Prompt for content	Prompt for style
`a {}`	`a dog in {} style`
`a {} in the jungle`	`a cat in {} style`
`a {} in the snow landscape`	`a dinosaur in {} style`
`a {} on the beach`	`a rabbit in {} style`
`a {} on a table`	`a capybara in {} style`
`a {} floating on a river`	`a dragon in {} style`
`a {} in downtown`	`a wolf in {} style`
`a {} with a mountain in the background`	`a fox in {} style`
`a {} with a wheat field in the background`	`a castle in {} style`
`a {} with autumn leaves`	`a robot in {} style`
`a {} with the Eiffel Tower in the background`	`a deer in {} style`
`a {} in the desert`	`a bedroom in {} style`
`a {} in space`	`a truck in {} style`
`a {} in the sky`	`a chair in {} style`
`a {} in the swimming pool`	`a car in {} style`
`a {} on the boat`	`a flower in {} style`
`a {} look out`	`a piano in {} style`
`a {} is lying down`	`a crown in {} style`
`a {} in sticker style`	`a bus in {} style`
`a {} in pixel art style`	`a person texting in {} style`
`a {} in anime style`	`a laptop in {} style`
`a {} in watercolor style`	`a lighthouse in {} style`
`a {} made of gold`	`a coffee cup in {} style`
`a {} made of silver`	`a spaceship in {} style`
`a {} made of glass, crystal`	`a train in {} style`

Figure 9. List of 50 prompts used in our experiments for evaluating content-style decomposition.

A. Appendix

This appendix includes our supplementary materials as follow:

- Validation Data in Appendix A.1
- CSD-100 Dataset Analysis in Appendix A.2
- Additional Qualitative Results in Appendix A.3
- More ablation experimental settings in Appendix A.4
- User Study in Fig. 10

A.1. Validation Data

To assess the effectiveness of our scale-wise detail analysis and ablation studies, we construct a validation dataset comprising 35 diverse concepts. These concepts are primarily curated from StyleDrop [41], B-LoRA [8], UnZipLoRA [26], and DreamBench++ [33]. This dataset allows us to systematically validate our assumptions regarding hierarchical scale representations and evaluate different experimental configurations. Fig. 12 provides an overview of the validation dataset, while Fig. 9 presents the set of 50 prompts used across all experiments.

A.2. CSD-100

The CSD-100 dataset in Fig. 11 is designed to provide a comprehensive benchmark for content-style decomposition (CSD) models.Fig. 13 illustrates the content and style distributions in the CSD-100 dataset. The dataset comprises 63 distinct objects (content) and 53 unique styles, providing a diverse set of content-style pairs for evaluating decomposition models. Fig. 13 (Left) depicts the frequency distribution of content categories, where most content types appear only once or twice, ensuring a broad variety of objects. Fig. 13 (Right) shows the distribution of styles, with certain styles occurring more frequently, particularly those with well-defined characteristics (e.g., line drawing, geometric shapes, abstract digital art). By balancing content
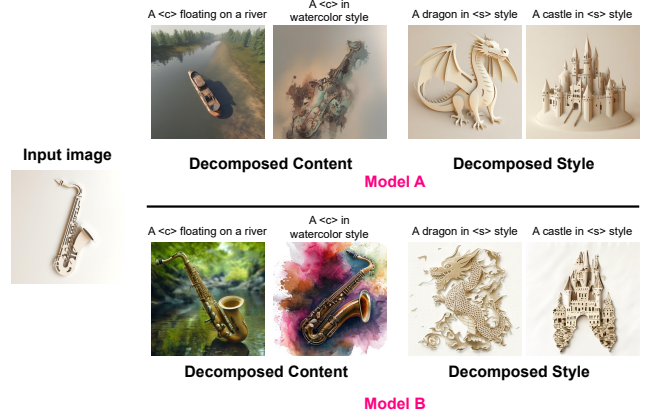

Figure 10. Our user study interface

# Augmented K, V	Content Align		Style Align		Text Align
	CSD-C	CLIP-I	CSD-S	DINO	CLIP-T
1	0.603	**0.751**	**0.564**	0.517	**0.330**
2	**0.608**	0.724	0.560	**0.521**	0.320
5	0.607	0.732	0.560	0.517	0.318

Table 6. Ablation study on the number of augmented K-V memories used in self-attention.

and style diversity while maintaining a representative distribution, CSD-100 serves as a well-curated benchmark for assessing the effectiveness of content-style decomposition models.

