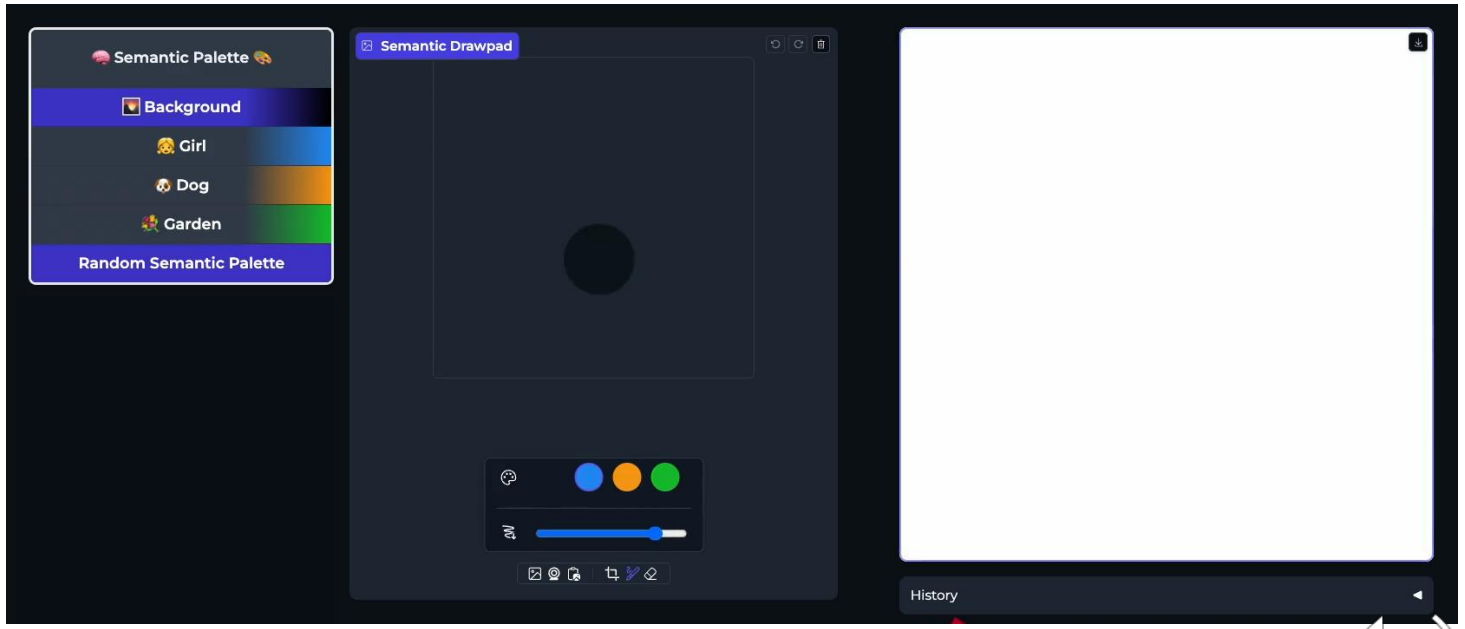
CVPR Nashville JUNE 11-15, 2025

Semantic Draw: Towards Real-Time Interactive Content Creation from Image Diffusion Models

Jaerin Lee, Daniel Sungho Jung, Kanggeon Lee, and Kyoung Mu Lee

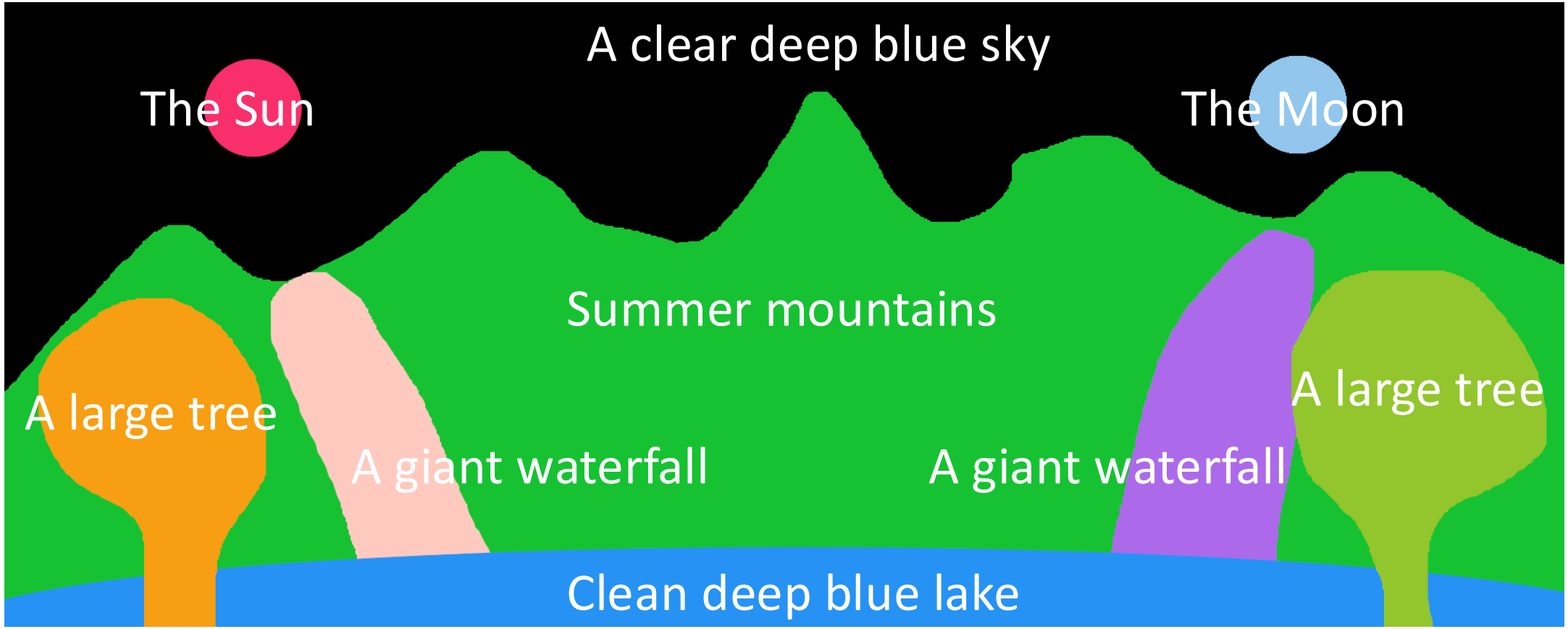
ComputerVisionLab
Seoul National University



Motivation

Irworobongdo, Korean royal folding screen, Joseon Dynasty.

