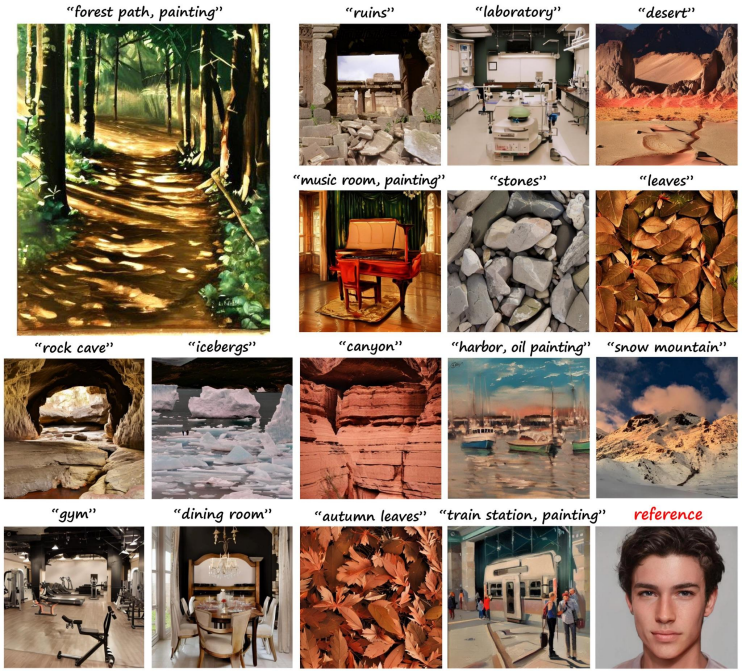
PTDiffusion: Free Lunch for Generating Optical Illusion Hidden Pictures with Phase-Transferred Diffusion Model

Xiang Gao, Shuai Yang, Jiaying Liu
Wangxuan Institute of Computer Technology, Peking University

CVPR 2025

Background

- Optical Illusion Hidden Picture



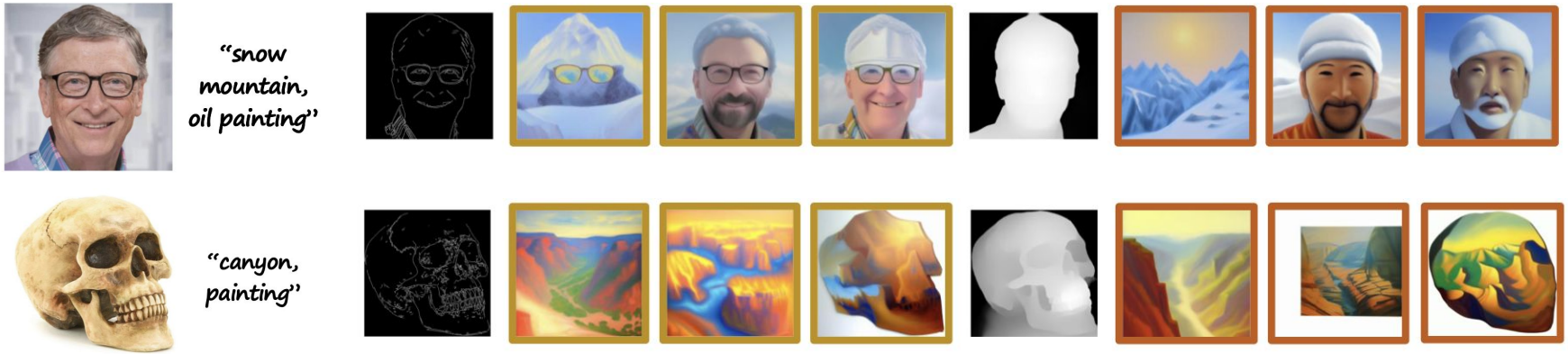
Evolution of Optical Illusion

- Early Computational Illusions
  - Geometric, color, and motion illusions
- Camouflage Generation
  - Re-texturing & Style Transfer
- Diffusion-Based Approaches
  - QR Codes: DiffQRCoder
  - Overlay & Multi-view: Diffusion Illusions, Visual Anagrams

