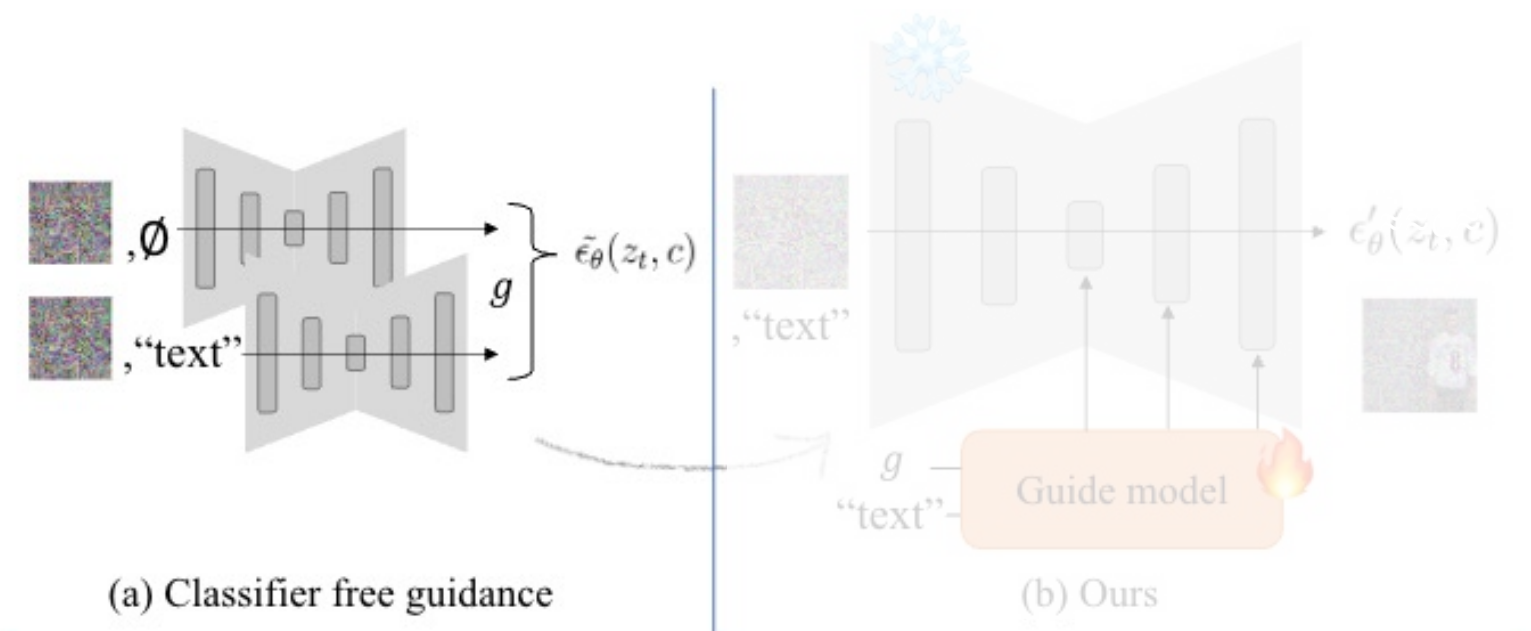# Plug-and-Play Diffusion Distillation

Yi-Ting Hsiao, Siavash Khodadadeh, Kevin Duarte, Wei-An Lin, Hui Qu, Mingi Kwon, Ratheesh Kalarot

The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024
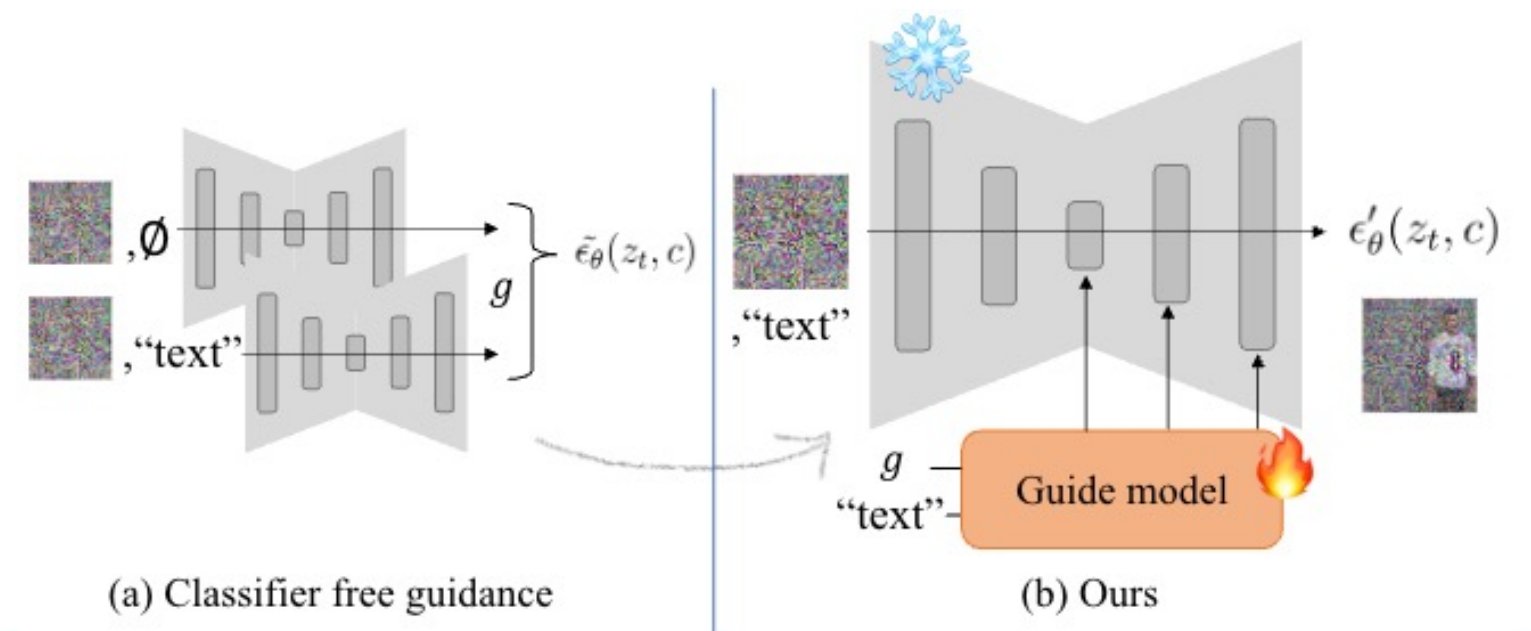
Methodology Overview

