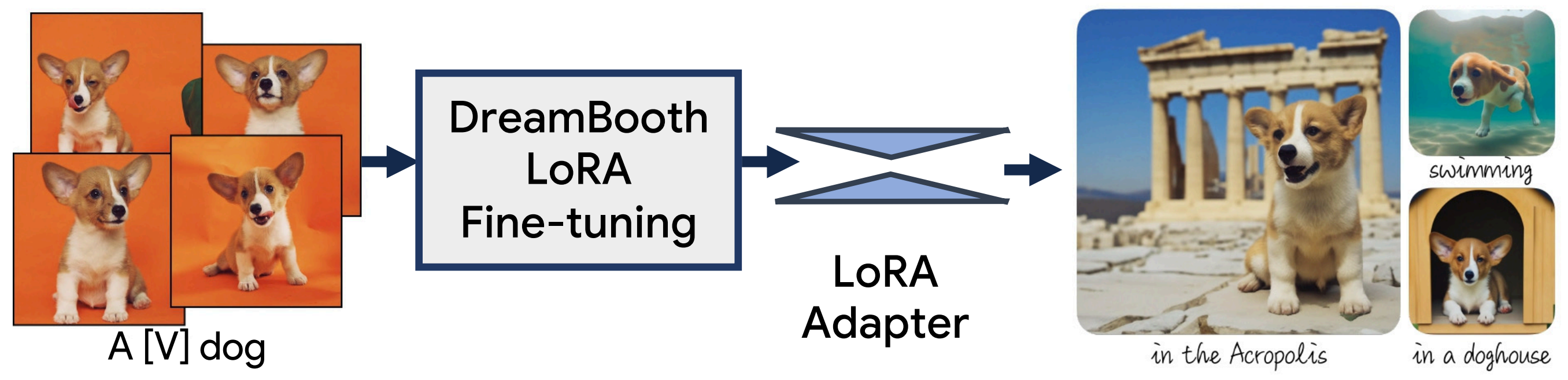
ZipLoRA: Any Subject in Any Style by Effectively Merging LoRAs

Viraj Shah[1,2], Nataniel Ruiz[1], Forrester Cole[1], Erika Lu[1], Svetlana Lazebnik[2], Yuanzhen Li[1], Varun Jampani[1]

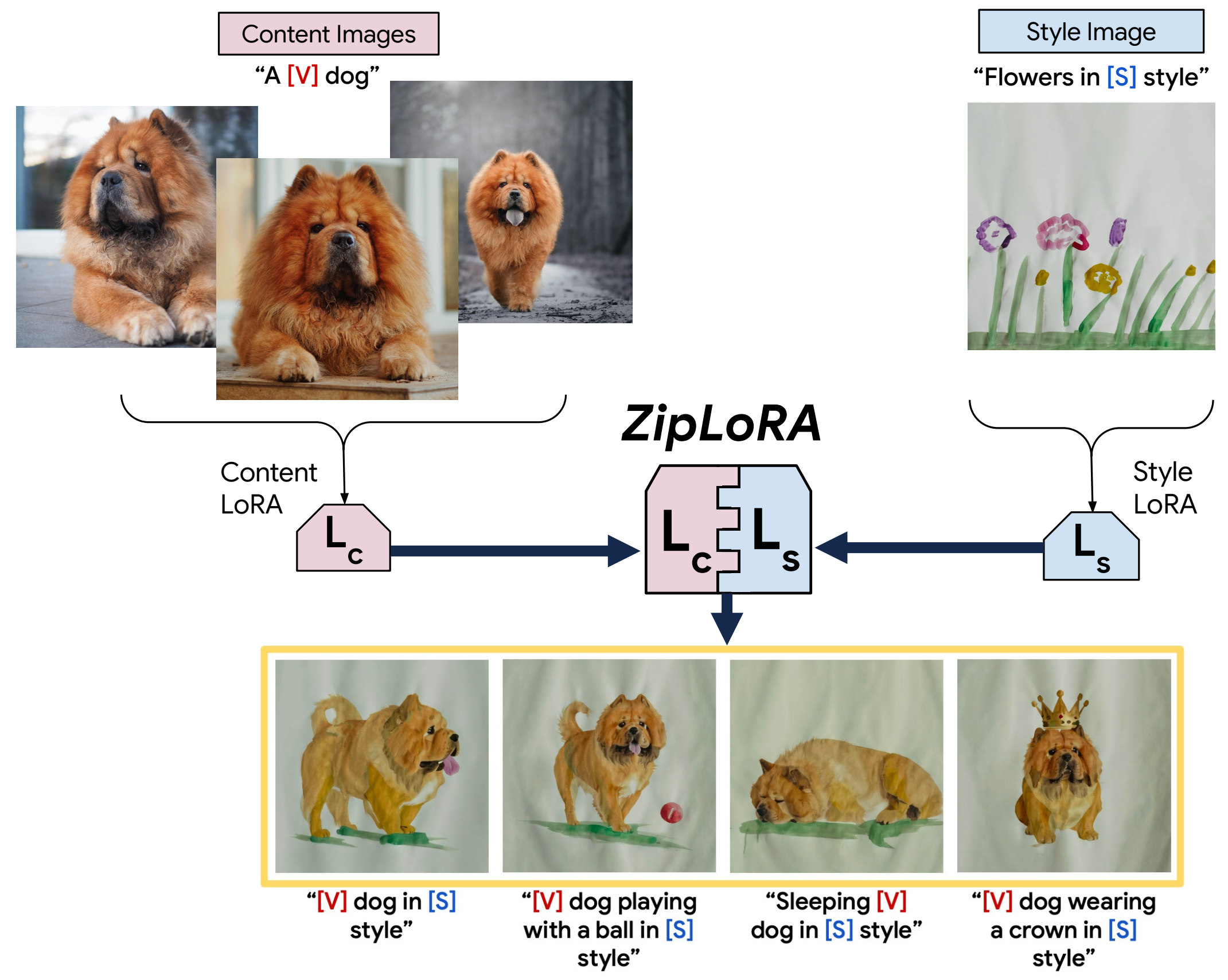
[1] Google Research, [2] UIUC {virajshah, natanielruiz}@google.com