Motivation



1280 X 768 prototyping of Irworobongdo with 9 prompt-mask pairs.

Motivation

SLOW!

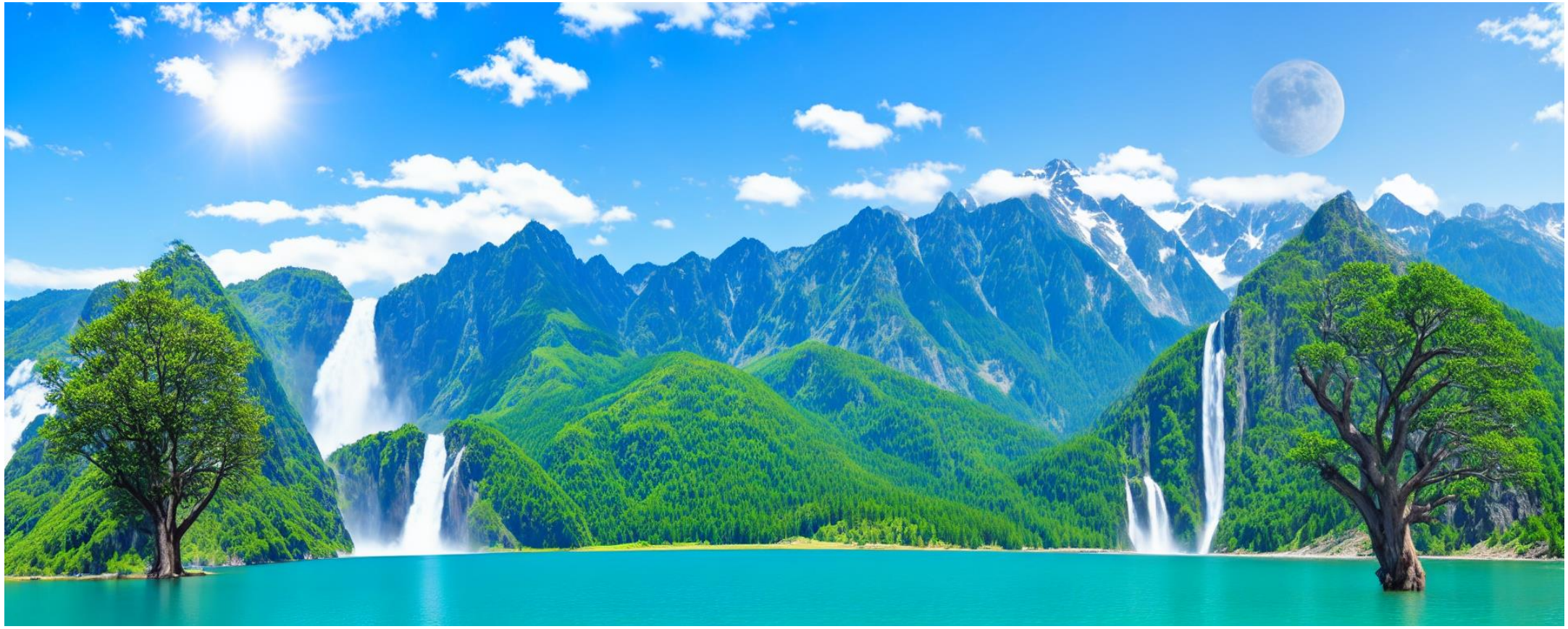
Inconsistent!

MultiDiffusion, took 51 minute 39 seconds.

Motivation

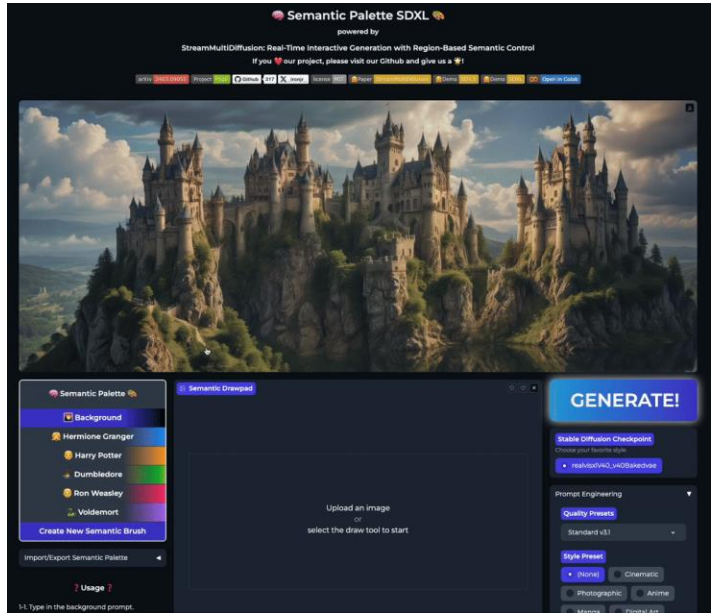
MD + LCM, took 4 minute 47 seconds.