(a) Classifier free guidance

(b) Ours

(c) Sampling with various models

Methodology Overview
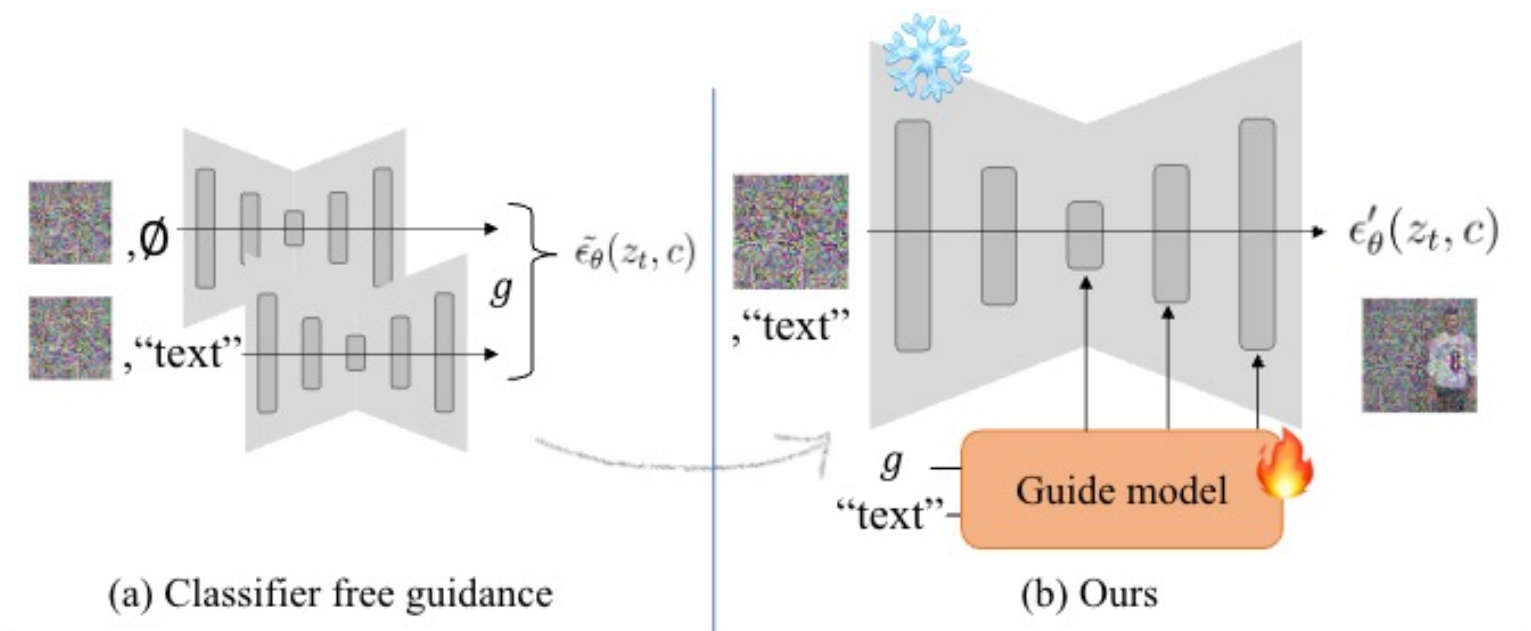

(a) Classifier free guidance

(b) Ours

(c) Sampling with various models

Methodology Overview

(a) Classifier free guidance

(b) Ours

(c) Sampling with various models

Diffusion model Distillation with ~51% FLOPs count decrease

tiny guide model (ours)

classifier-free guidance (teacher)

51% decrease

Method	FLOPs (trillion)	# of Params (million)
Ours-full	T (338.7) + 116.5	T (859) + 361
Ours-tiny	T (338.7) + 7.79	T (859) + 8.27
Teacher × 2	677.5	859

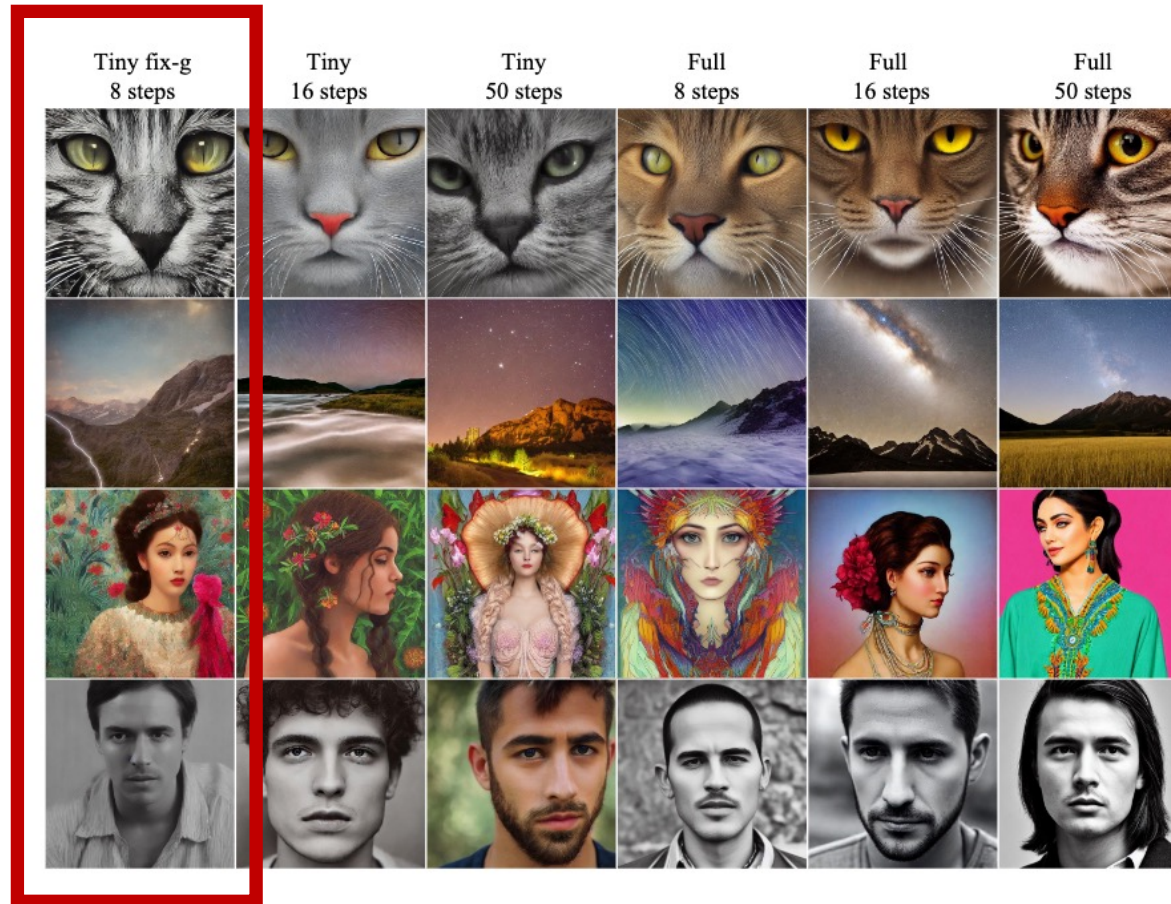# Diffusion model Distillation with ~50% FLOPs count decrease

