

Stylized Outputs from Our Method

Consistent layout, prompt-aligned, cross-style robustness.

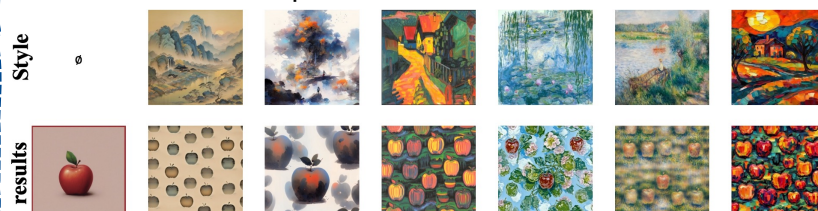


Challenges

- Style Overfitting:** ignoring text prompt variations.



- Visual Artifacts:** Spatial structure breaks down

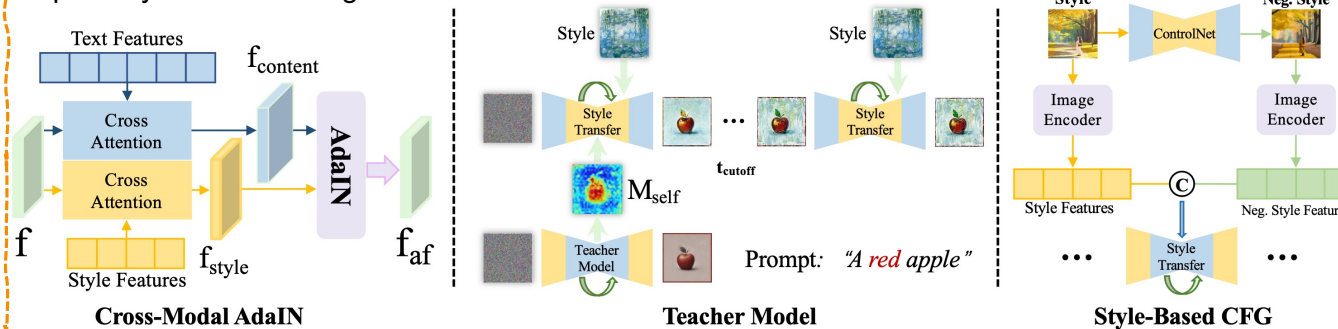


Why Do These Challenges Arise?

- The limitations of current fusion mechanisms (e.g., **additive attention**) and the absence of structured style encoders contribute to style overfitting.
- Visual artifacts, while observable via attention map breakdowns, remain a systemic challenge without clear causality.

Methods (Training-Free, Plug-and-Play Design)

We propose a training-free, plug-and-play framework that fuses style and text via AdaIN, stabilizes layout through early-stage self-attention transfer, and enables selective style control with a negative reference inspired by classifier-free guidance.



Qualitative Results of the Ablation Study



Quality Comparisons



Experiments

- Our method handles complex prompts with diverse objects and preserves both structure and style.



- Style-Based CFG selectively removes undesired attributes.



- Replacing Self-Attention preserves layout better than Cross-Attention in the teacher model design.



"A motorcycle"



- Integration with InstantStyle