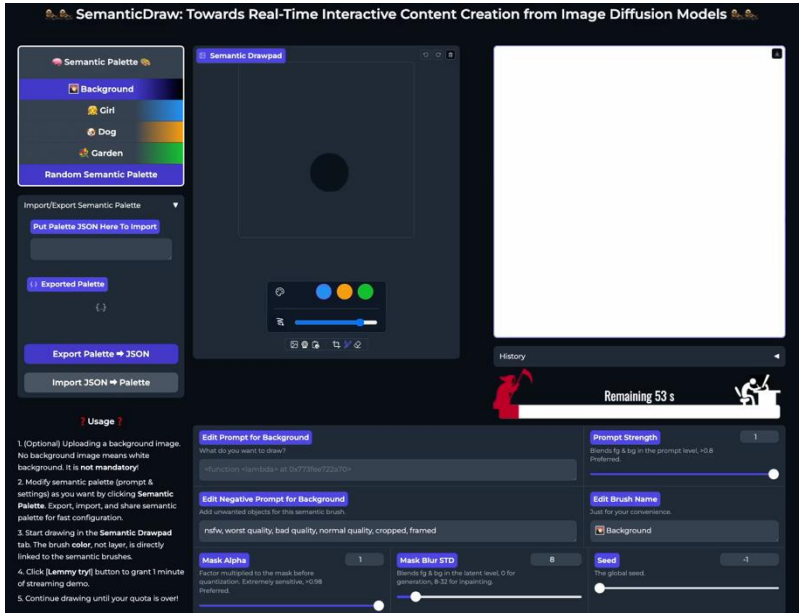
Motivation



Ours, took 59 seconds.

Our Solution





Establish compatibility:

Up to 50x faster

multi-prompt generation

Maximize throughput with

Multi-Prompt Stream Batch

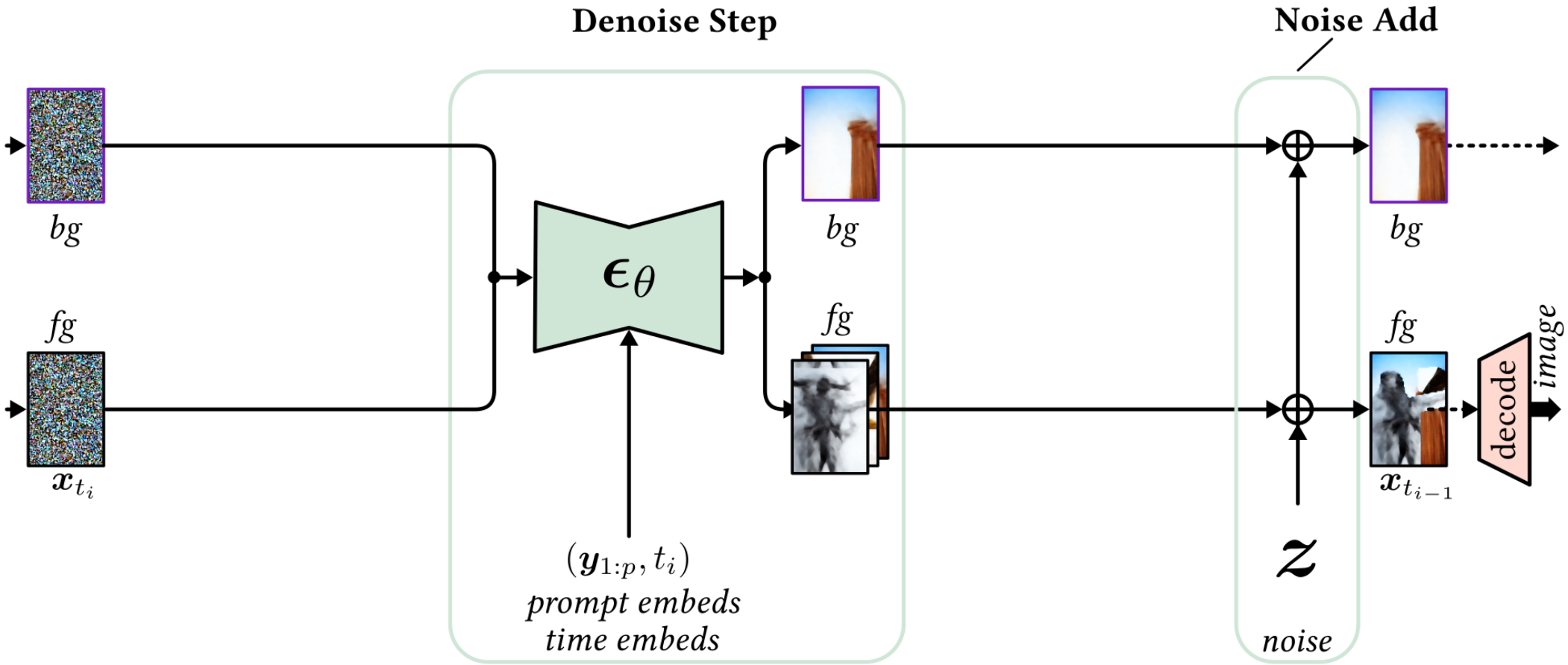
pipeline architecture

New Drawing Paradigm:

Semantic Draw

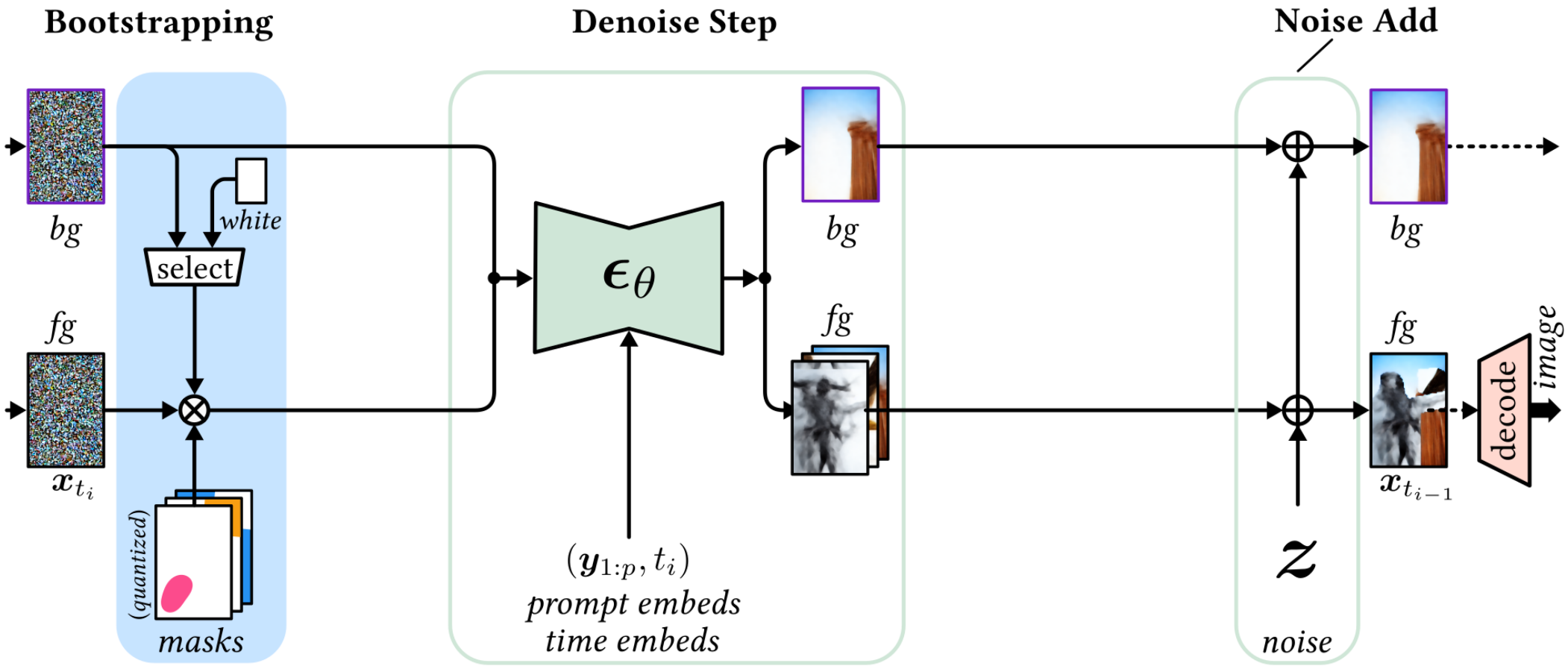
with demo applications

Method: Redesigning the Pipeline



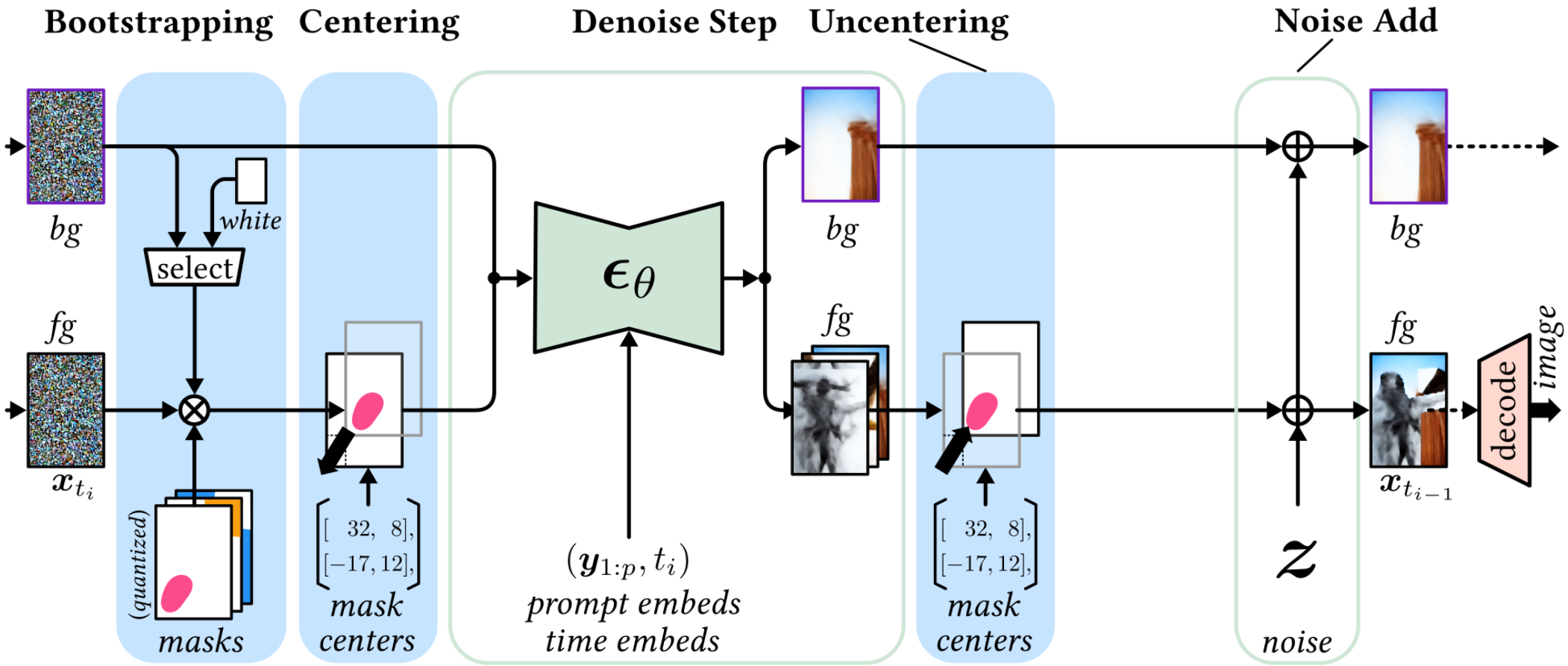
(a) Any image diffusion pipeline can be augmented with our compatibility module.

