SAGIVTECH IMVC 2026

SIGGRAPH ASIA 2025 香港 HONG KONG

OmnimatteZero
Fast Training-free Omnimatte with Pre-trained Video Diffusion Models

Dvir Samuel[1], Matan Levy[3], Nir Darshan[1], Gal Chechik[2,4], Rami Ben-Ari[1]

[1] OriginAI
[2] Bar-Ilan University
[3] The Hebrew University of Jerusalem
[4] NVIDIA Research

OriginAI

THE HEBREW UNIVERSITY OF JERUSALEM

NVIDIA

The “Omnimatte” Task

Core Task: A fundamental problem in video understanding and editing is isolating objects from their scenes.

Applications: background replacement, retiming, duplicating actors…

The “Omnimatte” Task

Decompose a video into layers: background layers and foreground layers **WITH** all of the object associated effects.

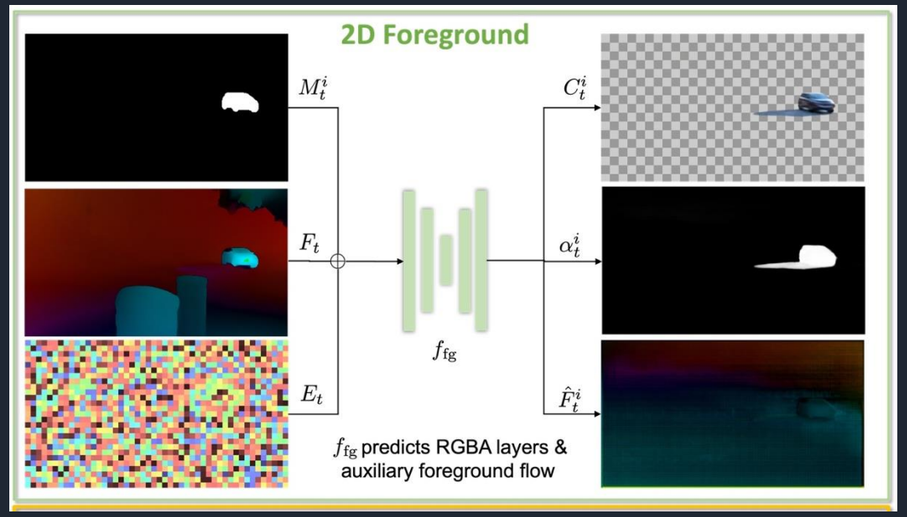
Generative Omnimatte Learning to Decompose Video into Layers; Lee et al.

Current methods

Require extensive training or costly self-supervised optimization for each video

Current methods

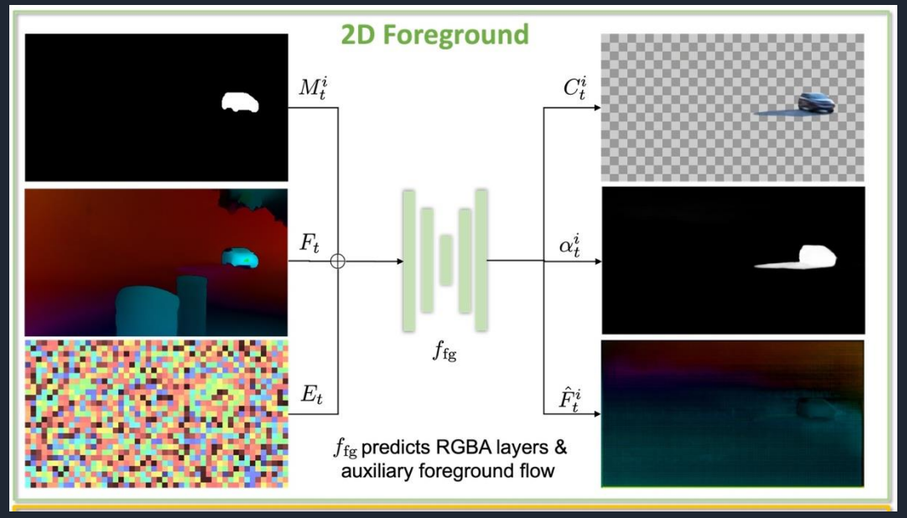
Require extensive training or costly self-supervised optimization for each video



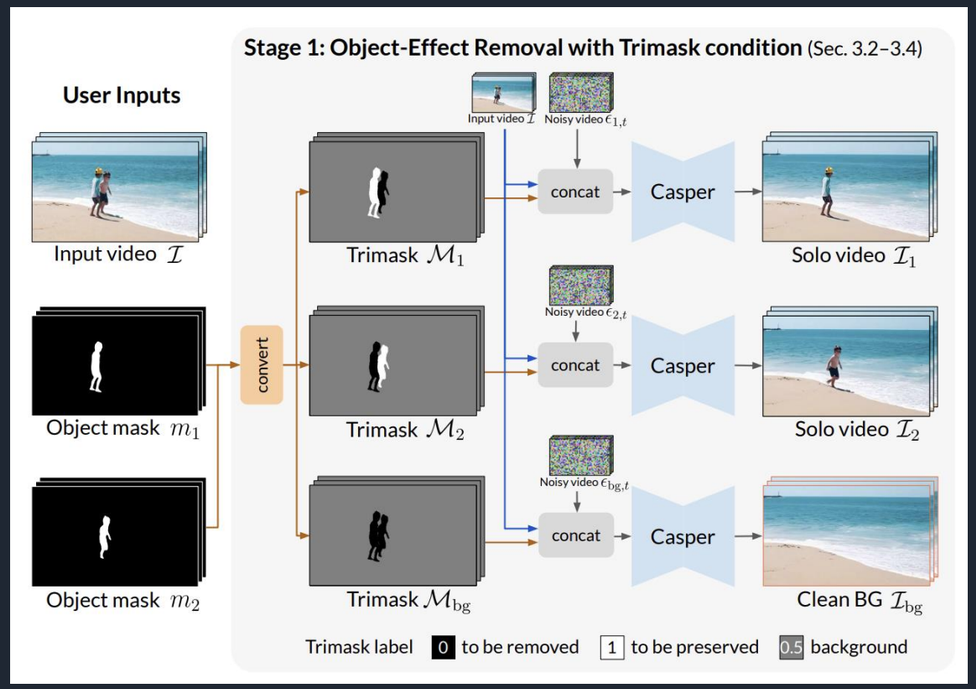
Optimization methods decompose the video into layers and then recompose them.