Figure 5. The results of the full guide model and tiny guide model under 8, 16, and 50 sampling steps.
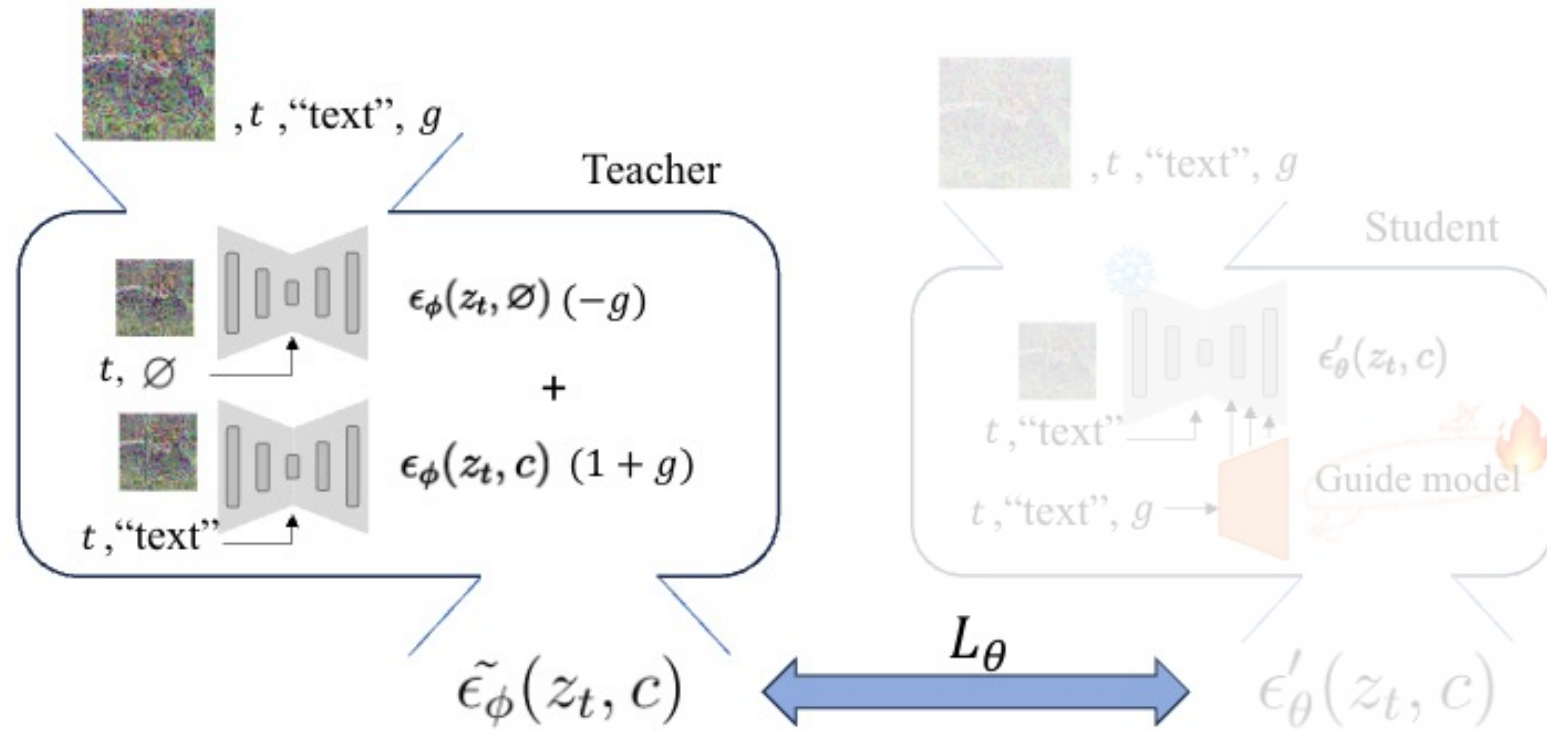
The base model is untouched, so we can distill different fine-tuned model without retraining
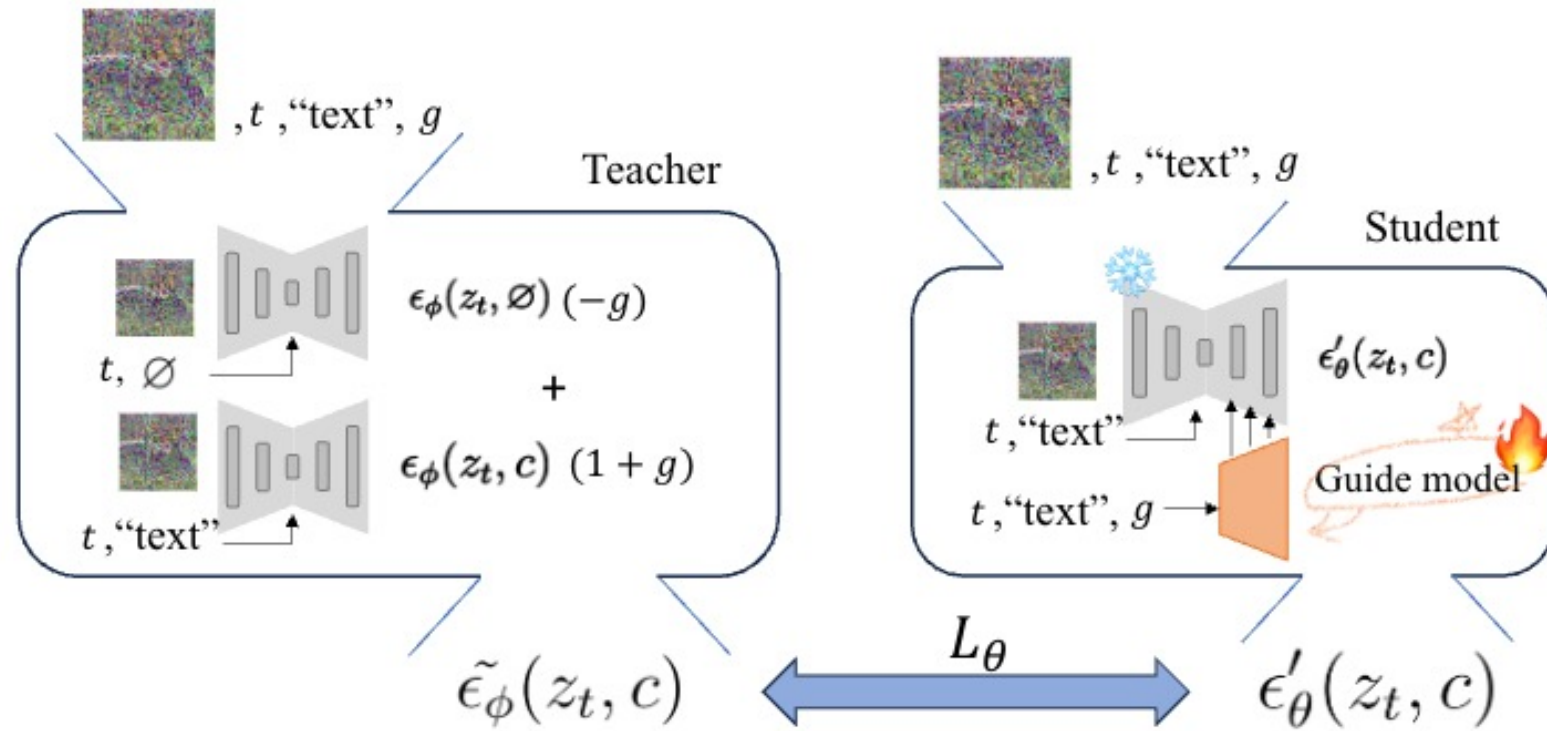

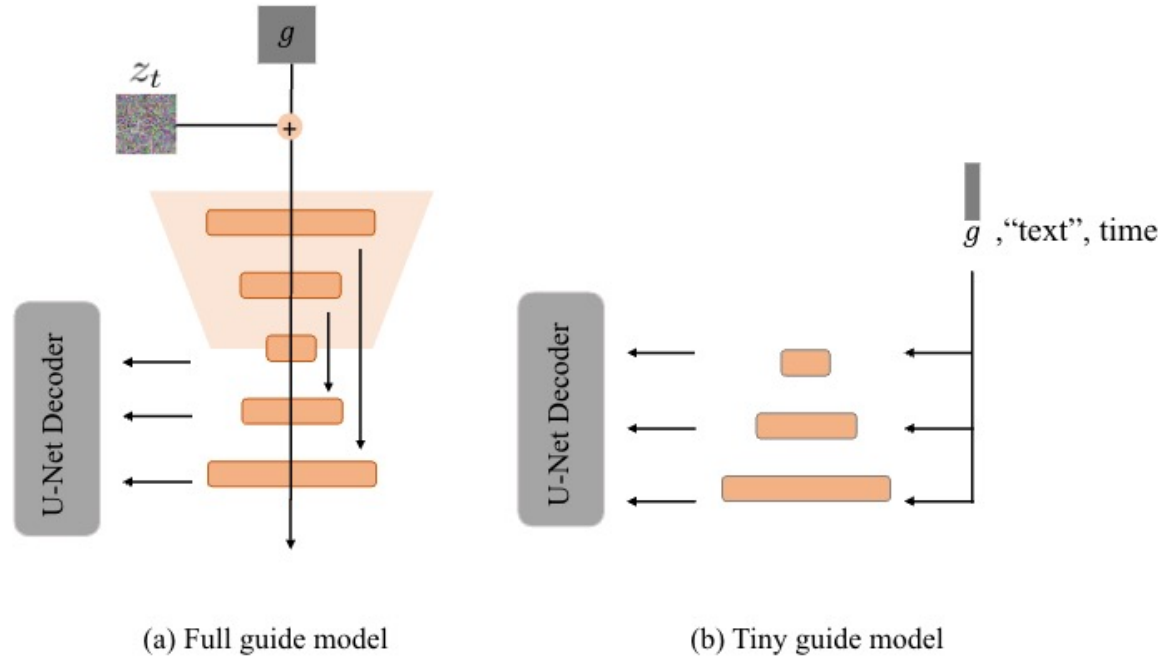Figure 8. Plug in the tiny guide model with different fine-tuned U-Nets.

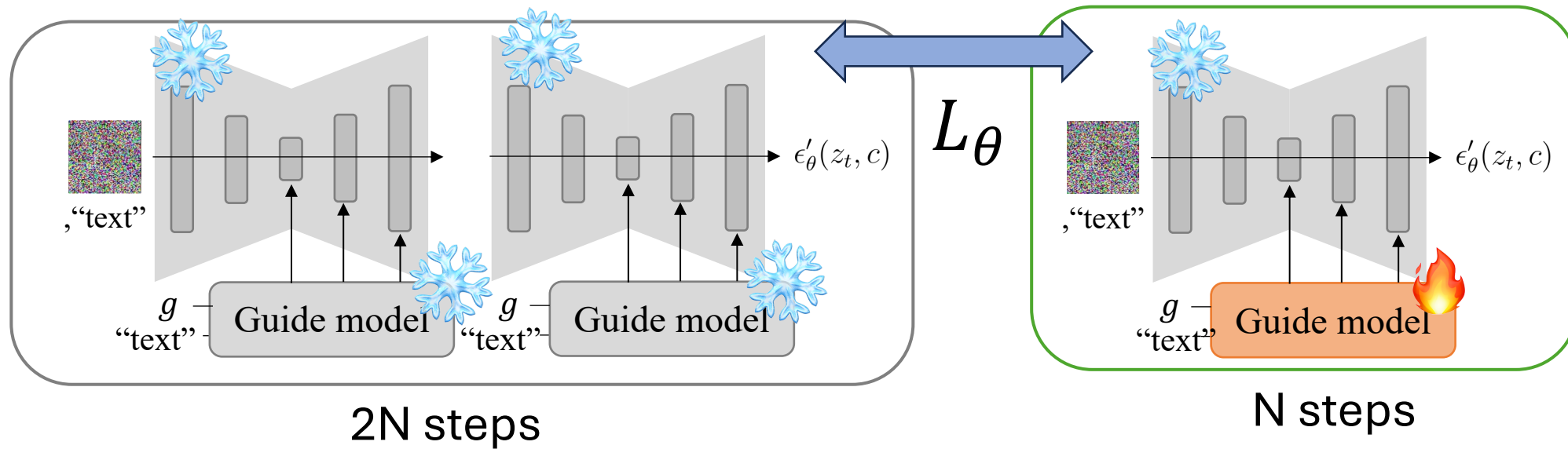# Stage 1 distillation: classifier-free guidance distillation