Introduction

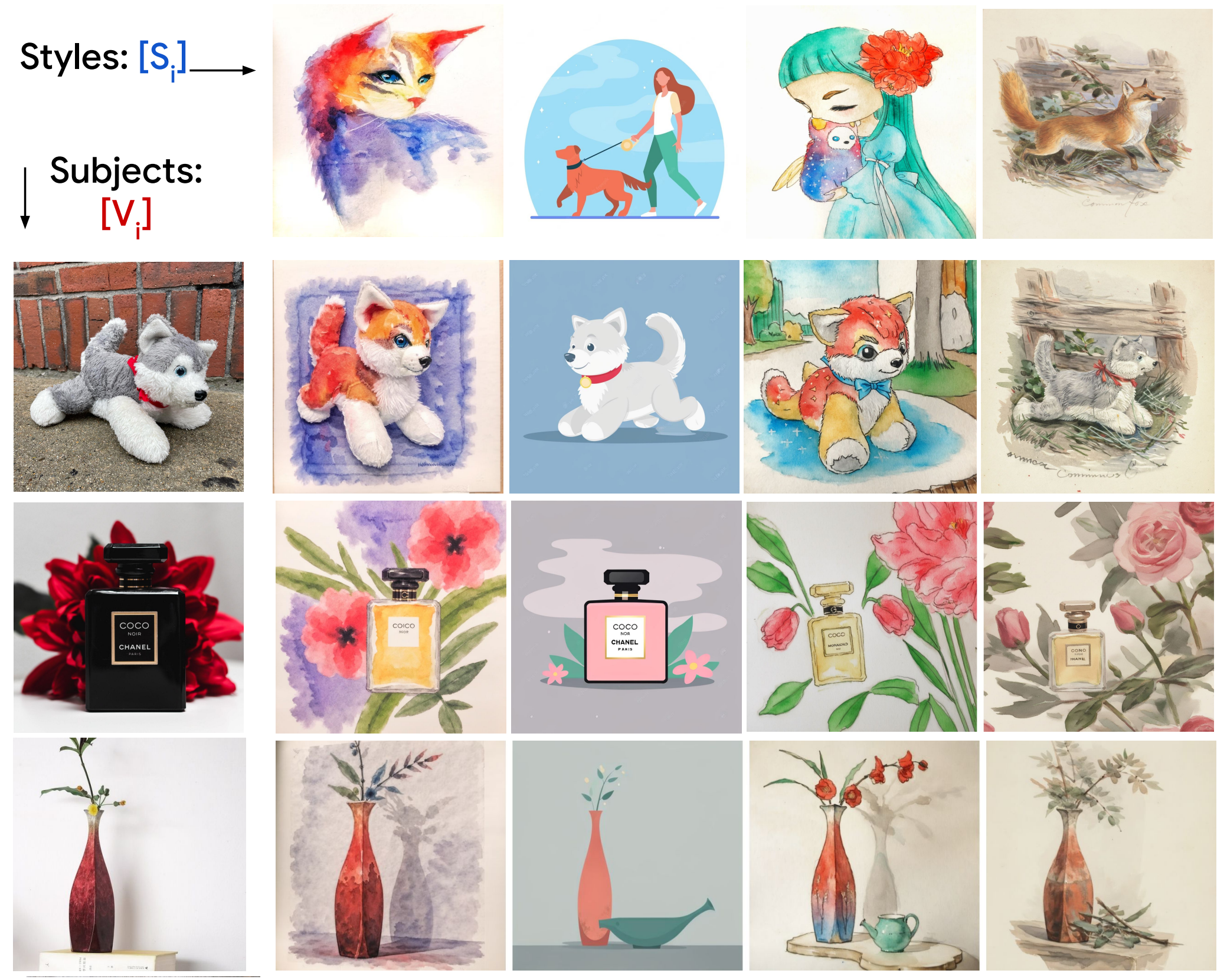
- Personalization methods like DreamBooth fine-tune diffusion models to obtain novel renditions of specific concepts, such as objects, or artistic styles.
- Preferred way for efficient fine-tuning is to use Low Rank Adaptation (LoRA).



