TAUE: Training-free Noise Transplant and Cultivation Diffusion Model

Daichi Nagai*[1], Ryugo Morita*[2], Shunsuke Kitada[1], Hitoshi Iyatomi[1]
[1]Hosei University, Tokyo, Japan [2]RPTU Kaiserslautern-Landau & DFKI GmbH, Kaiserslautern, Germany
*Equal contribution

CVPR JUNE 3-7, 2026 DENVER COLORADO

Background

Why?

Existing T2I models produce non-layered images
- 😭 Difficult post-editing & object replacement
- 😭 Limited flexibility in practical creative workflows

However ...

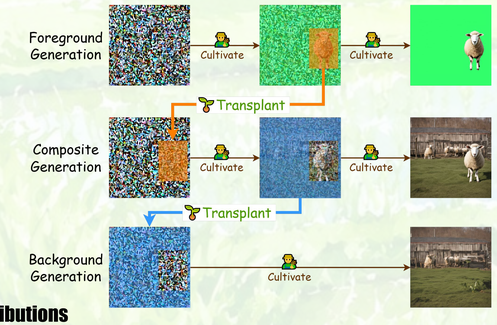
Existing methods w/ fine-tuning [Zhang+ ACM Trans. Graph.'24]
- Requires layer-annotated datasets
- Incurs substantial computational cost

Existing approaches w/o fine-tuning [Quattrini+ ECCVW'24]
- Limited to generating foreground layers only

Our Insight

TAUE achieves training-free layer generation
- ✨ High consistency of semantics and structure
- 💡 Key idea: transplanting **seedling noise** across layers

Seedling noise is dug up and transplanted 🌱
↳ An intermediate latent representation containing semantic and structural information

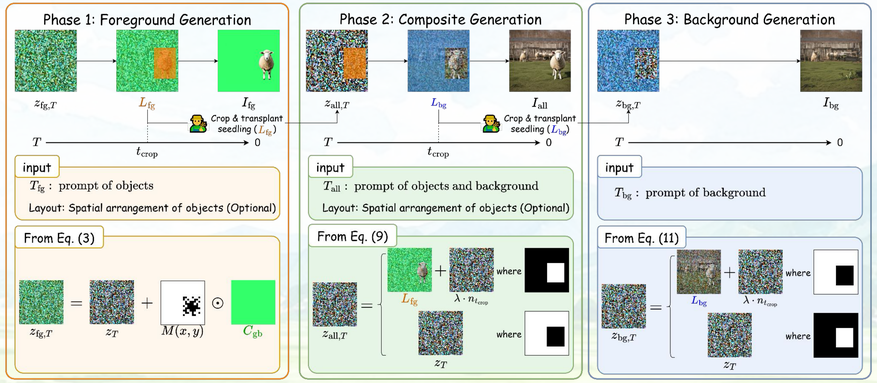


Contributions

We introduce the training-free layer generation framework
- **Enables interactive image editing**
- **Facilitates practical creative workflows**

†**TAUE** is inspired by the Japanese rice-planting process ("Taue"), where seedlings are transplanted for growth. Similarly, our method achieves consistent generation by transplanting intermediate latents across denoising trajectories.

Pipeline of TAUE



Phase 1: Foreground Generation

1. Generate a foreground image following TKG-DM [Morita+ CVPR'25]
2. Extract the foreground seedling noise L_{fg} from the mid stage of the denoising process

Phase 2: Composite Generation

1. Generate a composite image using initial noise transplanted with L_{fg}
2. Through NTC, the transplanted L_{fg} guides the denoising process
3. Extract the background seedling noise L_{bg} from the mid stage of the denoising process

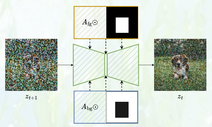
Phase 3: Background Generation

1. Generate a background image using initial noise transplanted with L_{bg}

NTC: Noise Transplant & Cultivation

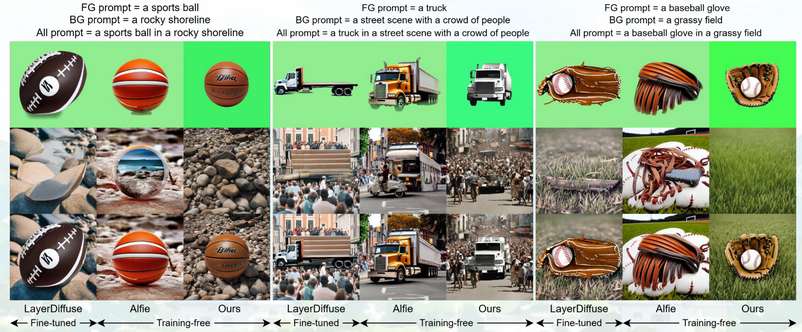
- Enhance high-frequency components of L_{fg}
- The pre-scheduled noise in the L_{fg} region is progressively removed throughout denoising
- Foreground and background attn. are separately applied to their corresponding regions
  - Specifically, for multi-object generation, object-specific attn. prevents object swapping and unintended merging

➡ Enhances foreground generation



➡ **Obtain the foreground, background, and composite images!**

Experiments & Results



	Overall Quality			Layer-wise Reconstruction Quality					
	FID ↓	CLIP-I ↑	CLIP-S ↑	$PSNR_{fg}$ ↑	$PSNR_{bg}$ ↑	$SSIM_{fg}$ ↑	$SSIM_{bg}$ ↑	$LPIPS_{fg}$ ↓	$LPIPS_{bg}$ ↓
Fine-tuning Methods									
LayerDiffuse	61.46	**0.653**	0.312	14.78	**32.76**	0.828	**0.957**	0.323	0.039
Training-free Methods									
Alfie + inpainting	85.93	0.644	0.302	15.32	27.45	0.778	0.947	0.254	**0.019**
Ours	**60.53**	0.646	**0.323**	**20.46**	25.86	**0.901**	0.895	**0.137**	0.106

Qualitative:
Consistent layer generation with fewer artifacts

Quantitative:
Competitive performance across multiple metrics

Application