A.3. Additional Qualitative Results

We present additional qualitative results of our method, as shown in Fig. 14

A.4. More ablation experimental settings

Study on the Number of Augmented K-V Memories. The results in Tab. 6 show that increasing the number of K-V memories does not consistently improve performance across all metrics. We hypothesize that adding more K-V pairs complicates model distribution alignment, leading to diminishing returns. As a result, we adopt a single K-V memory as the optimal setting.

A.5. User Study

We conducted a user study with 100 participants to compare our model against an alternative method. Each participant answered 15 questions, selecting their preferred output based on five multiple-choice criteria. For each question, users evaluated two examples per output group, focusing on content and style (see Fig. 10). To ensure fairness, model outputs were anonymized and their order randomized.
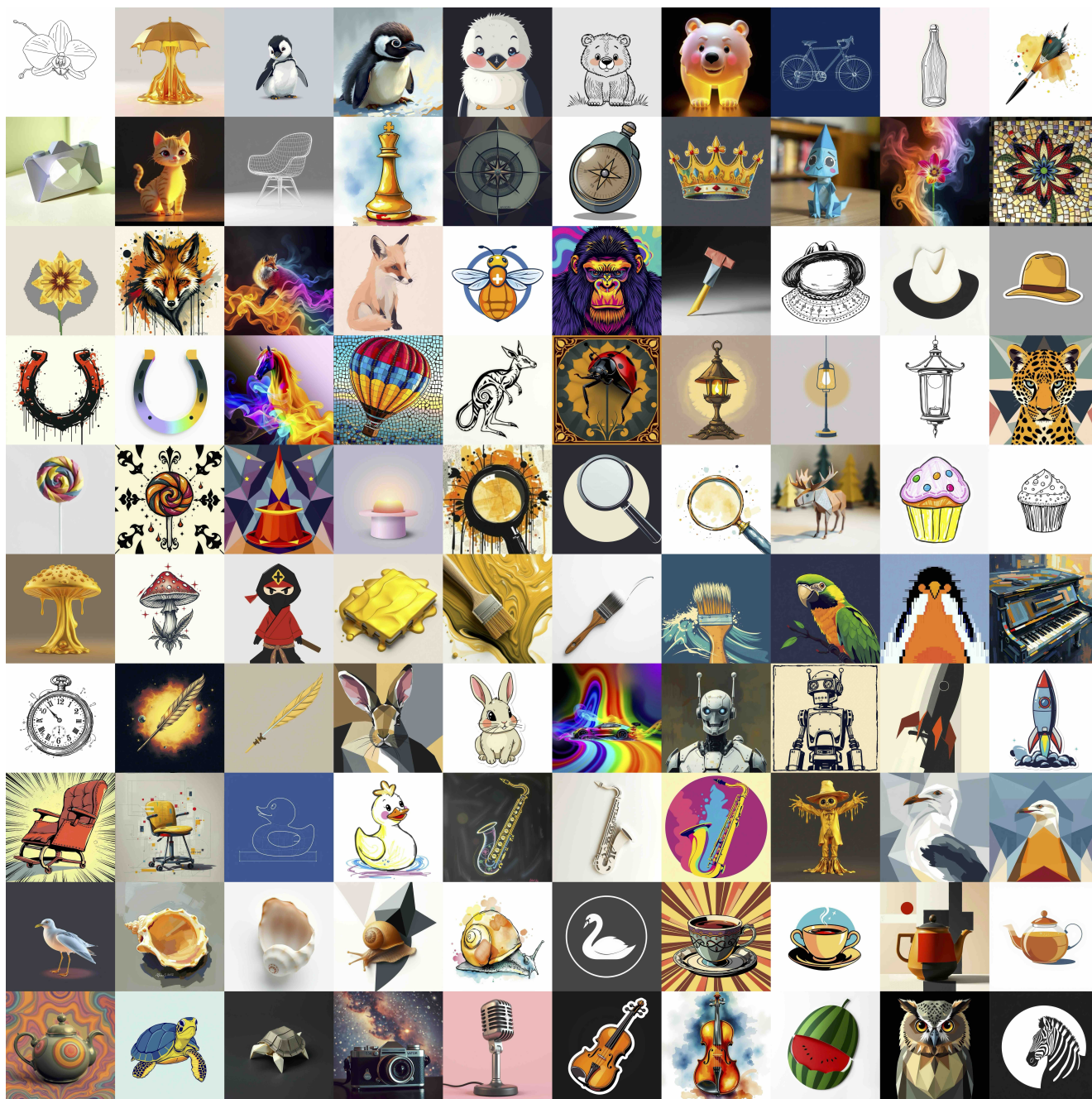
Figure 11. Visualization of the full CSD-100 dataset, showcasing its diverse content and style pairings