- While personalization methods work for subjects and styles independently, a key unsolved problem is to generate a specific user-provided *subject* in a specific user-provided *style*.

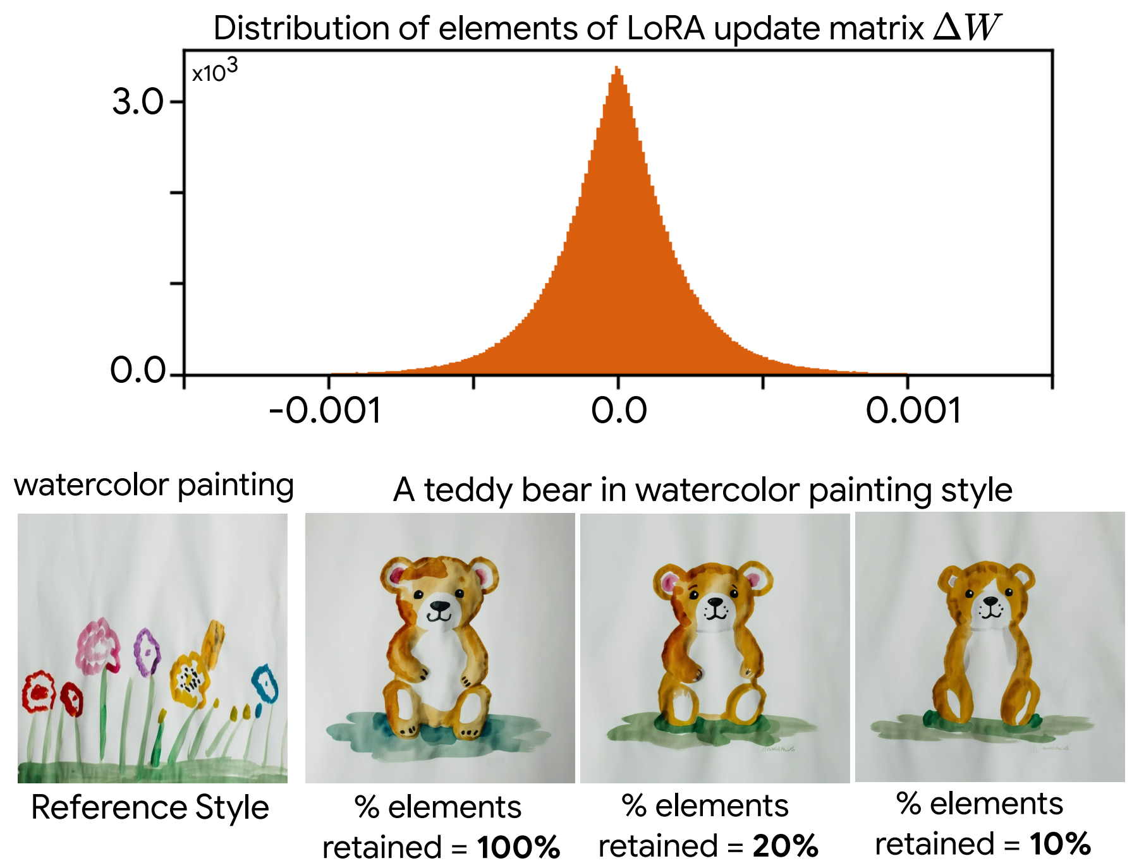