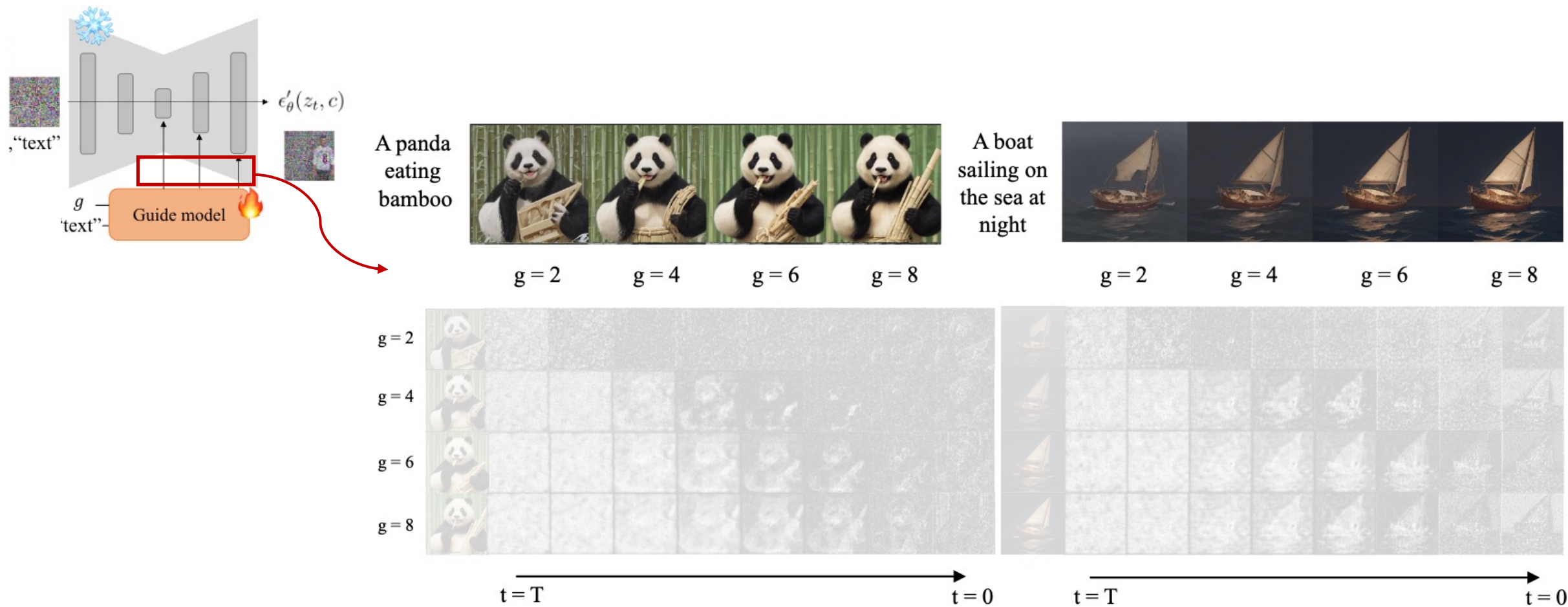# Stage 1 distillation: classifier-free guidance distillation

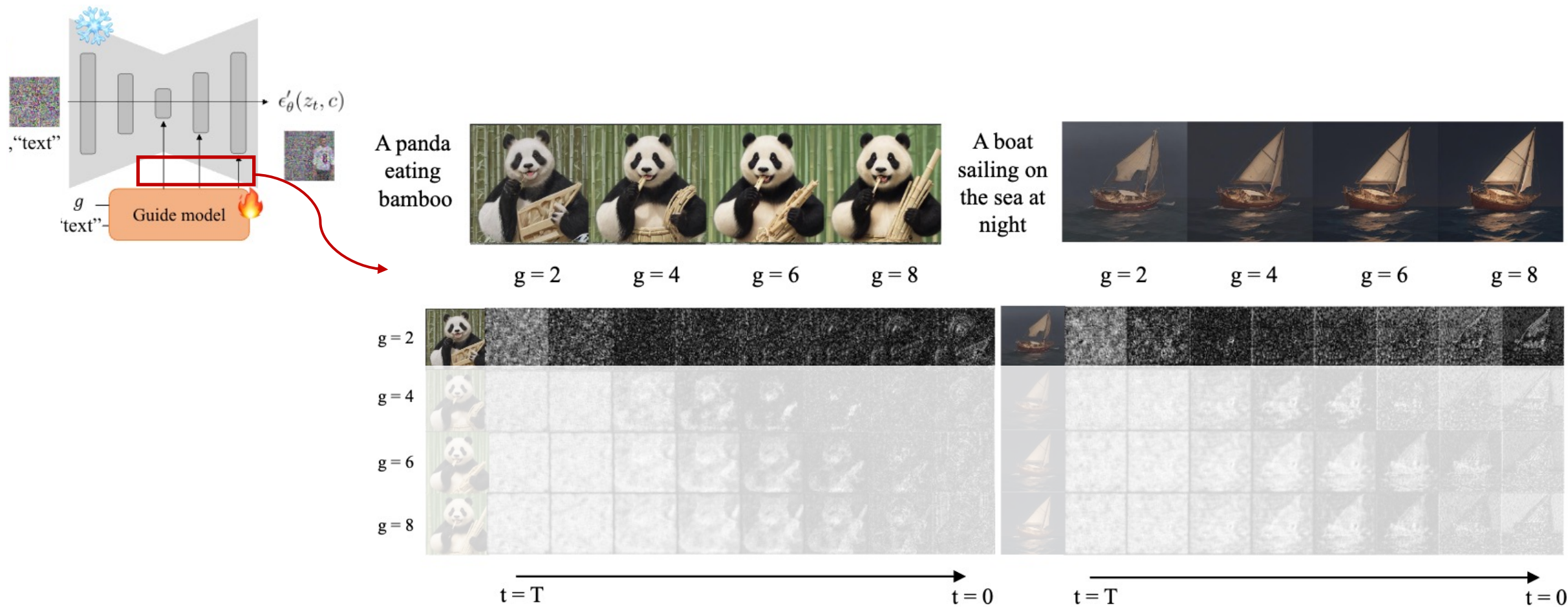# Tiny guide model is introduced to reduce the model size