Method: Redesigning the Pipeline



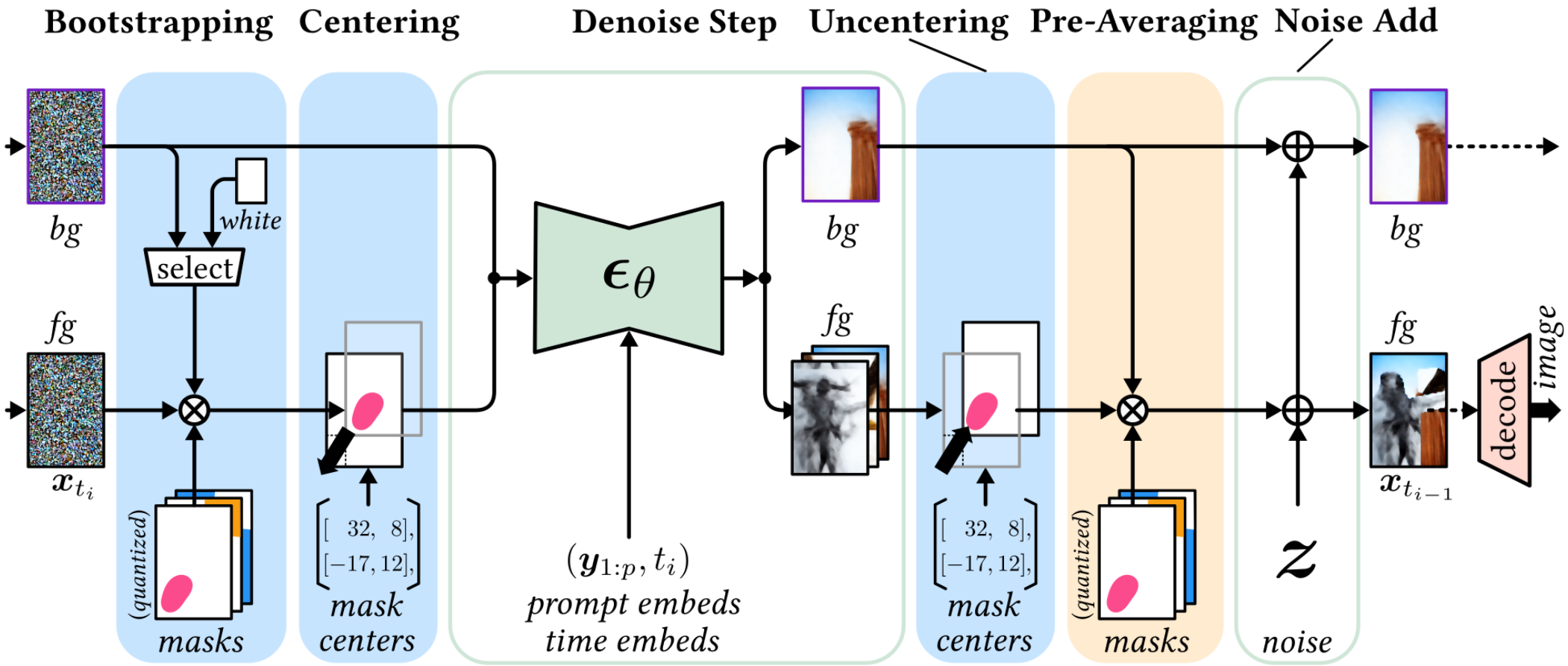
(a) Any image diffusion pipeline can be augmented with our compatibility module.

Method: Redesigning the Pipeline



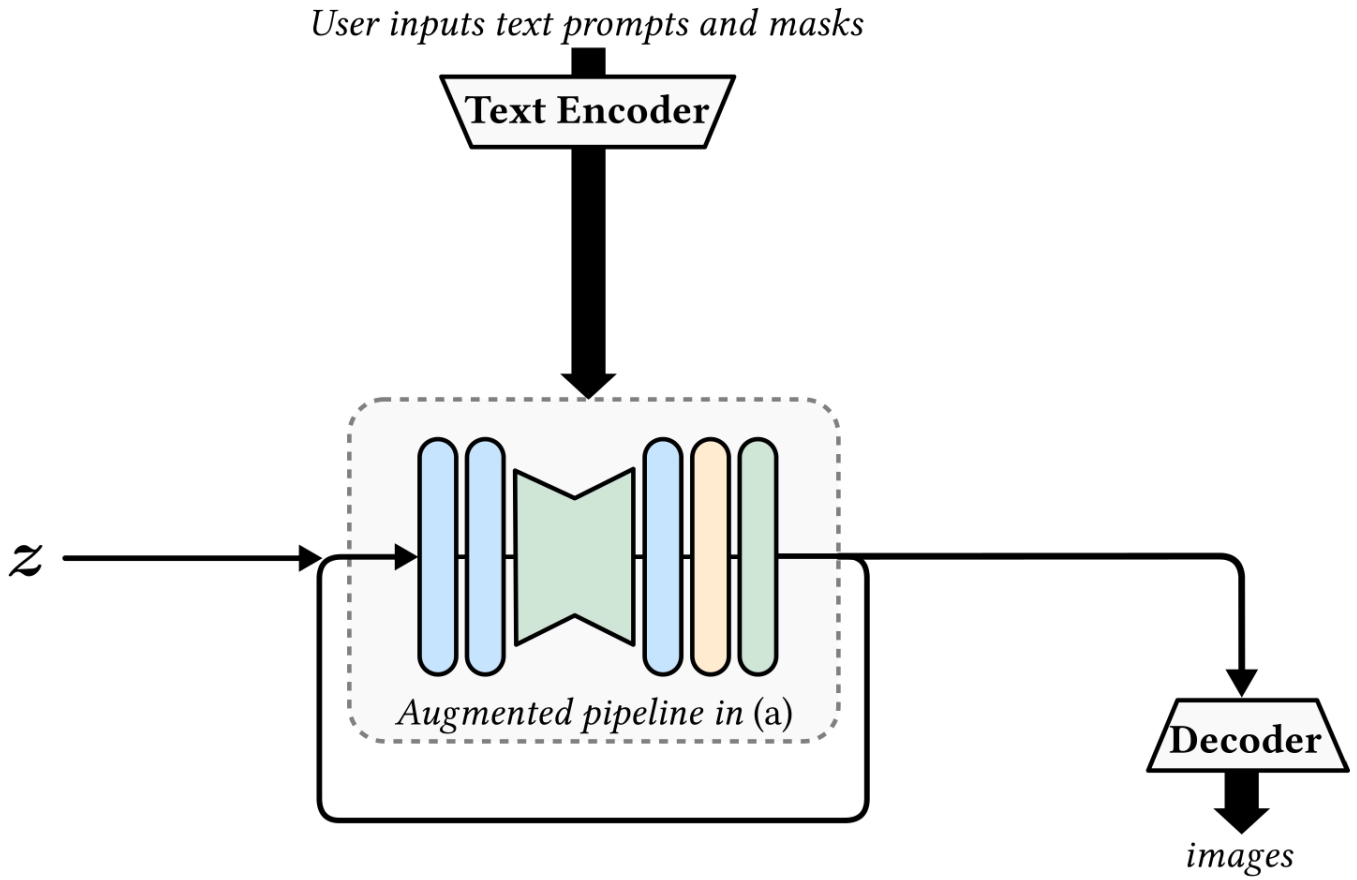
(a) Any image diffusion pipeline can be augmented with our compatibility module.