- **ZipLoRA** can merge independently trained subject and style LoRAs, enabling consistent, customizable stylization and recontextualization for a variety of combinations.



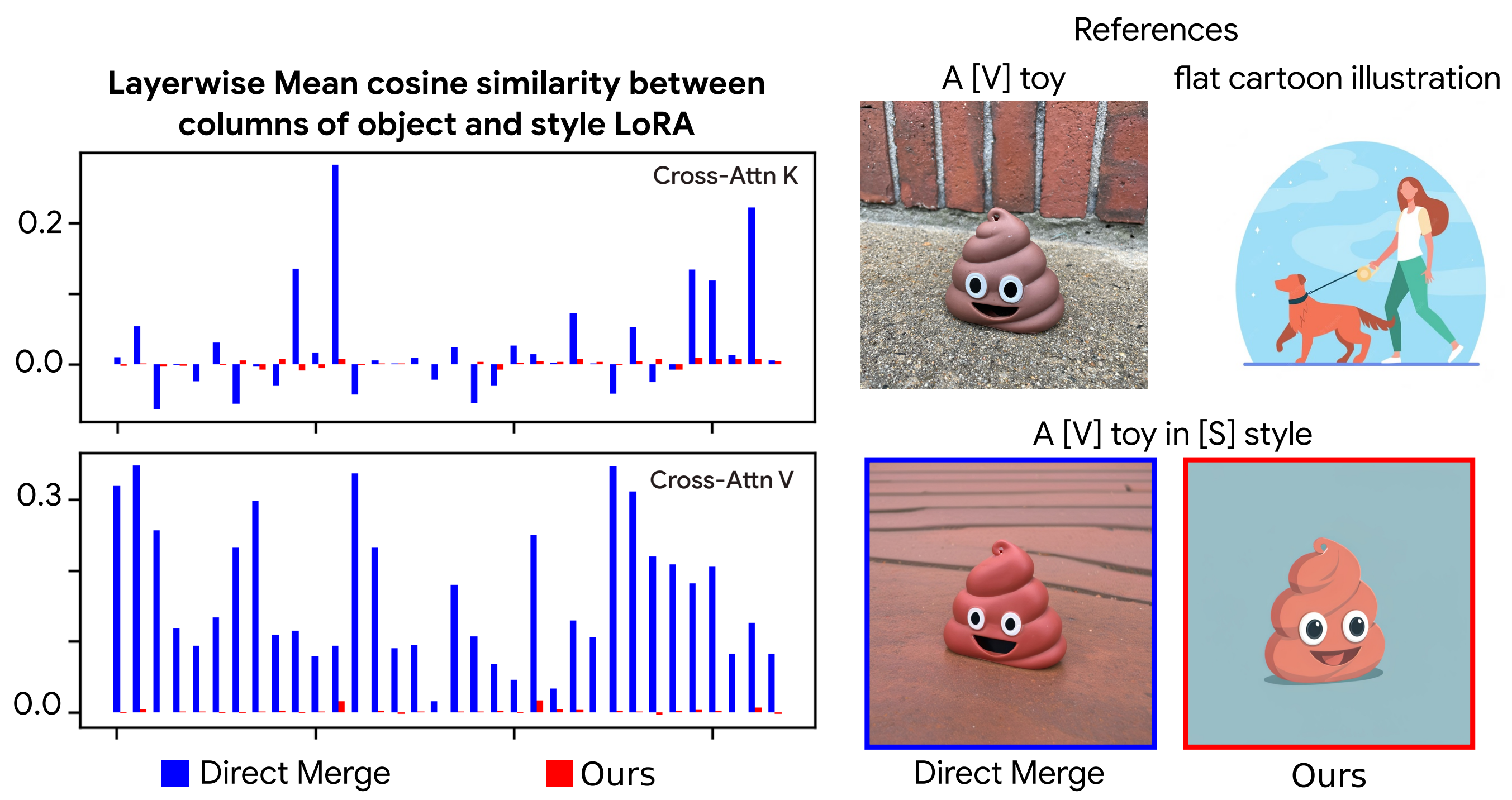
"[V$_j$] subject in [S$_i$] style"