(a) Full guide model
(b) Tiny guide model

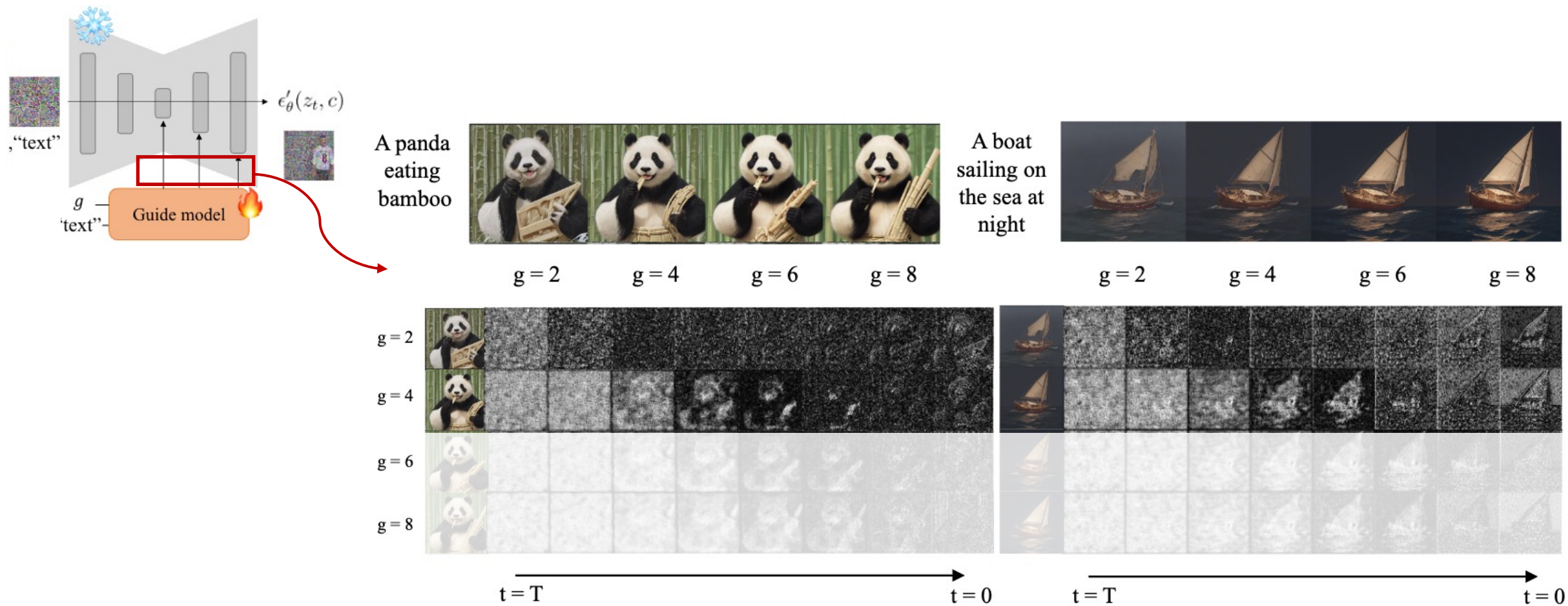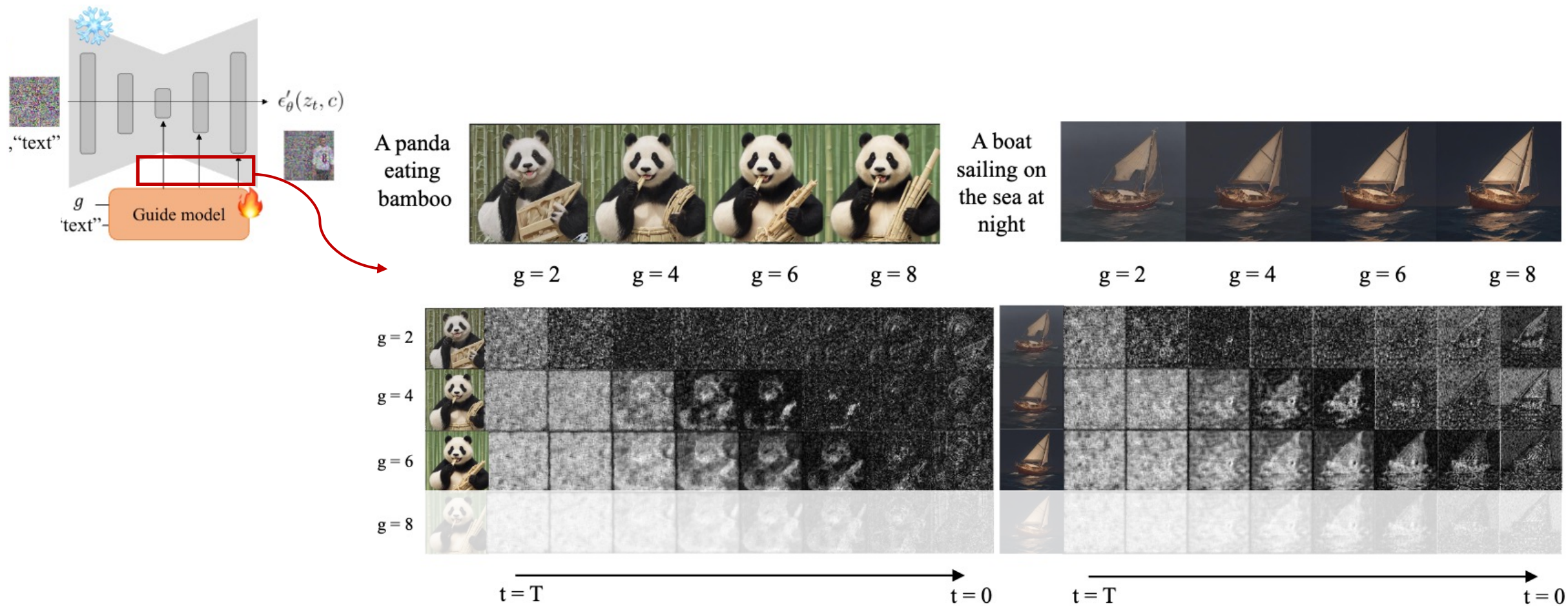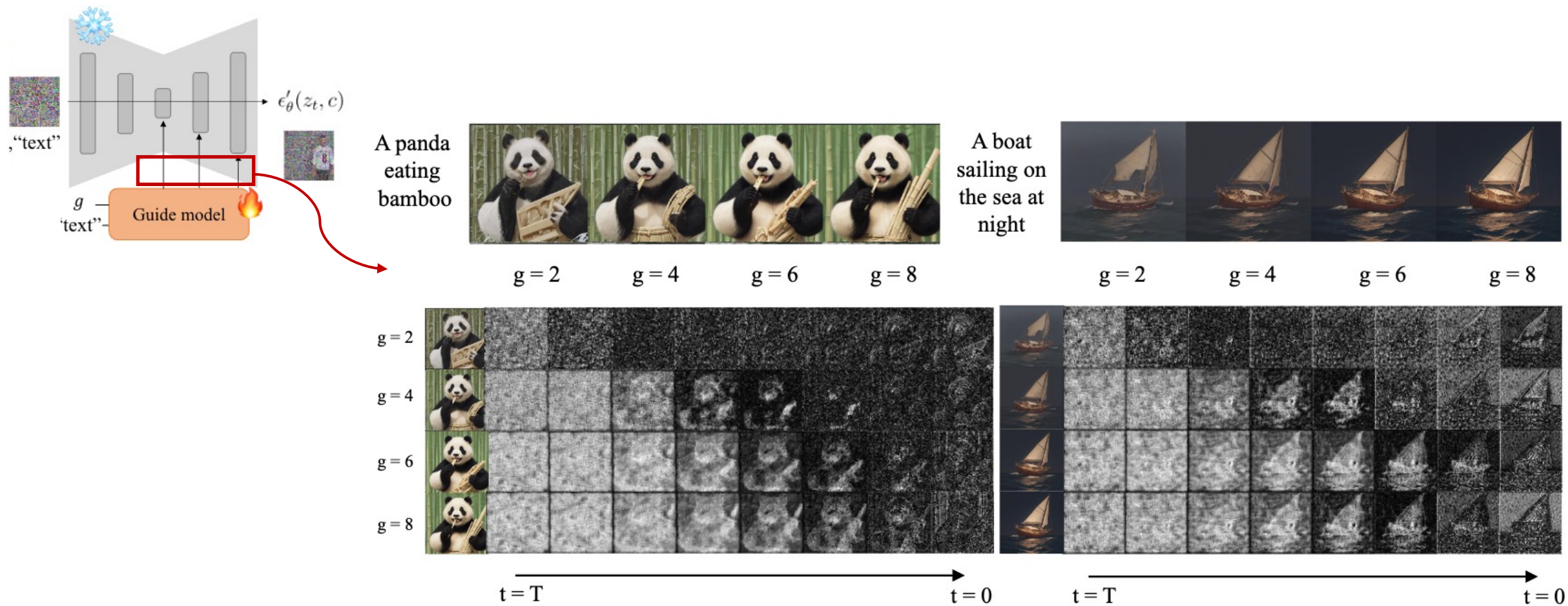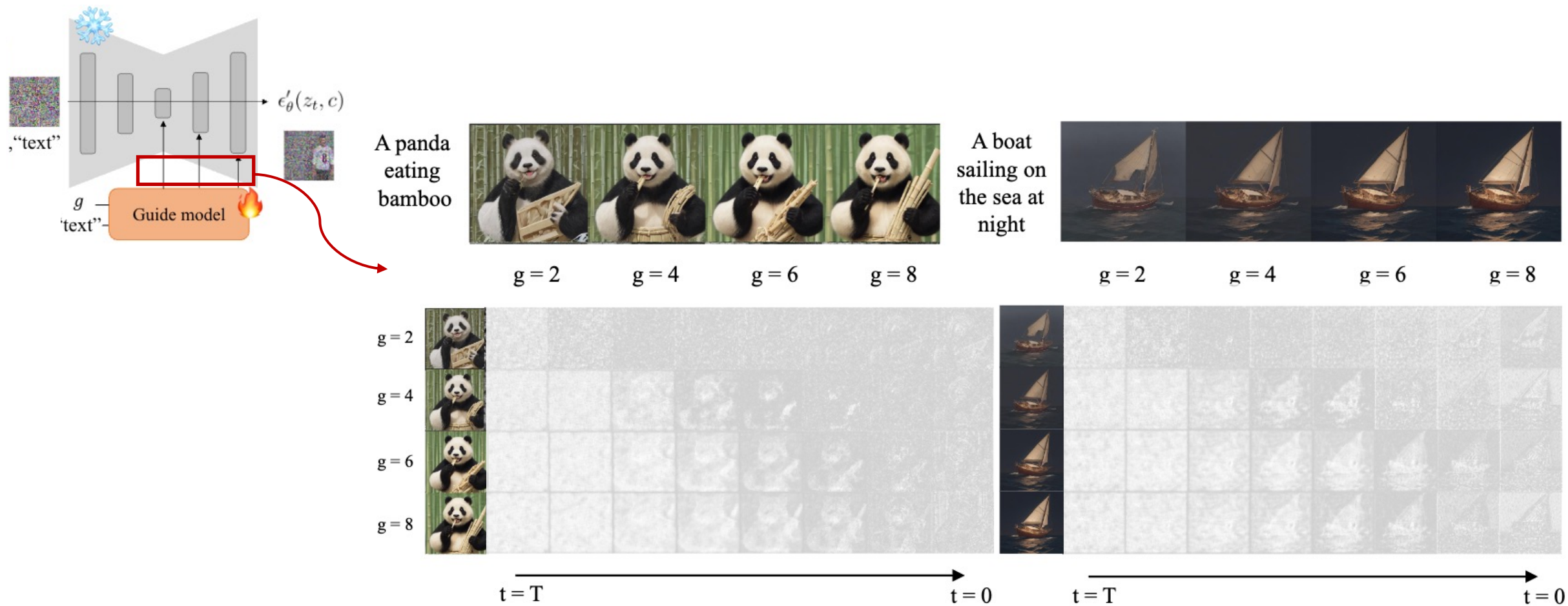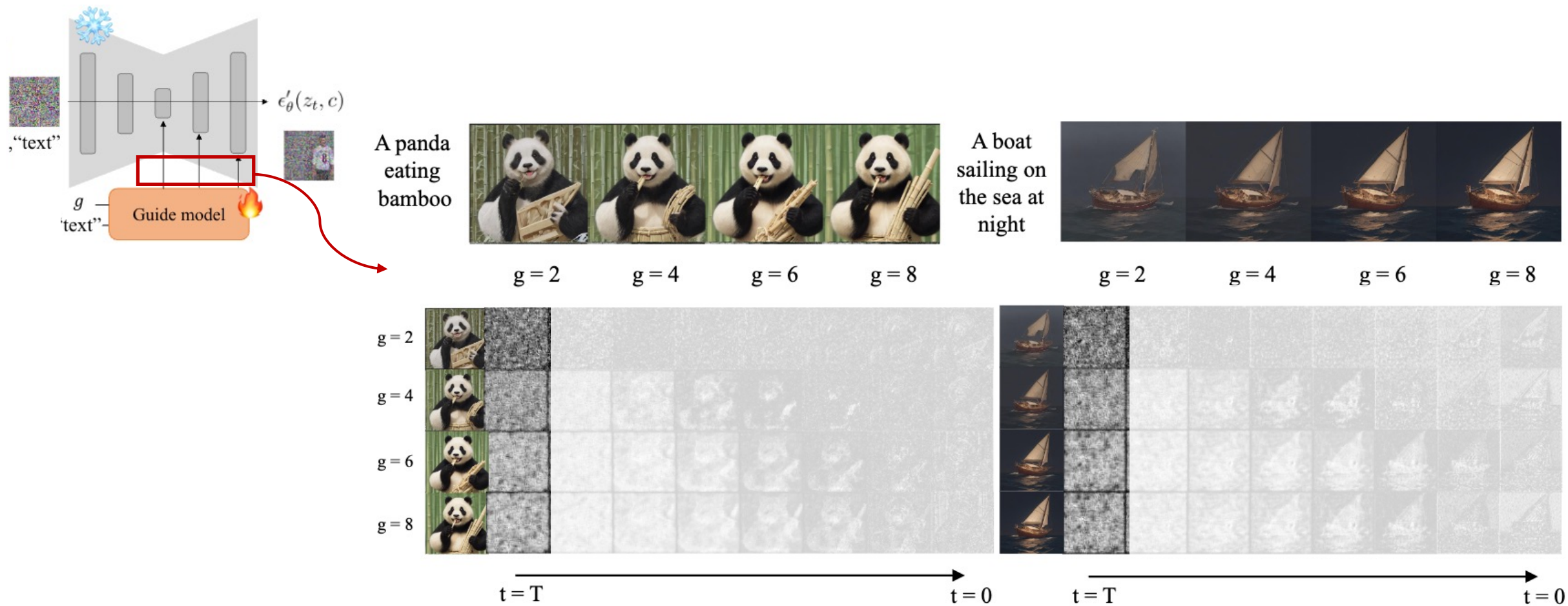# Stage 2 distillation: sampling step distillation

Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation
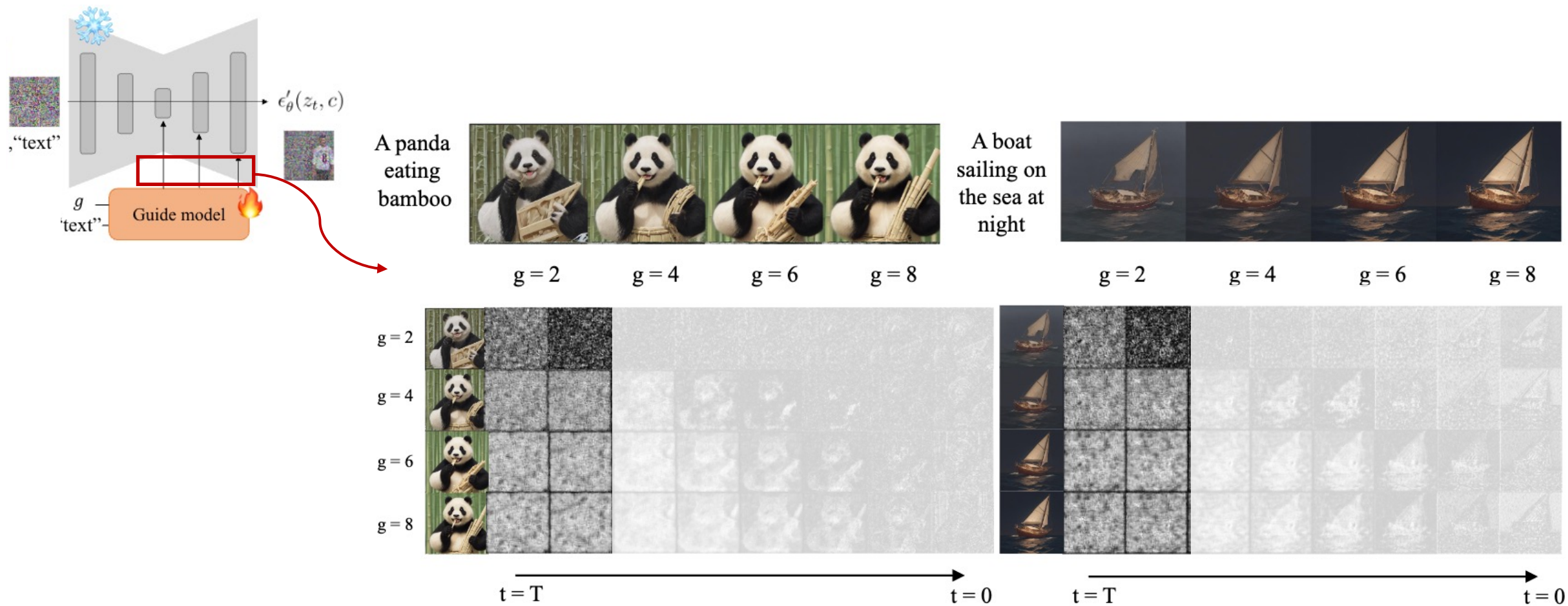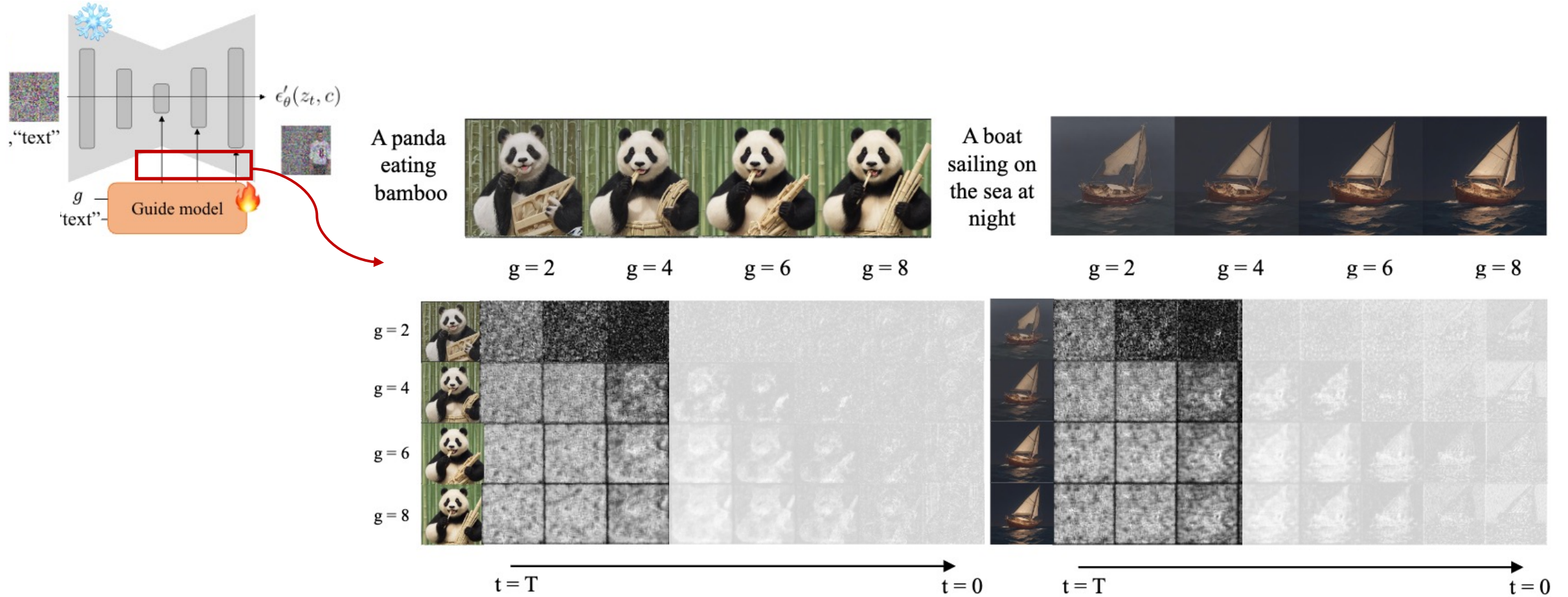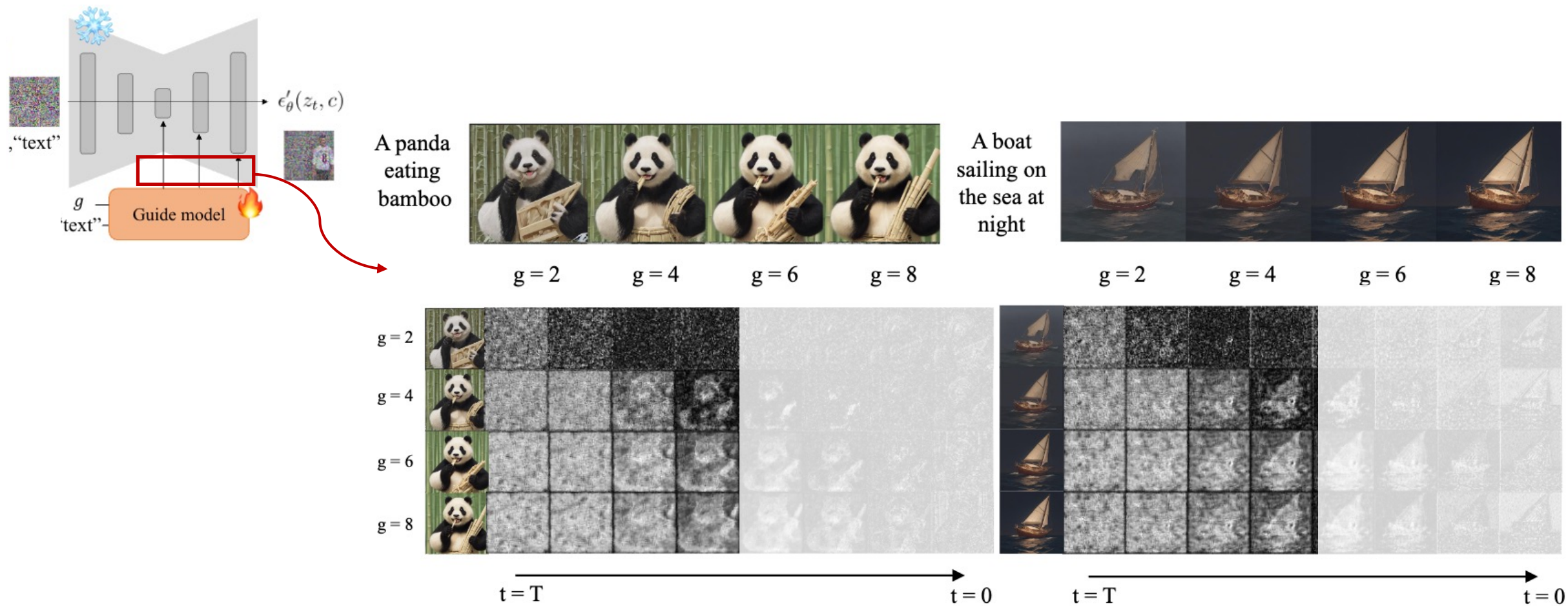
Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