Limitations of Existing Methods

Text-guided image generation: ControlNet, T2I-Adapter

- Use extra trained networks for structural control (edges, depth)
- Structure Over-constraint

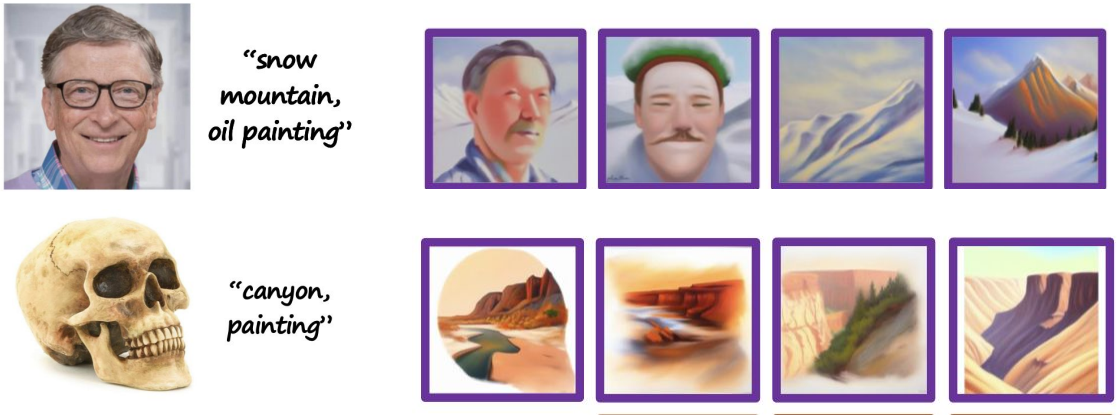


ControlNet-canny (input, 0.3, 0.4, 0.5 control weight) ControlNet-depth (input, 0.3, 0.4, 0.5 control weight)

Limitations of Existing Methods

Image-to-image(I2I) translation: SDEdit, Attention-based methods

- SDEdit translates a reference image by noising it to an intermediate step followed by text-guided denoising
- Over-binding / Structure-Semantic Conflict



SDEdit (0.7, 0.75, 0.8, 0.85 denoising strength)

Goal

- Pioneer generating optical illusion hidden pictures from the perspective of **text-guided Image-to-Image (I2I) translation**
- Translate an input reference image into an illusion picture
  - Semantic: Follow the text prompt description
  - Structural: Clearly manifests clues of the reference image
- **Three key differences from existing methods**
  - Visual Discernibility
  - Generative Backgrounds
  - Seamless Dissolution

Key Insight from Signal Processing

- **Phase**: Structure and contours
- **Amplitude (Magnitude)**: Style and semantic strength