Key Insights



1. LoRA weight matrices are sparse

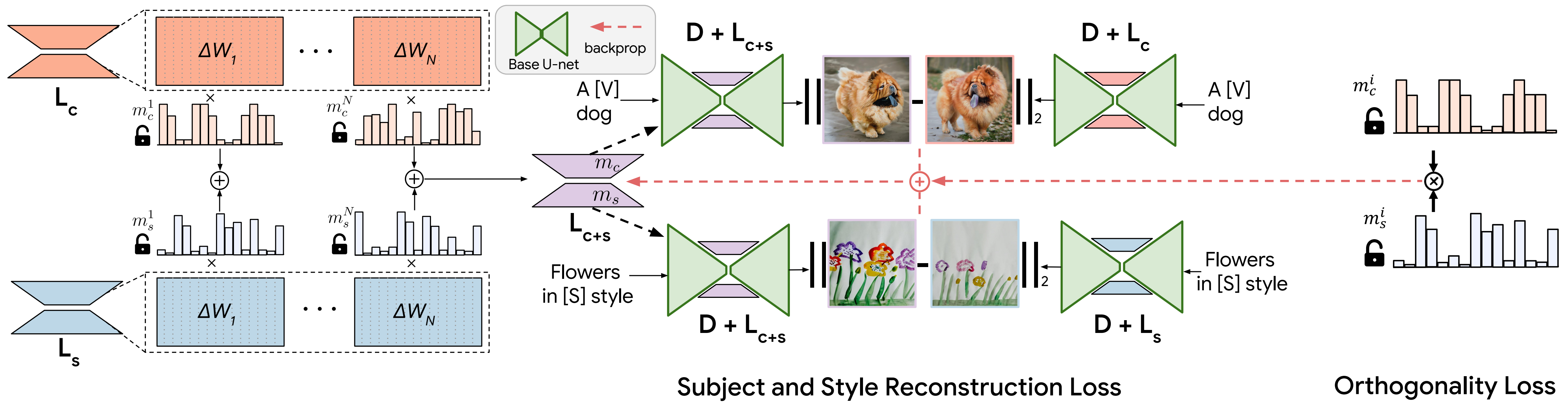
- Most elements in the LoRA weight matrix have very small magnitude and have little effect on generation quality and fidelity.



2. Highly aligned LoRA weights merge poorly

- Directly summing columns of weight matrices that are *highly aligned* degrades performance of the merged model.