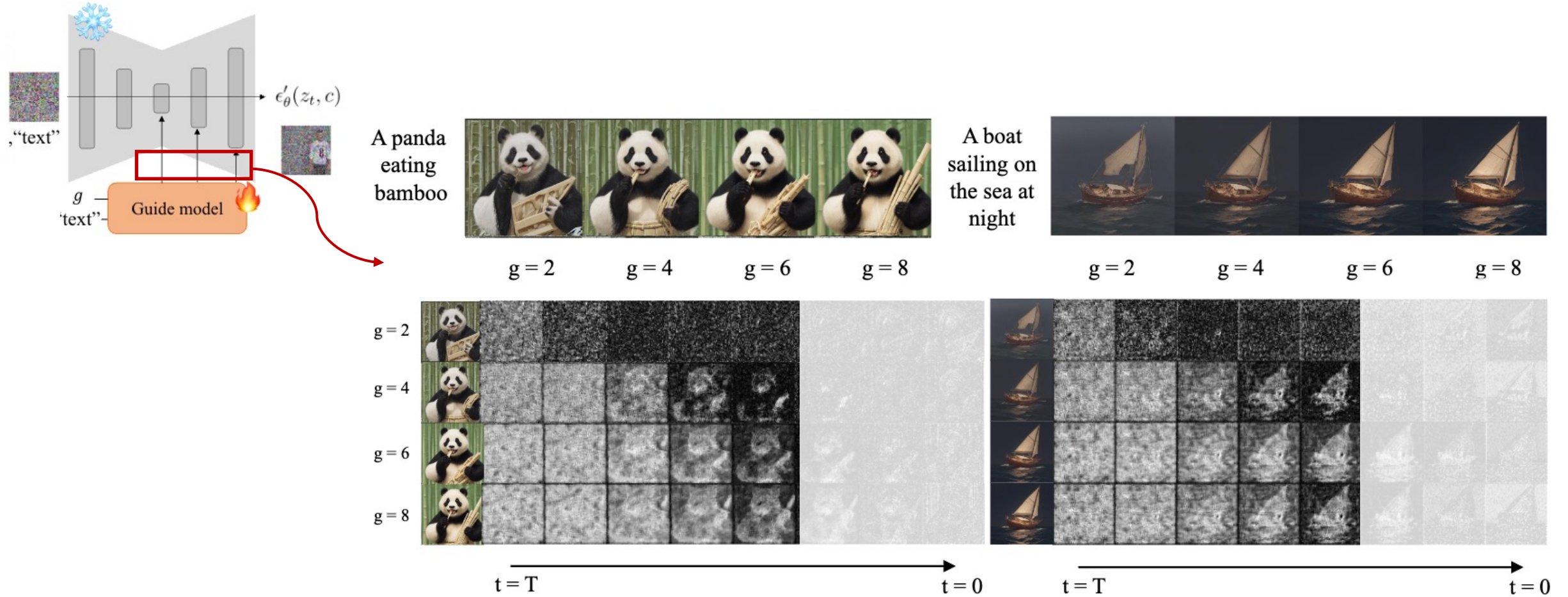# Visualizing Guide model Latent Representation