Method: Redesigning the Pipeline



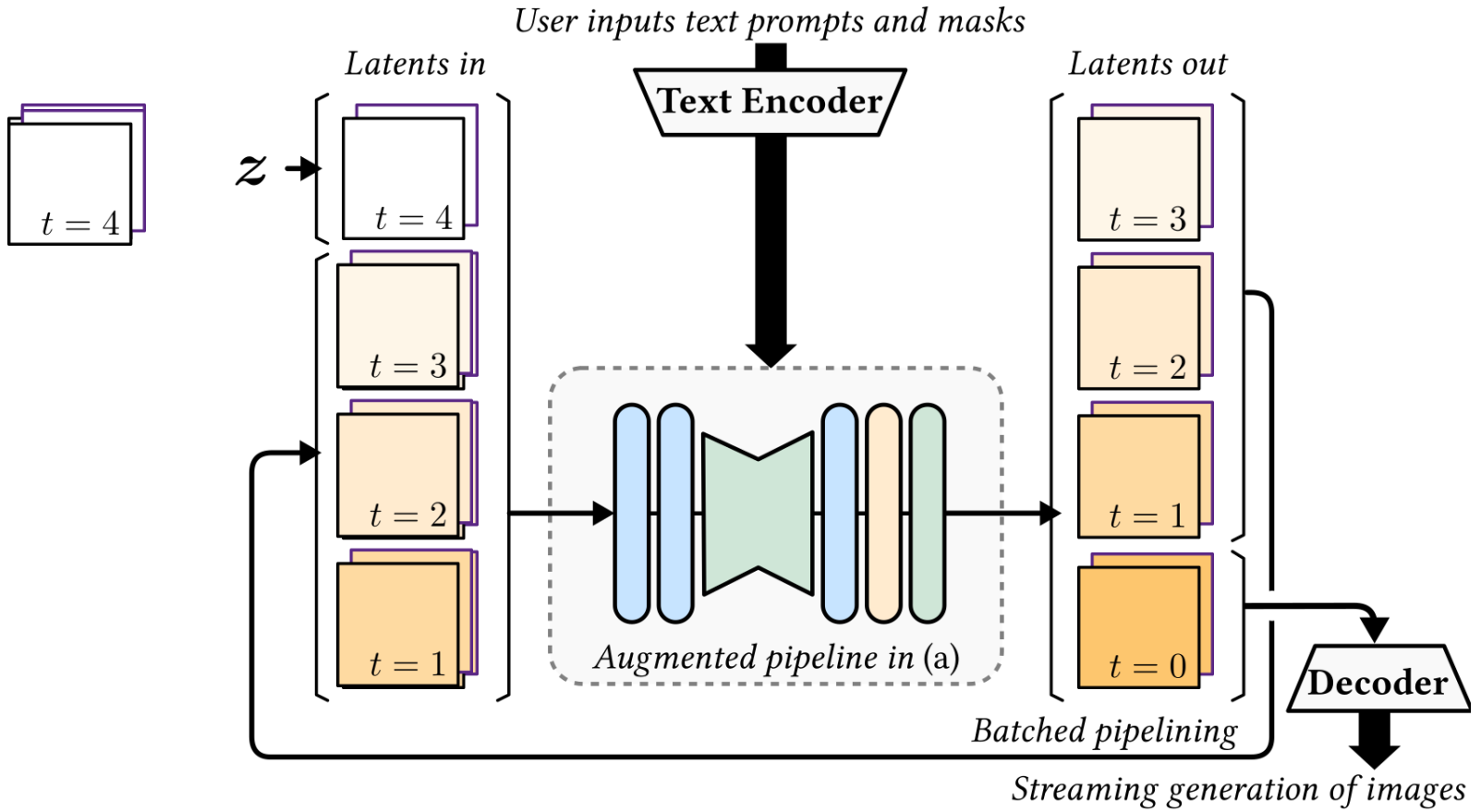
(a) Any image diffusion pipeline can be augmented with our compatibility module.