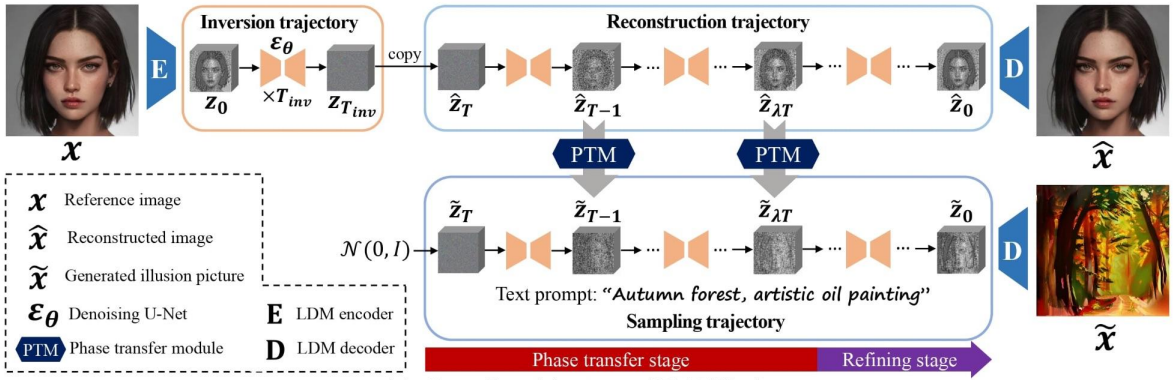
Phase

Magnitude

Phase-Transferred Diffusion Model

Three Diffusion Trajectories

1. Inversion Trajectory
2. Reconstruction Trajectory
3. Sampling Trajectory

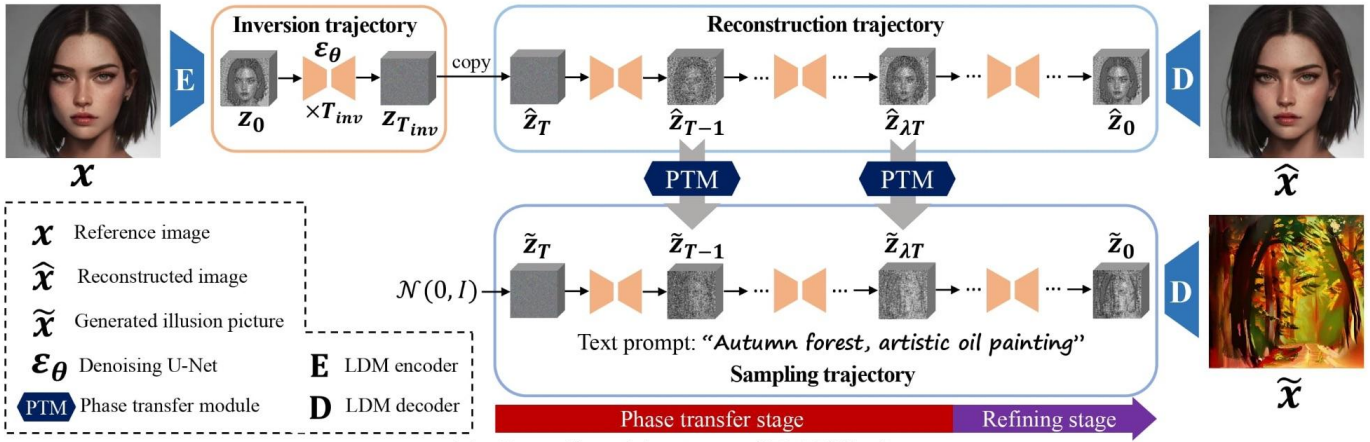


(a). Overall architecture of PTDiffusion

Phase-Transferred Diffusion Model

1. Inversion Trajectory

- Inverts reference image X to noise z_{Tinv} using DDIM Inversion

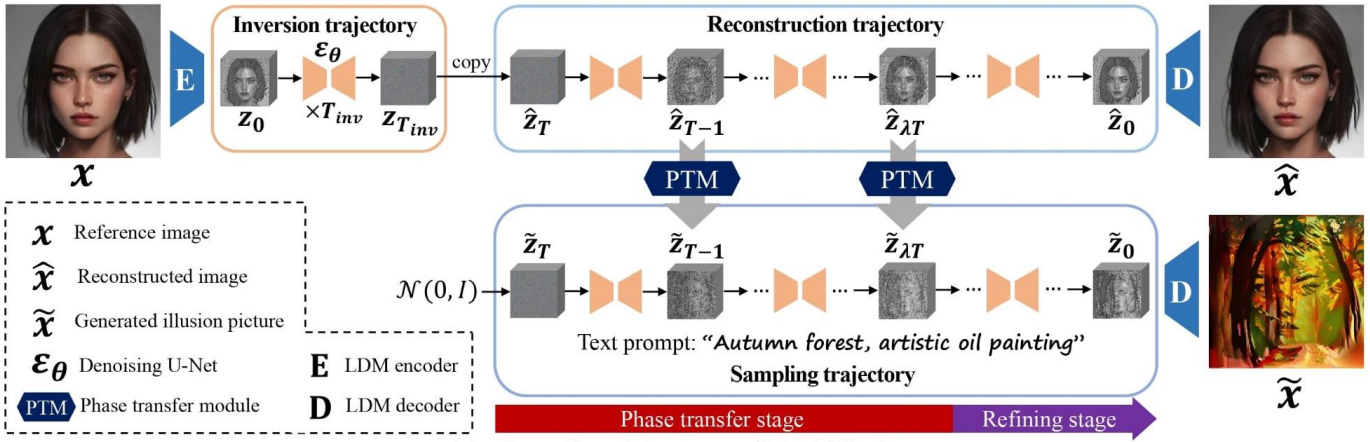


(a). Overall architecture of PTDiffusion

Phase-Transferred Diffusion Model

2. Reconstruction Trajectory

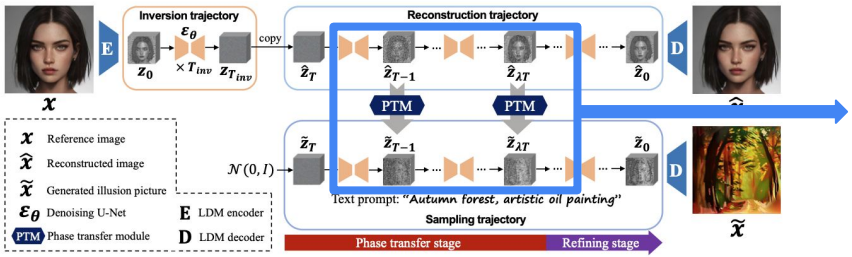
- Reconstructs X to obtain structural guidance features $\hat{z}_T$