StyleDrop, B-LoRA

UnZipLoRA

DreamBench++

External sources

Figure 12. Overview of the validation dataset, consisting of 35 curated concepts sourced from existing personalization benchmarks.
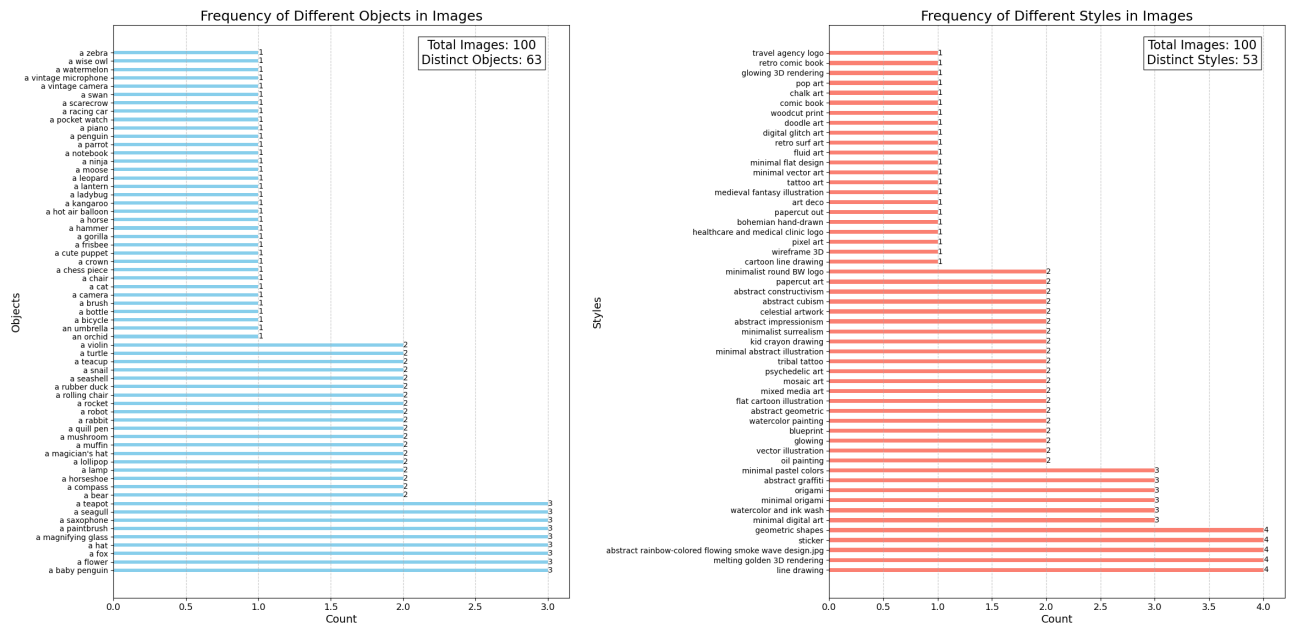
Figure 13. Distribution of content and style in the CSD-100 dataset

Content

...made of glass ...in anime style ...on the boat

...in wheat field ...made of crystal ...in bathroom

...in watercolor ...as sticker ...floating on water

...in space ...is swimming ...as sticker

...made of gold ...with autumn leaves ...on the beach

Input

Style

A castle in ... A deer in ... A girl in ...