Method



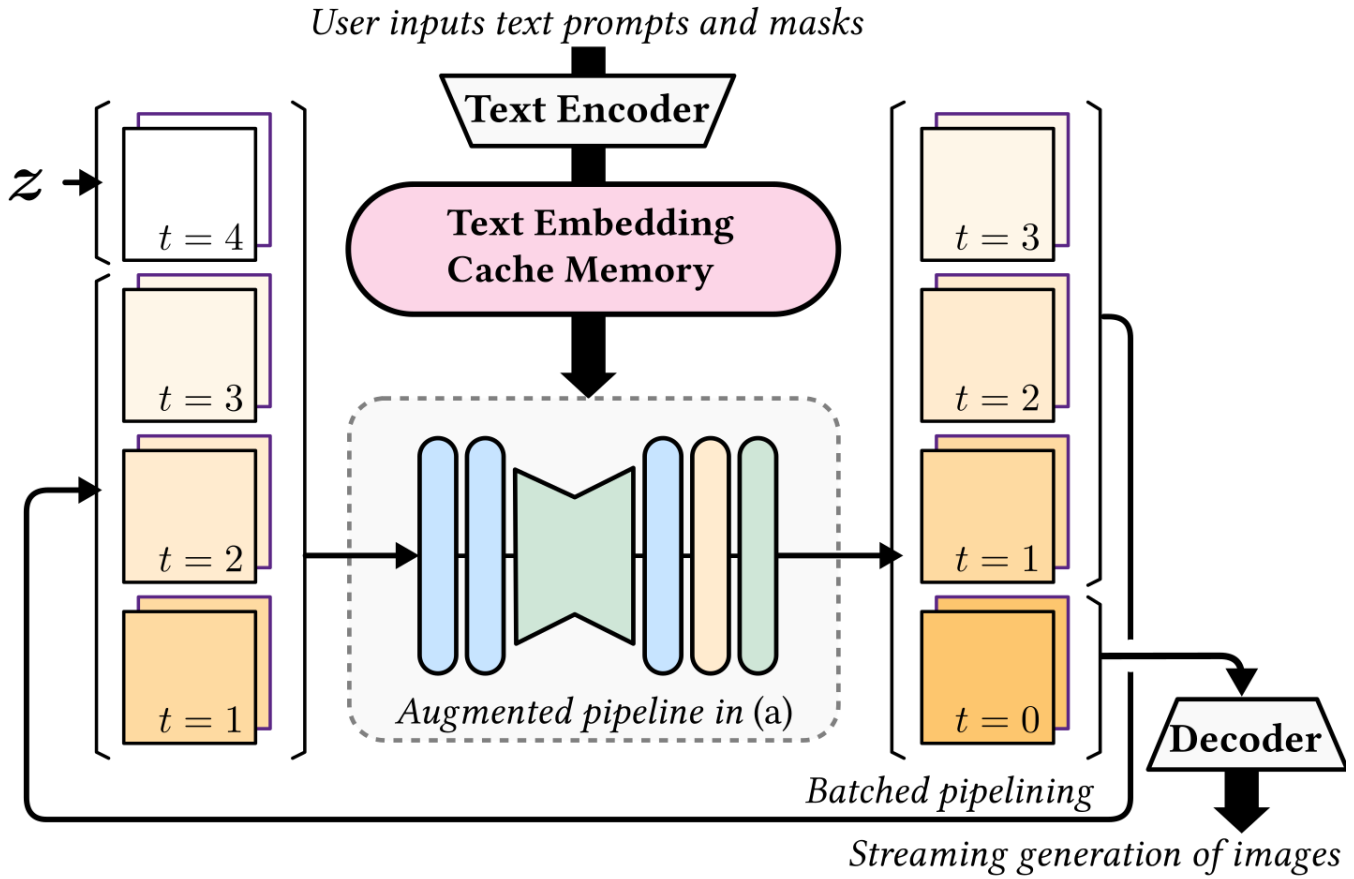
(b) Streaming by aggregating the regional latents of different timesteps. Cache data to remove redundancy.

Method



(b) Streaming by aggregating the regional latents of different timesteps. Cache data to remove redundancy.

Method



(b) Streaming by aggregating the regional latents of different timesteps. Cache data to remove redundancy.

Results

- New wine in new wineskin; New powerful tools like diffusion models deserves new type of applications.

Method	Sampler	FID ↓	IS ↑	$CLIP_{fg}$ ↑	$CLIP_{bg}$ ↑	Time (s) ↓
SD1.5 (512 × 512)						
MultiDiffusion (Ref.)	DDIM [47]	70.93	16.24	24.09	27.55	14.1
MultiDiffusion (MD)	LCM [31]	270.55	2.653	22.53	19.63	1.7
SemanticDraw (Ours)	LCM [31]	93.93	14.12	24.14	24.00	1.3

Method	Sampler	FID ↓	IS ↑	$CLIP_{fg}$ ↑	$CLIP_{bg}$ ↑	Time (s) ↓
SDXL (1024 × 1024)						
MultiDiffusion (Ref.)	DDIM [47]	73.77	16.31	24.16	28.11	50.6
MultiDiffusion (MD)	EulerDiscrete [18]	572.95	1.328	21.02	17.36	4.3
SemanticDraw (Ours)	EulerDiscrete [18]	84.27	15.04	24.19	24.22	3.6

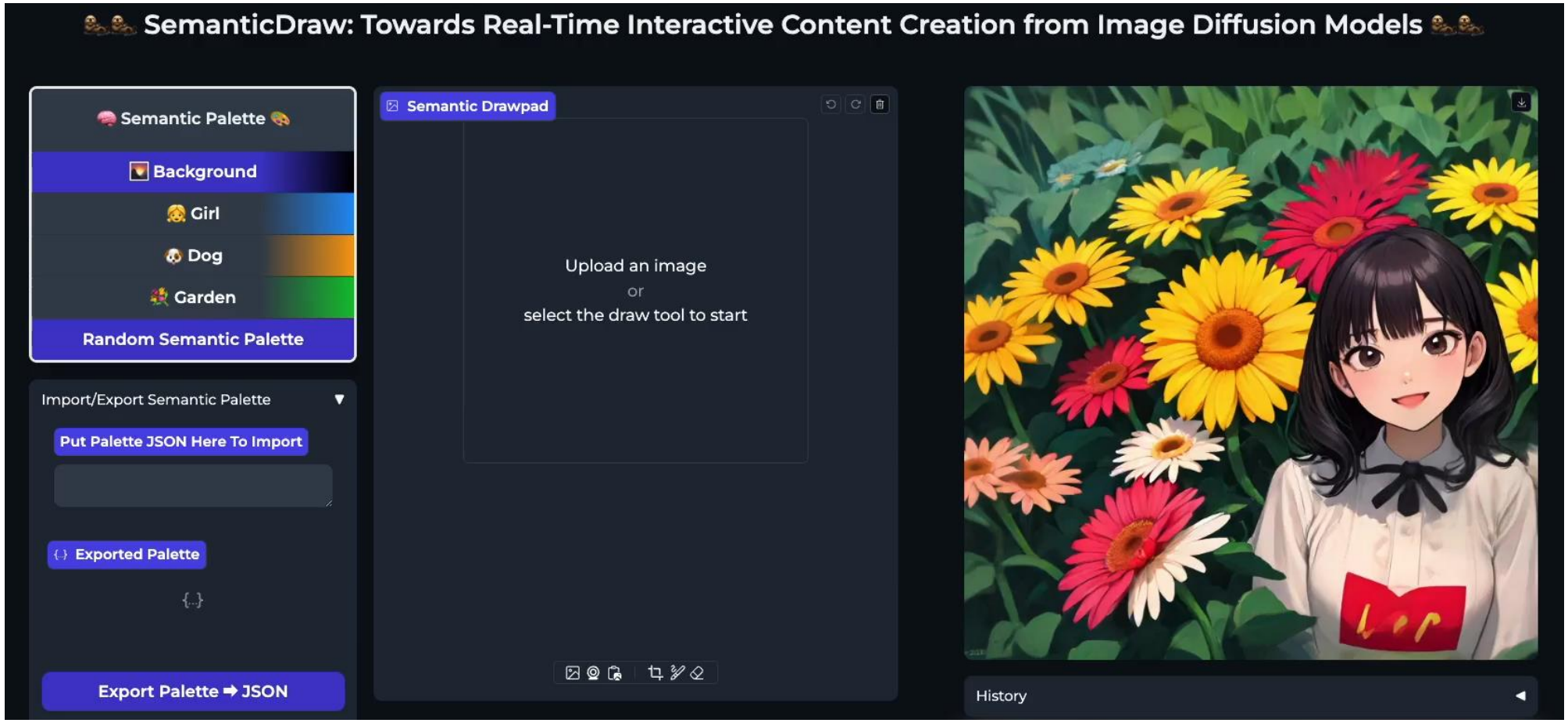
Method	Sampler	FID ↓	IS ↑	$CLIP_{fg}$ ↑	$CLIP_{bg}$ ↑	Time (s) ↓
SD1.5 (512 × 512)						
MultiDiffusion (Ref.)	DDIM [47]	70.93	16.24	24.09	27.55	14.1
MultiDiffusion (MD)	Hyper-SD [40]	168.34	10.12	20.08	15.90	1.7
SemanticDraw (Ours)	Hyper-SD [40]	98.60	14.90	24.48	23.31	1.3