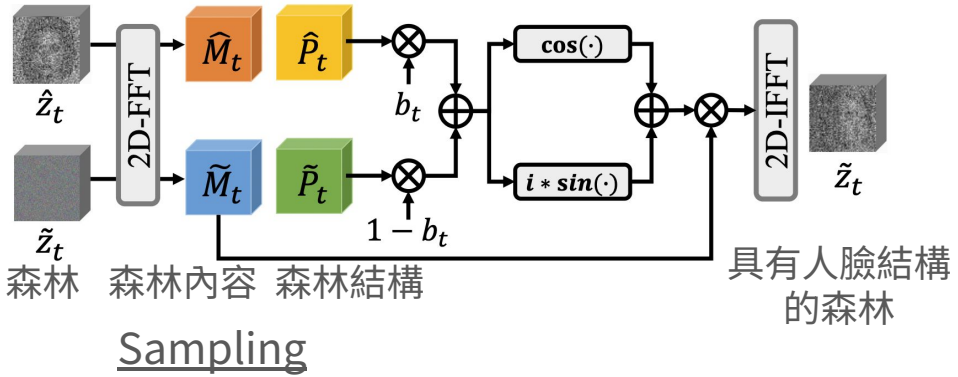


(a). Overall architecture of PTDiffusion

Phase-Transferred Diffusion Model

- **Phase Transfer Module (PTM)**

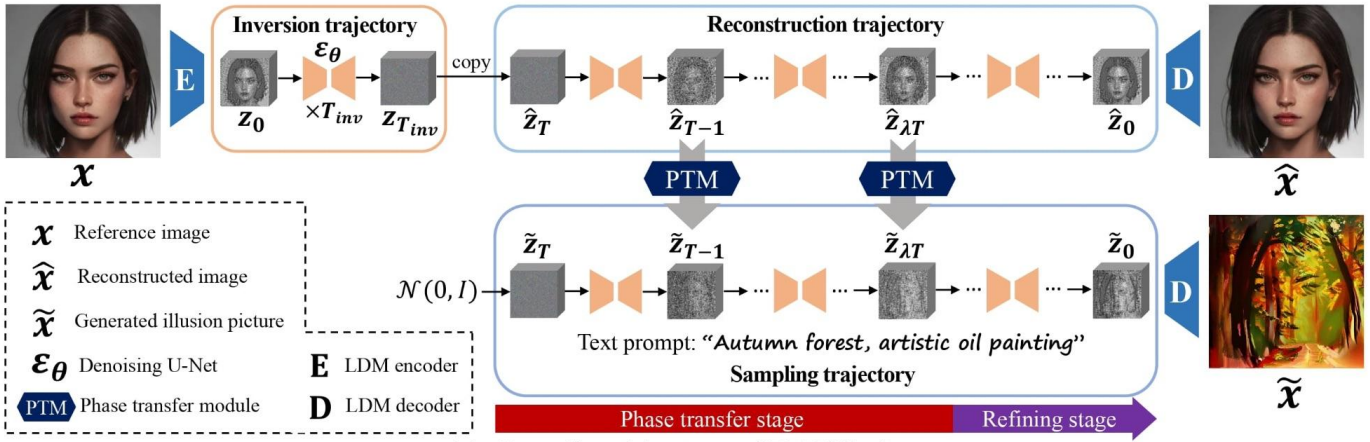




Phase-Transferred Diffusion Model

3. Sampling Trajectory

- Samples the final illusion image x~ from initialized random noise z~T guided by the target text prompt using DDIM sampling

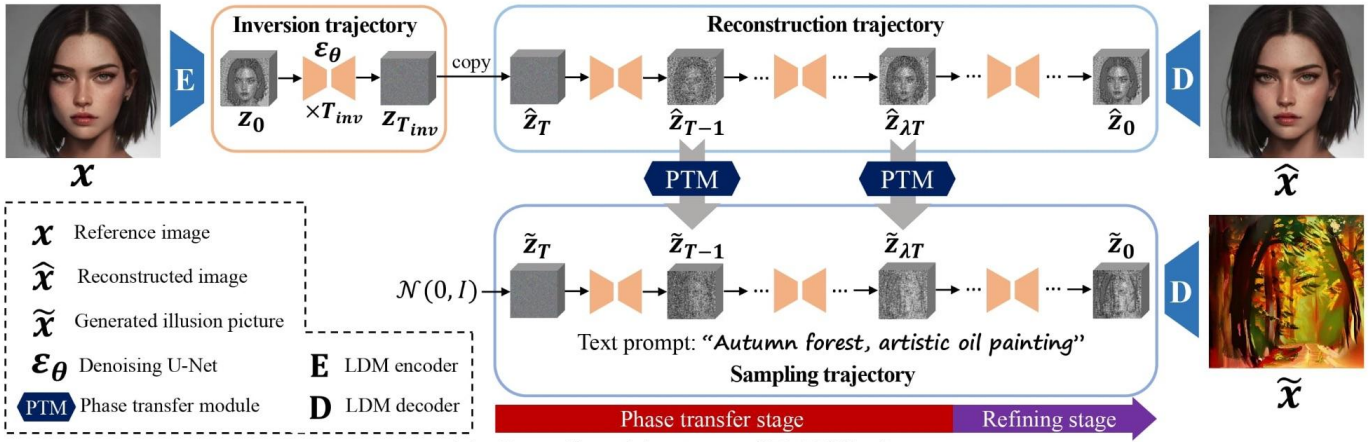


(a). Overall architecture of PTDiffusion

