, we employ 20 text prompts, and the evaluation prompts for each concept are presented in Fig. 1.

Prompts for characters	Prompts for pets	Prompts for seats
photo of a {}. {} selfie standing under the pink blossoms of a cherry tree. {} in a chef's outfit, cooking in a kitchen. {} paddling a canoe on a tranquil lake. {} playing with their pet dog. Photo of {} taking a shot in basketball. {} selfie with eiffel tower in the background. {} in an astronaut suit, floating in a spaceship. {} dressed in a firefighter's outfit, a raging forest fire in the background. {} wearing Victorian-era clothing, reading a book in a classic British library. {} dressed as a knight, standing in a medieval castle. long shot of {} wearing a classic cowboy outfit, riding on a horse A pointillistic painting of {}. A Monet-inspired painting of {} standing near a blooming lily pond. Anime avatar of {}. A watercolor painting of {} with mountains in the background. {} painted with thick, dramatic brush strokes, in the style of Edvard Munch. {} as a character in the Sherlock Holmes' movie. {} as Neo in the matrix movie. {} impersonating a character from the Harry Potter movie.	Photo of a {}. {} swimming in a pool. {} at a beach with a view of the seashore. {} in times square. {} is wearing sunglasses. {} is wearing a sombrero. {} is in a construction outfit. {} is playing with a ball. {} is wearing headphones. {} oil painting ghibli inspired. Painting of {} at a beach by artist claude monet. {} digital painting 3d render geometric style. Georgia O'Keeffe style {} painting. a {} in chef outfit. A {} is reading a book. a {} is acting in a play wearing a costume. a barking {}. a sleeping {}. a sad {}. a sculpture of {}.	Photo of a {}. {} near a pool. {} at a beach with a view of the seashore. {} in a garden. {} in grand canyon. {} in front of a medieval castle. {} and a coffee table. floor lamp on the side of {}. {} and an orange sofa. {} and a table with chocolate cake on it. {} oil painting ghibli inspired. {} painting by artist claude monet. a watercolor painting of {} in a forest. A digital illustration of the {}. Georgia O'Keeffe style {} painting. An orange {}. A pink {}. A red color {}. {} crochet. An egg chair in the style of {}.

Fig. 1: Summary of our evaluation prompts for each concept.

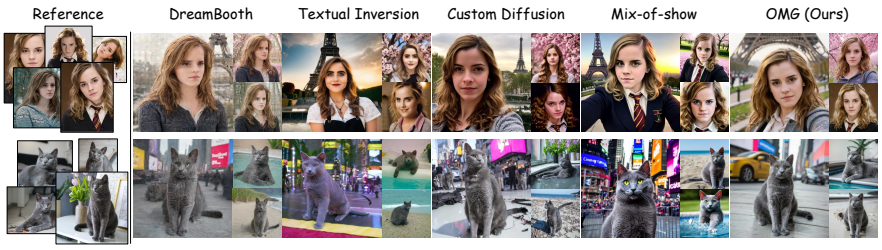

Fig. 2: More comparison results of OMG with other methods on the single-concept customization.

Fig. 3: More experimental results of OMG on the multi-concept customization.

B More Qualitative Results

We additionally present supplementary experimental results for concept customization. Single-concept customization results compared with other methods are depicted in Fig. 2. Multi-concept customization results for characters are shown in Fig. 3. Moreover, further experimental results combining characters and objects are presented in Fig. 4.

C Combination with ControlNet

The proposed method is versatile and practical, allowing for combination with various conditions using ControlNet. To validate its effectiveness, we conduct experiments by combining ControlNet with different conditions, including human pose, canny edge, and depth maps. The results are presented in Fig. 5.

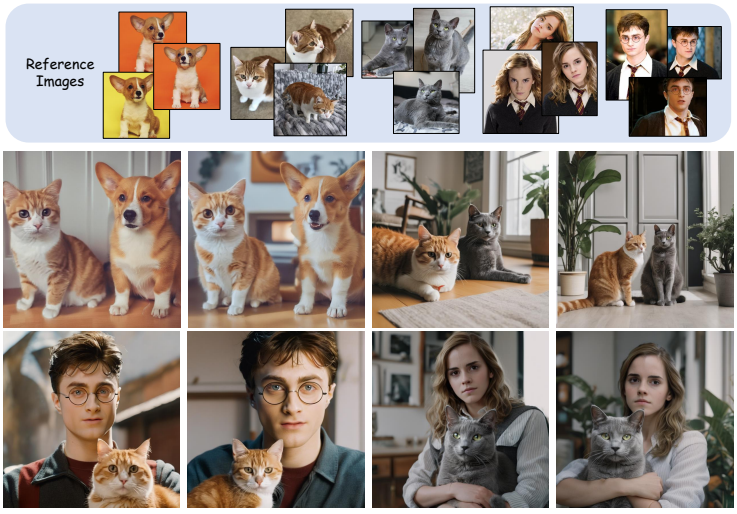

Fig. 4: Multi-concept customization results in combining characters and objects.