Method	Sampler	FID ↓	IS ↑	$CLIP_{fg}$ ↑	$CLIP_{bg}$ ↑	Time (s) ↓
SD3 (1024 × 1024)						
MultiDiffusion (Ref.)	FlowMatch [9]	166.42	8.517	20.66	16.39	46.3
MultiDiffusion (MD)	FlashFlowMatch [7]	209.36	5.347	19.83	14.48	4.0
SemanticDraw (Ours)	FlashFlowMatch [7]	79.2	17.41	23.59	27.83	3.2

Results

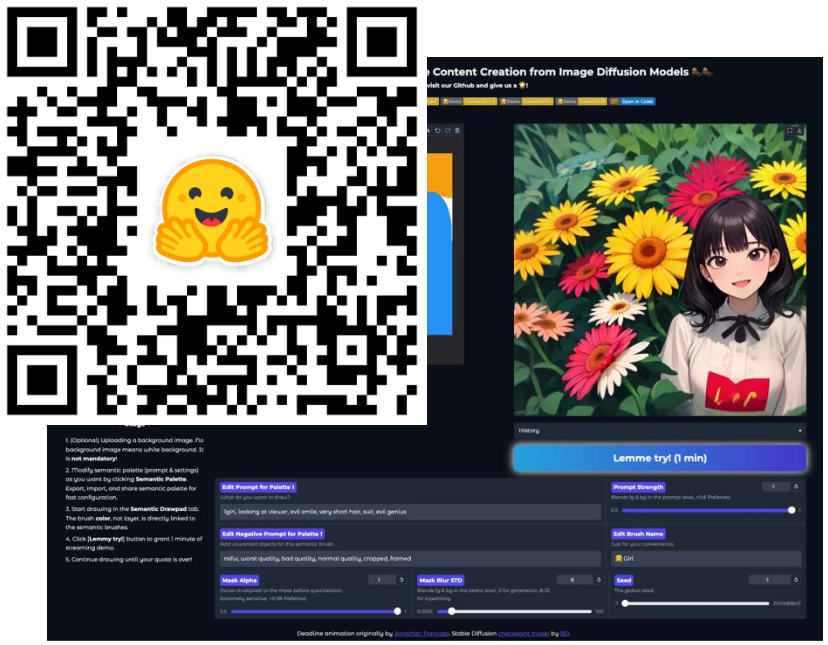
Ours, took 59 seconds.

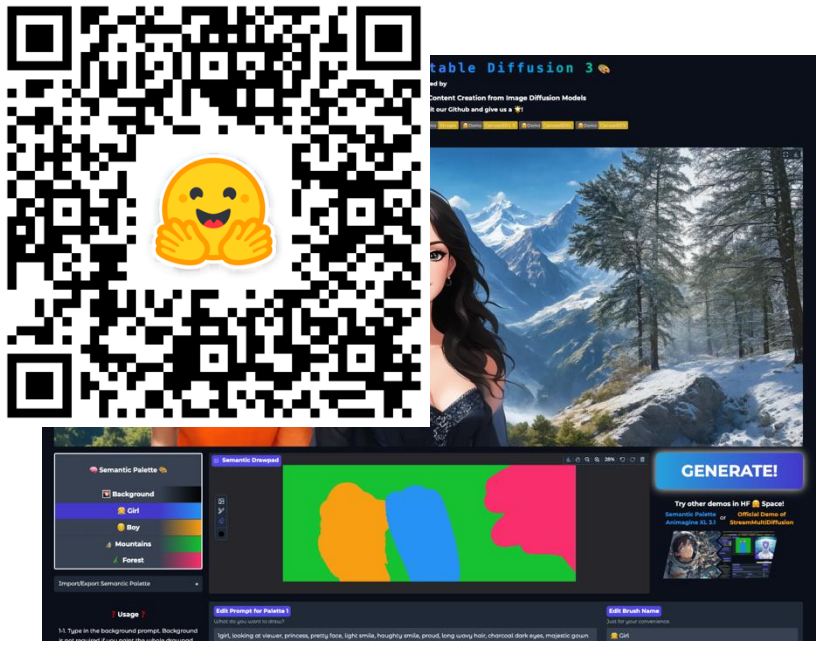
Application



Take Home Message

- New wine in new wineskin; New powerful tools like diffusion models deserves new type of applications.





Thank You

ironjr@snu.ac.kr

ComputerVisionLab
Seoul National University

CVPR Nashville JUNE 11-15, 2025