Phase-Transferred Diffusion Model

3. Sampling Trajectory

(1) Phase transfer stage
(2) Refining stage



(a). Overall architecture of PTDiffusion

Phase-Transferred Diffusion Model

- **Asynchronous Phase Transfer**

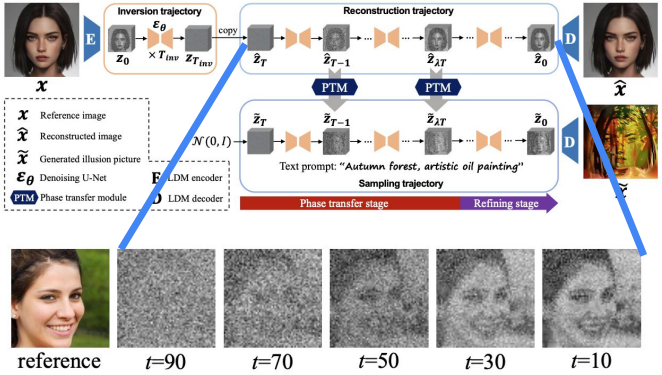


Figure 4. Visualization of the guidance features $\{\hat{z}_t\}$ along the 100-step reconstruction trajectory. The structural information of $\hat{z}_t$ becomes increasingly distinct as the denoising proceeds.