A boy wears crown in ... A castle in ... A capybara in ...

A flower in ... A cat in ... A chair in ...

A dragon in ... A castle in ... A robot in ...
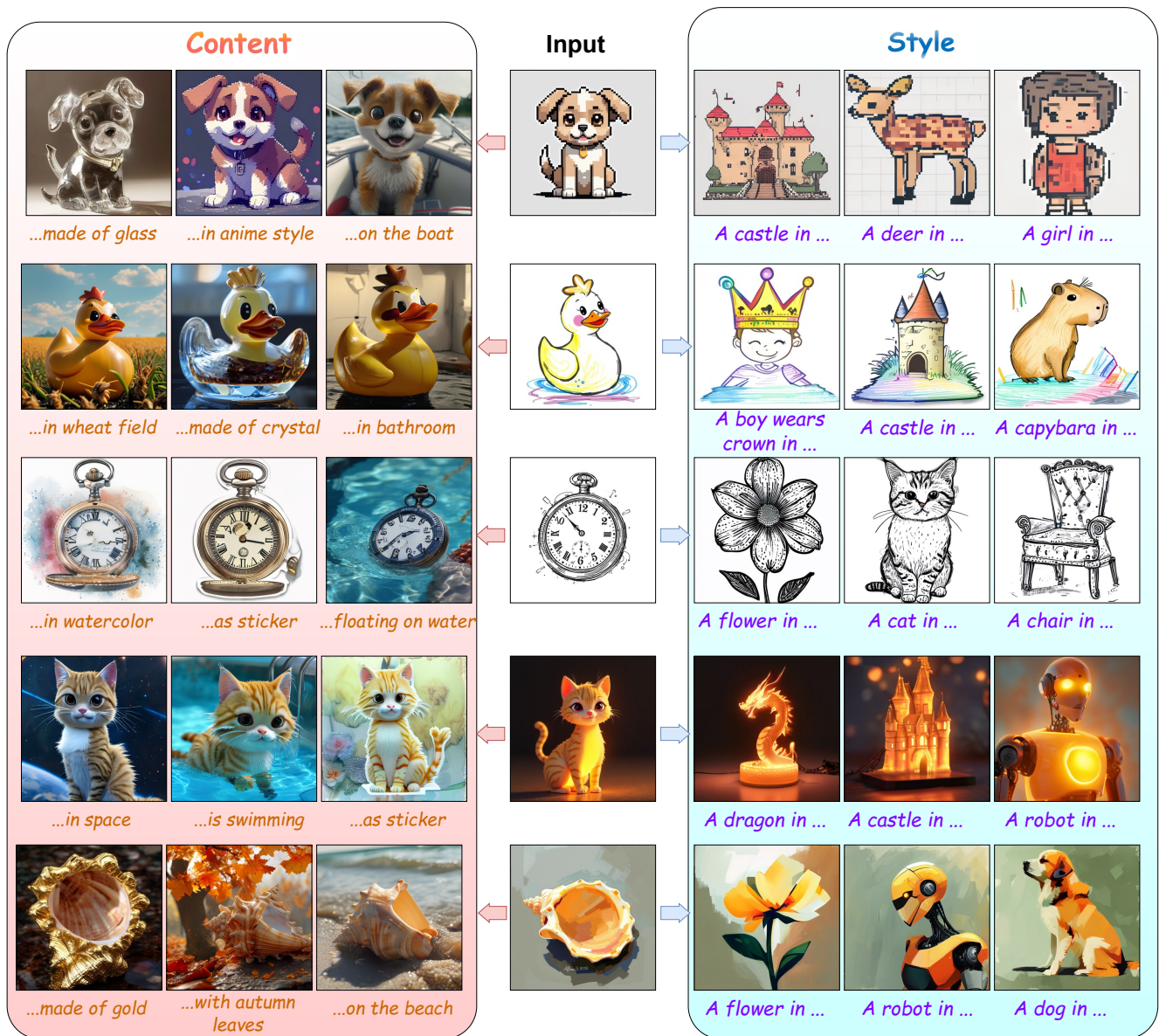
A flower in ... A robot in ... A dog in ...

Figure 14. Additional qualitative results from our CSD-VAR