

Problem Definition

- Generating sequential visual narratives requires maintaining persistent character identity across multiple frames.
- Standard text-to-image models struggle with two key challenges simultaneously:

1. Personalized Subject Representation:

Accurately rendering a specific user-provided character in every frame.

2. Cross-Image Consistency:

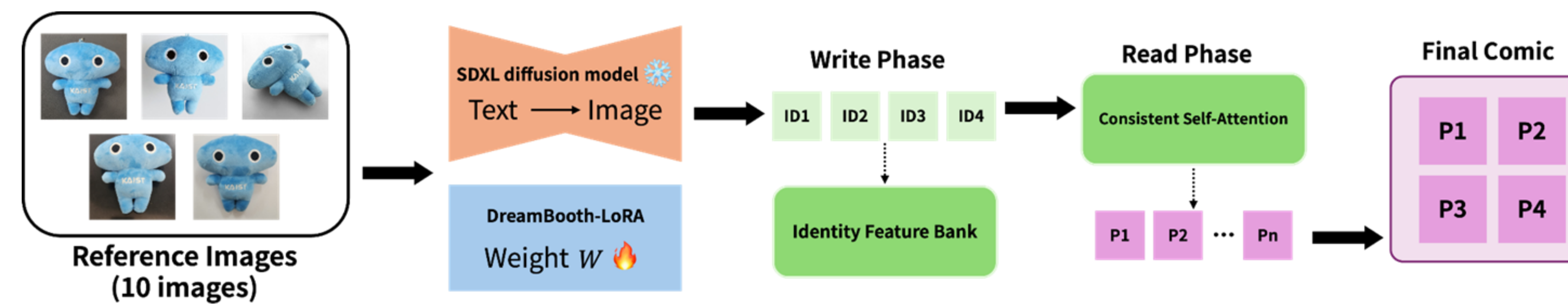
Preserving visual identity and context coherence throughout panels without drifting.



Key Ideas

- Unified Framework (StoryBooth):** Integrates subject-driven fine-tuning (DreamBooth-LoRA^[1]) with training-free attention mechanisms (StoryDiffusion^[2]) for consistent comic generation.
- Log-Normal Timestep Sampling:** Improves personalization during DreamBooth training by biasing sampling toward semantically rich timesteps, preventing overfitting and capturing fine details.

Method



1. Personalization Stage (Left Block):

- Fine-tune SDXL using **DreamBooth with LoRA** on few reference images (e.g., 3-5) to learn the specific subject token.
- Apply **Log-Normal Timestep Sampling** during training to enhance subject fidelity.

2. Write Phase (Middle Block):

- Generate initial "identity reference images" using the personalized LoRA model.
- Extract and store their self-attention features into a shared **Identity Feature Bank**.

3. Read Phase (Right Block):

- Generate subsequent comic panels.
- The model attends to both its current features and the stored features in the Identity Bank via **Consistent Self-Attention**, ensuring visual continuity.

References

- [1] N. Ruiz et al., "DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation", CVPR 2023.
[2] Y. Zhou et al., "StoryDiffusion: Consistent Self-Attention for Long-Range Image and Video Generation", NeurIPS 2024.

Acknowledgements

Jaeyoung, Jaewoo: Result analysis, generation, Sangoh: Code implementation

Results

- Consistent Character Identity:** The specific subject (blue "Nupjuki" character) is preserved across panels.
- Narrative Continuity:** The generated sequence successfully tells a visual story while maintaining stylistic coherence across multiple panels.