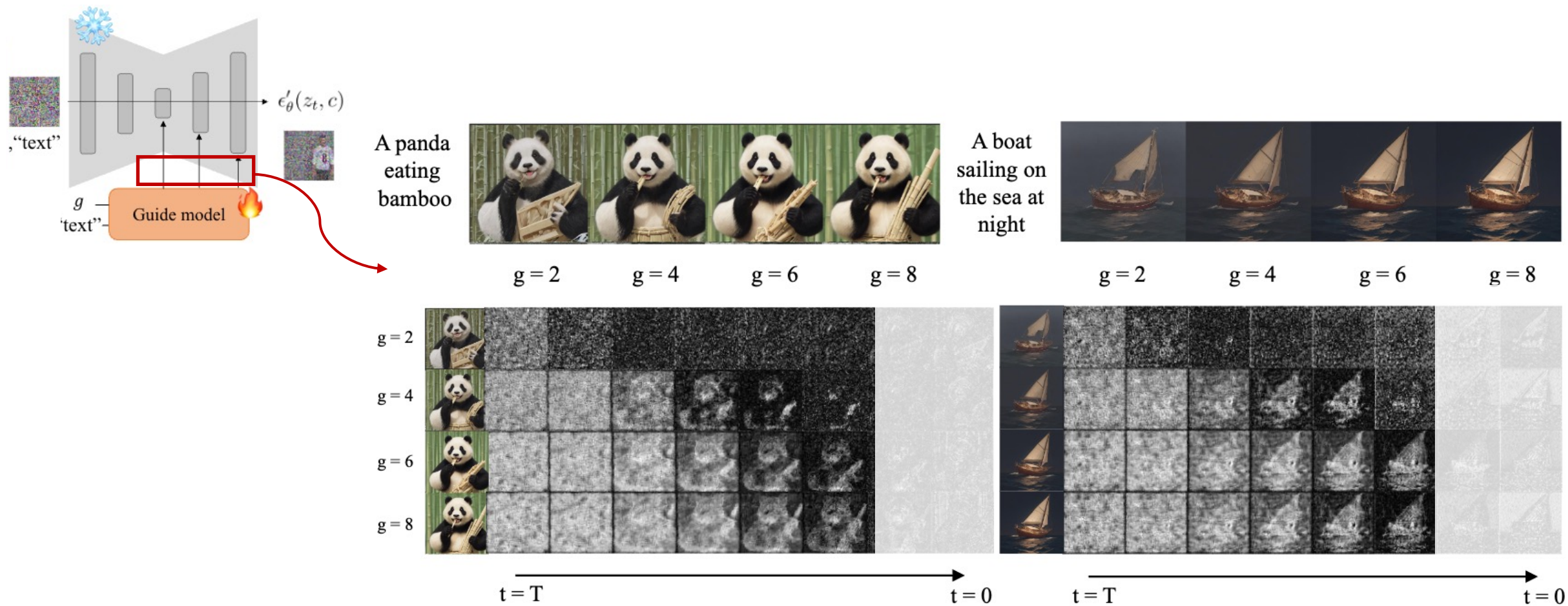# Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

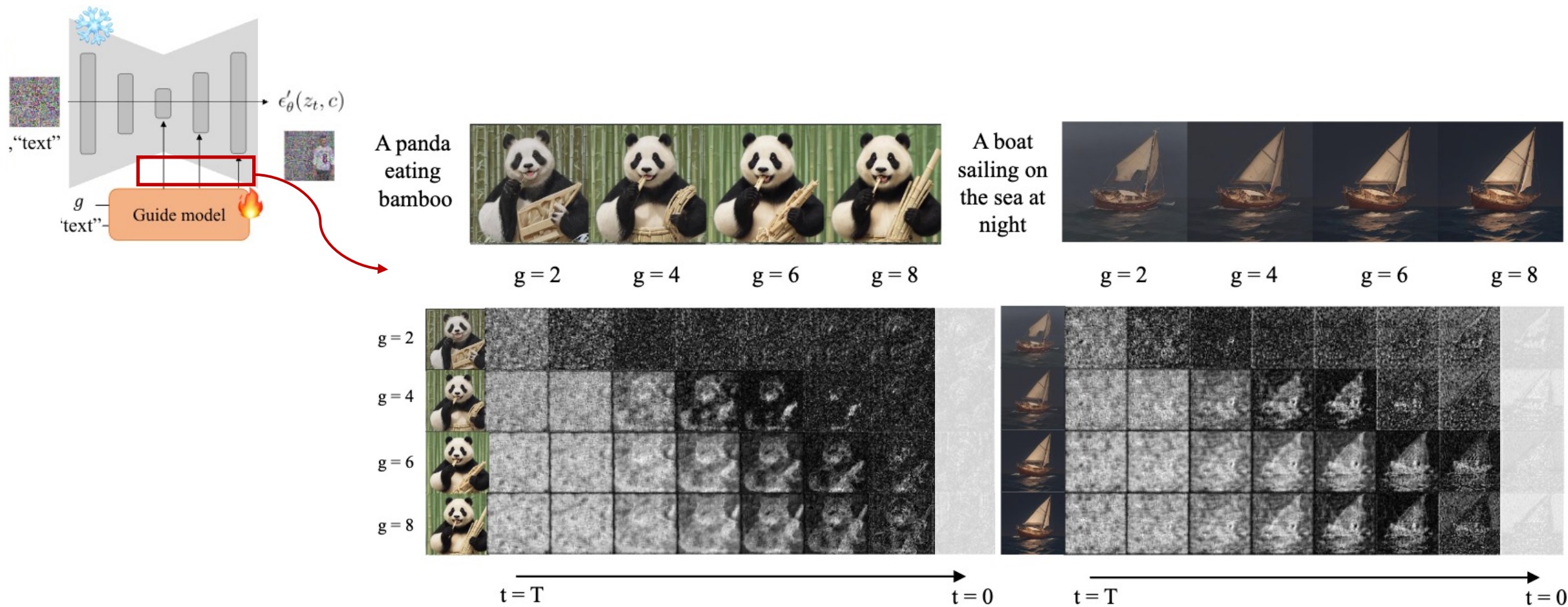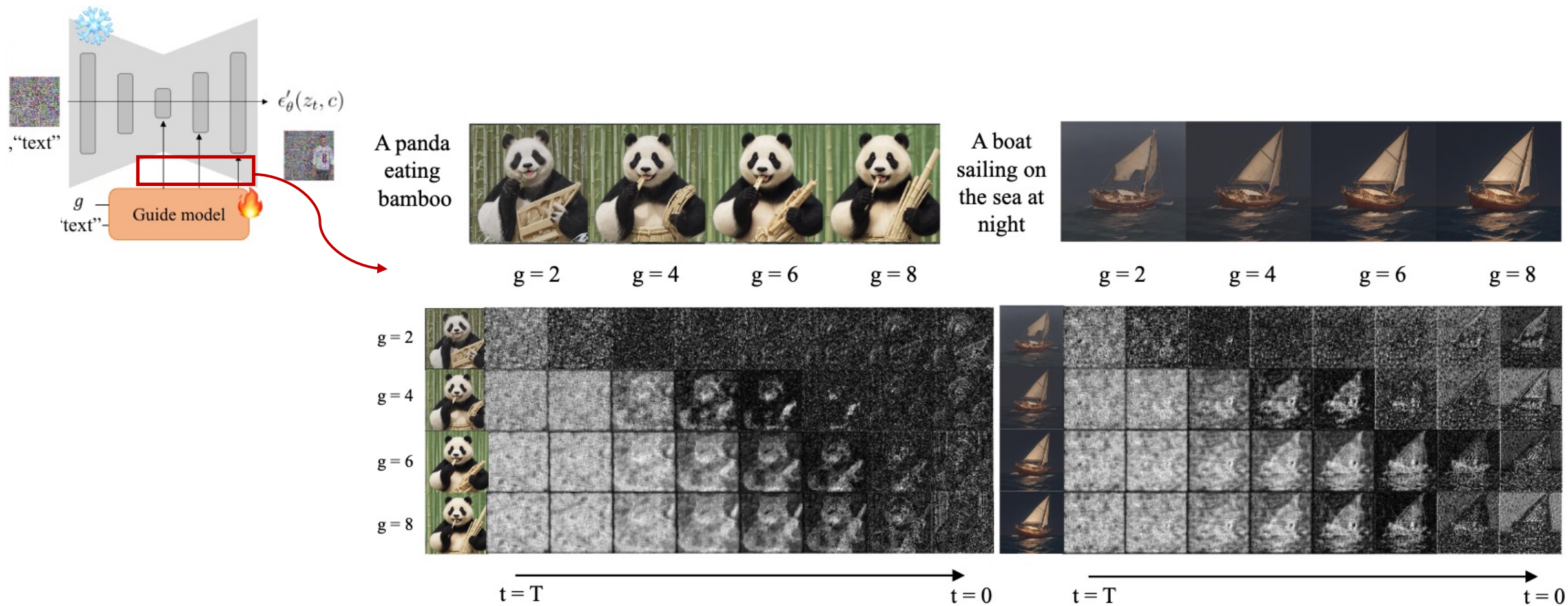# Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

Visualizing Guide model Latent Representation

https://5410tiffany.github.io/plug-and-play-diffusion-distillation.github.io/