Method



Subject and Style Reconstruction Loss

Orthogonality Loss

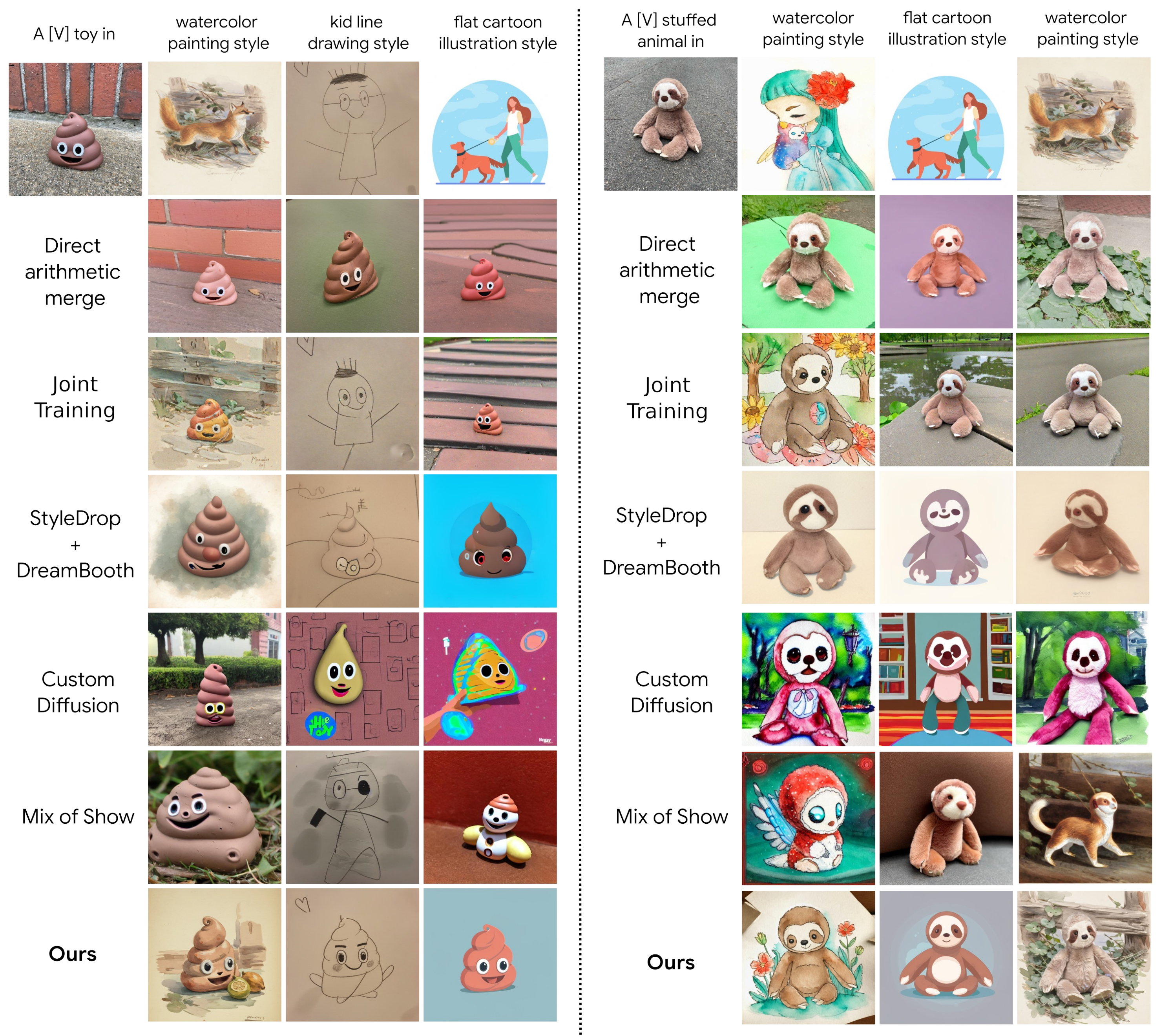
- ZipLoRA keeps the original LoRA weights frozen and multiply each column with a learnable merger coefficient m_c^i / m_s^i .

Orthogonality Loss minimizes the interference between two LoRAs by making m_c and m_s *disjoint*

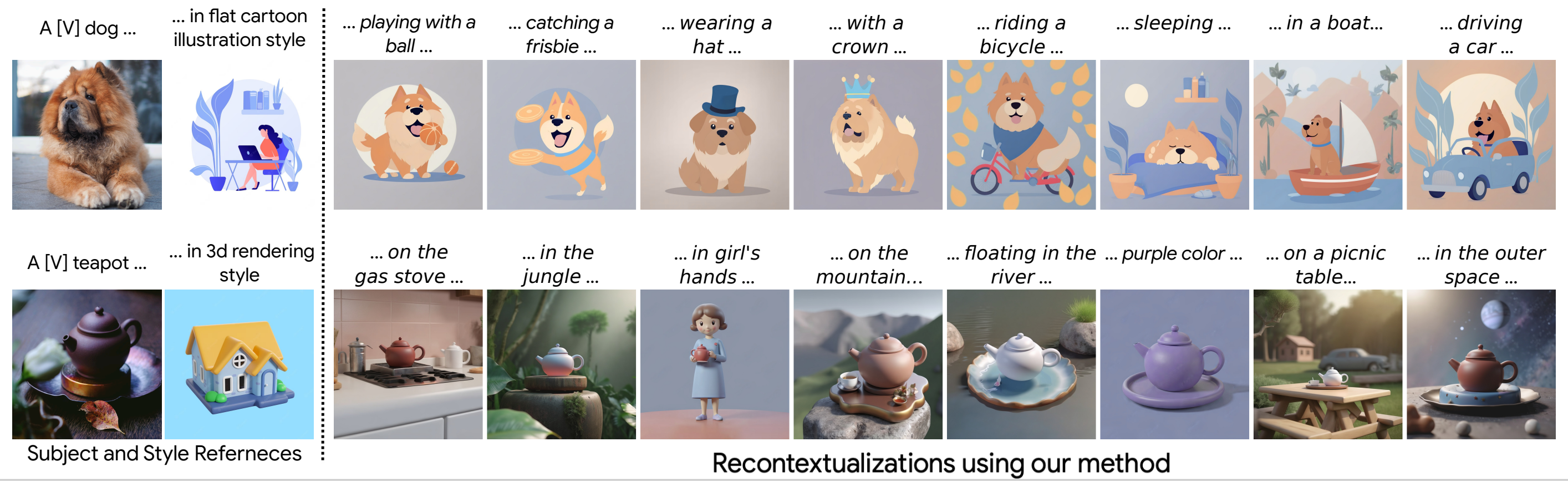
Reconstruction Loss conserves the ability to generate the subject and style independently