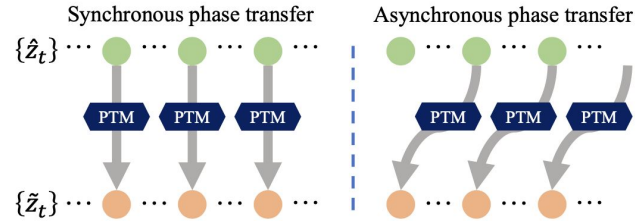


Figure 5. Illustration of the asynchronous phase transfer which transfers phase across diffusion features at different time steps.

Key Advantages of PTDiffusion

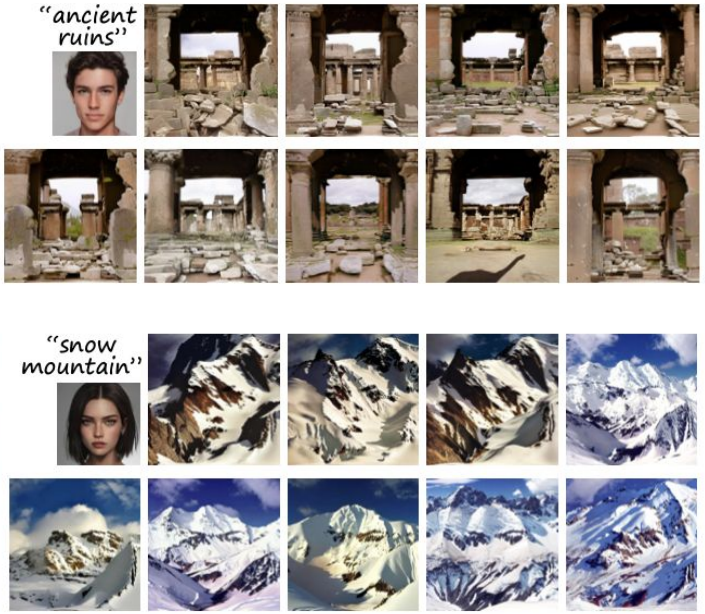
- Zero Training Required
- Optimization-Free Efficiency
  - No Fine-tuning
  - No Online Optimization
- Superior Performance
  - Better than existing I2I methods

Experimental Setup

- Model: Stable Diffusion v1.5.
- Hardware: Single NVIDIA GeForce RTX 3090 Ti GPU.
- Parameters:
  - DDIM Inversion: 1000 steps.
  - Sampling: 100 steps.
  - Guidance Scale: w=7.5.
  - Refining Stage: λ=0.4 (No phase transfer performed during the last 40% of steps).