Table 1: Quantitative results for ablation study.

Method	OMG w RCS	OMG w/o CNB	OMG w CNB
IDA	0.141	0.518	**0.524**

D Combination with Style LoRAs

The proposed method can be combined with different style LoRAs. To validate the effectiveness, we conduct experiments to combine OMG with various LoRAs related to styles, including Anime Sketch style, Comic Book Style, and Cyberpunk Style. We utilize different single-concept customization models, such as LoRA and InstantID. The experimental results are presented in Fig. 6, which demonstrate the effectiveness of our method in combining with different styles.

E More Results of Layout Preservation

The layout preservation operation in OMG plays a crucial role in maintaining the image layout. To validate the effectiveness of layout preservation, we provide additional examples. As illustrated in Fig. 7, each row showcases the generation results of the first stage, with and without using layout preservation. The absence of layout preservation can result in the loss of objects or alterations to the overall image structure. This underscores that layout preservation contributes to generating a more reasonable image layout and enhances the overall image quality.

We also conducted ablation experiments to evaluate the effectiveness of layout preservation and compare concept noise blending (CNB) with regionally controllable sampling (RCS). We calculated the average Identity Alignment scores

Fig. 5: The experimental results of OMG combined with ControlNet under different conditions, including human pose, canny edge, and depth maps.

(IDA) of multiple concepts. The experimental results are presented in Tab. 1. The results demonstrate that layout preservation is crucial for improving Identity Alignment scores, and concept noise blending is more effective than regionally controllable sampling.

F Comparison of different base models for OMG combined with InstantID

It's interesting to note the InstantID is sensitivity to the base model and CFG scale. As shown in Fig. 8, the choice of base model and CFG scale significantly

Fig. 6: The experimental results of OMG combined with different style LoRAs.

Fig. 7: Ablation of layout preservation. Generating images with layout preservation can preserve image structure and enhance image quality.

affects the quality of the generated images. Increasing the CFG scale can lead to over-saturation in the results, and the best outcome is achieved when the CFG scale is set to 3. Additionally, using YamerMIX-v8 as the base model seems to help alleviate the over-saturation issue. These findings provide valuable insights for optimizing the InstantID method and achieving better image generation quality.

Fig. 8: Comparison results of different base models for OMG combined with InstantID.

G Comparison of different base models for OMG combined with LoRA.

To verify the robustness of OMG combined with LoRA, we conducted experiments on other models, including Realism, YamerMIX-v8, and Playground-v2. The experimental results are presented in Fig. 9. Our proposed method can generate images with multiple concepts utilizing different base models, where these concepts can belong to the same or different class, demonstrating the robustness of this approach.

H Compare with more single image based methods.

In addition to InstantID, we also compared our method with other single image based methods, including IP-Adapter and PhotoMaker. The comparison results

Fig. 9: Comparison results of different base models for OMG combined with LoRA

Fig. 10: Comparison with more single image based methods.

are presented in Fig. 10. Our proposed method, OMG, achieves the best identity preservation compared to PhotoMaker and IP-Adapter.

I Model compression.

We performed SVD and compared the visualization results of different LoRA ranks, including 0, 1, 16, 64, and 256. As expected, the required storage space increased with the increase of ranks. The visualization results of different ranks are presented in Fig. 11.

J Limitation and Future Work

While our method provides an occlusion-friendly framework for multi-concept personalization with robust identity preservation and harmonious illumination, there are several limitations to consider. Firstly, OMG may face challenges in generating high-quality small-face regions due to information loss in the VAE. Besides, the computational intensity associated with noise fusion from multiple single-concept models are noteworthy consideration, leading to slower generation.

Fig. 11: Comparison with different LoRA ranks.

Future research efforts to improve the sample speed of OMG, especially in achieving rapid, high-fidelity sampling, are commendable. The planned exploration of combining OMG with other accelerated methods to generate high-quality images with a few-step inference reflects a proactive approach toward addressing these challenges and enhancing the overall efficiency of the method.