Results

Stylization: Qualitative Comparison



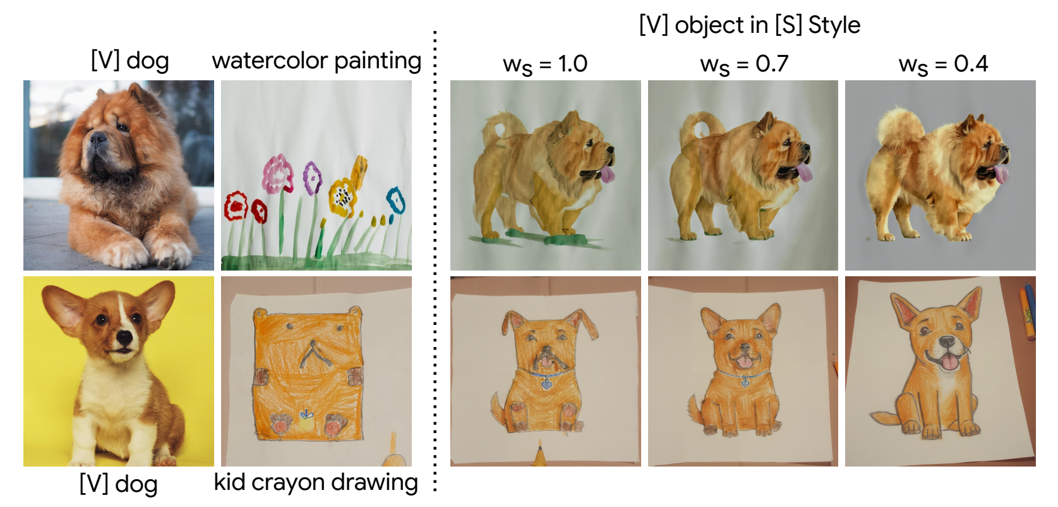
Recontextualization



Subject and Style Referneces

Recontextualizations using our method

Style Controllability



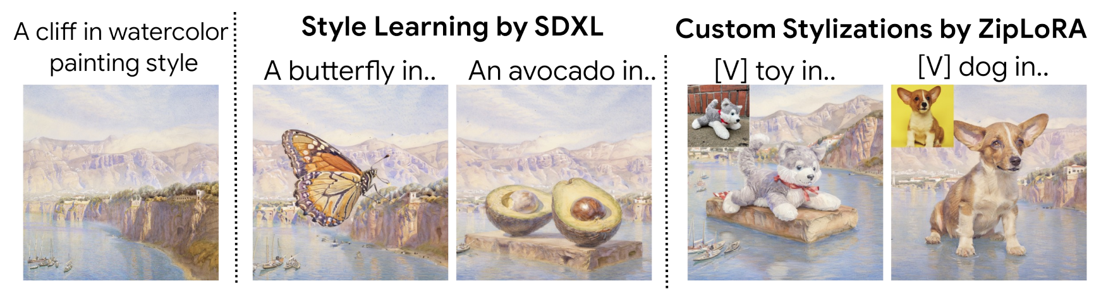
ZipLoRA allows for controlling the extent of stylization

User Study Results

% preference for ZipLoRA over:

Direct Merge	Joint Training	StyleDrop	Mix of Show	Custom Diffusion
82.7%	71.1%	68.0%	87.3%	88.1%

Limitations



For a few styles, content leaks in stylization outputs

Experimental Setup:

- SDXL is used as the base model.
- m_c^i / m_s^i are initialized as all *1*s.
- ZipLoRA typically takes <100 iterations to converge.

More results available at:

https://ziplora.github.io