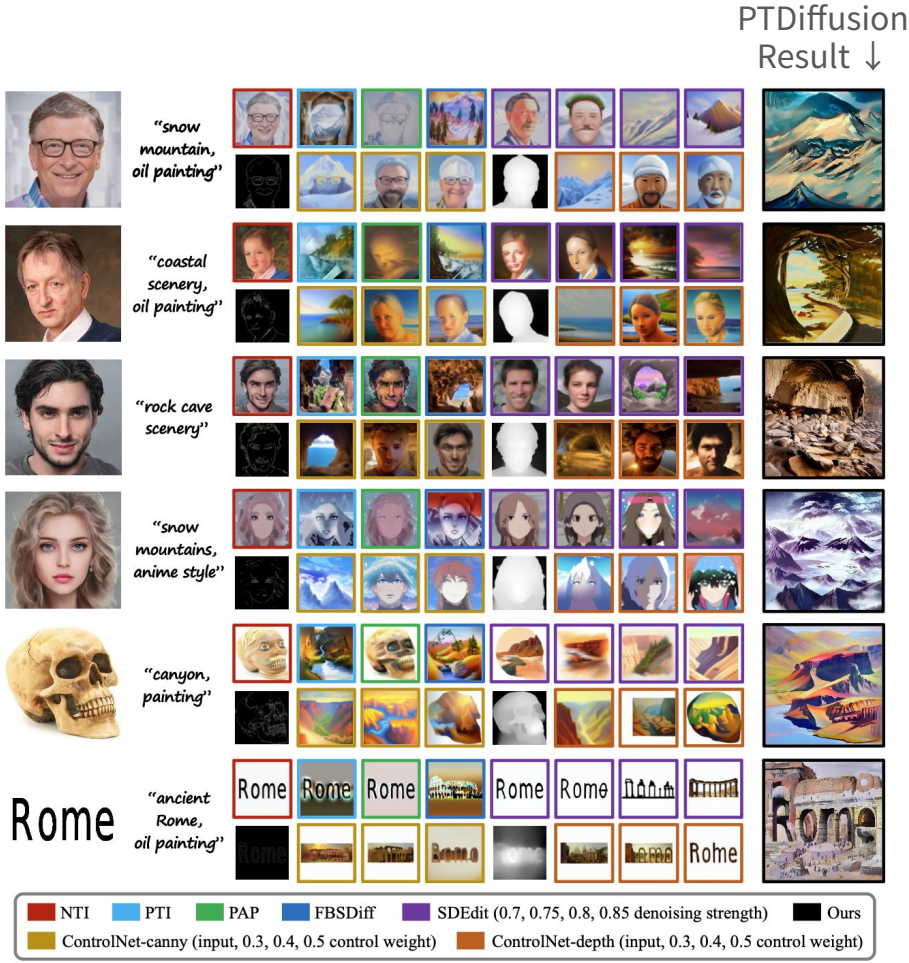
Results


reference
"island"
"Grand Canyon"
"icebergs"
"rock cave"
reference
"snow mountain"
"mountain stream, water color"
"London Tower Bridge, oil painting"
"Venice, oil painting"
reference
"park, oil painting"
"wave, oil painting"
"farmland, oil painting"
"Japanese alley, oil painting"


"ancient ruins"
"snow mountain"

Comparison with SOTA

- **Text-guided I2I**
  - Null-text Inversion (NTI)
  - PAP
  - Prompt-tuning Inversion (PTI)
  - FBSDiff
  - SDEdit
- **Controllable T2I**
  - ControlNet (w/ canny edge, depth map)



Ablation study

Quantitative evaluation

- Metrics
  - Aesthetic Score: Image quality
  - CLIP Score: Text consistency
  - LPIPS: Deviation from source image

Method	Aesthetic Score (↑)	CLIP Score (↑)	LPIPS (↑)
NTI [31]	6.24	0.23	0.21
PTI [5]	6.18	0.28	0.52
PAP [45]	6.09	0.25	0.40
FBSDiff [10]	5.96	0.29	0.56
SDEdit [30]	6.10	0.29	0.49
ControlNet [48]	6.05	0.26	0.47
PTDiffusion (ours)	**6.37**	**0.31**	**0.64**

- User study
  - Contextual Naturalness
  - Illusion Balance

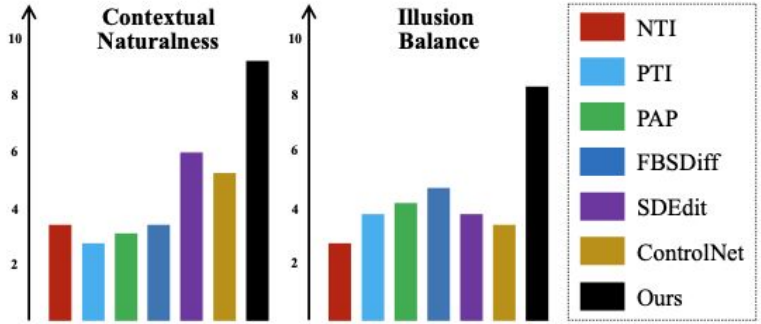


Figure 12. Average user ratings to different methods.

Conclusion

- **Pioneering Work:** The first text-guided I2I approach specifically for optical illusion hidden pictures.
- **Core Technology:** Proposed **PTDiffusion**, solving the structure-semantic over-binding issue via Phase Transfer.
- **Flexible Control:** Asynchronous Transfer enables precise control over the discernibility of hidden content.
- **Free Lunch:** Training-free, Optimization-free, and Plug-and-play

Limitation

- Inference Speed
- Dependence on SD 1.5
- Prompt Sensitivity

End