Current methods

Require extensive training or costly self-supervised optimization for each video



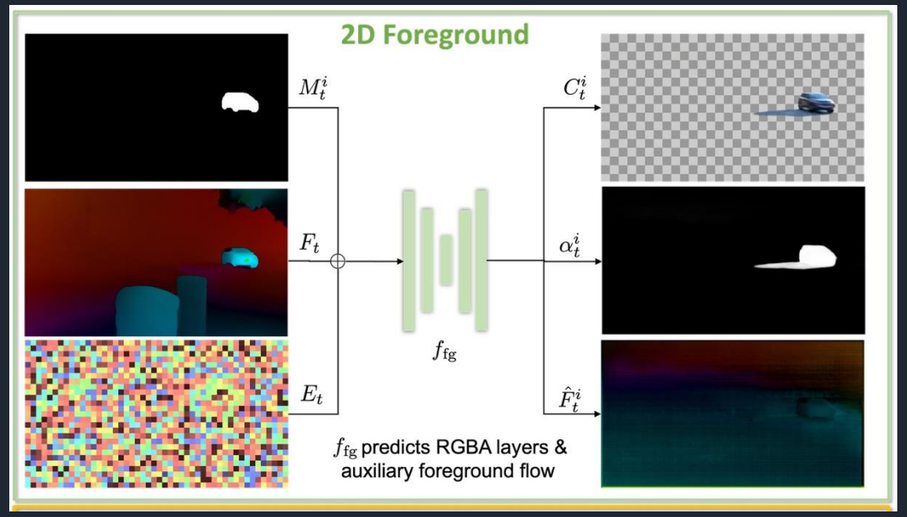
Optimization methods decompose the video into layers and then recompose them.



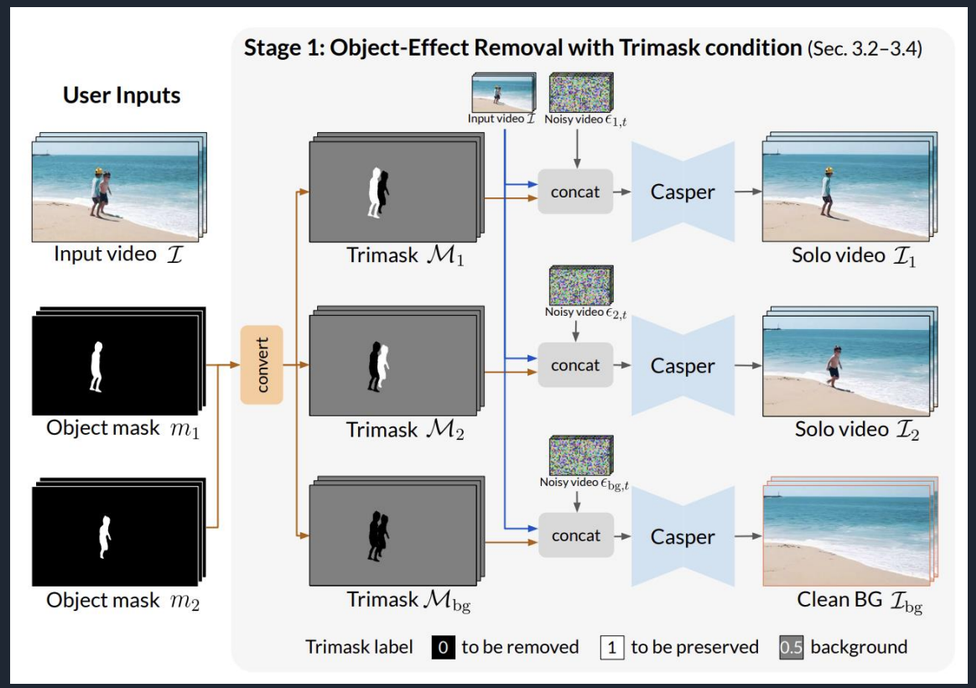
Training based methods trains a network to output each layer separately.

Current methods

Require extensive training or costly self-supervised optimization for each video



Optimization methods decompose the video into layers and then recompose them.



Training based methods trains a network to output each layer separately.

Challanges

1. **High computational cost:** hours to train/optimize per short video
2. **Slow inference/rendering:** ~3–9 seconds per frame.
3. **Limited generalization:** struggles with new videos and out-of-distribution objects.

OmniMatteZero in three steps

1

Object Removal

Remove all objects to create clean background.

2

Foreground Extraction

Isolate objects with their effects.

3

Layer Composition

Seamlessly composite onto new backgrounds.

Training Free!

No Optimization!!

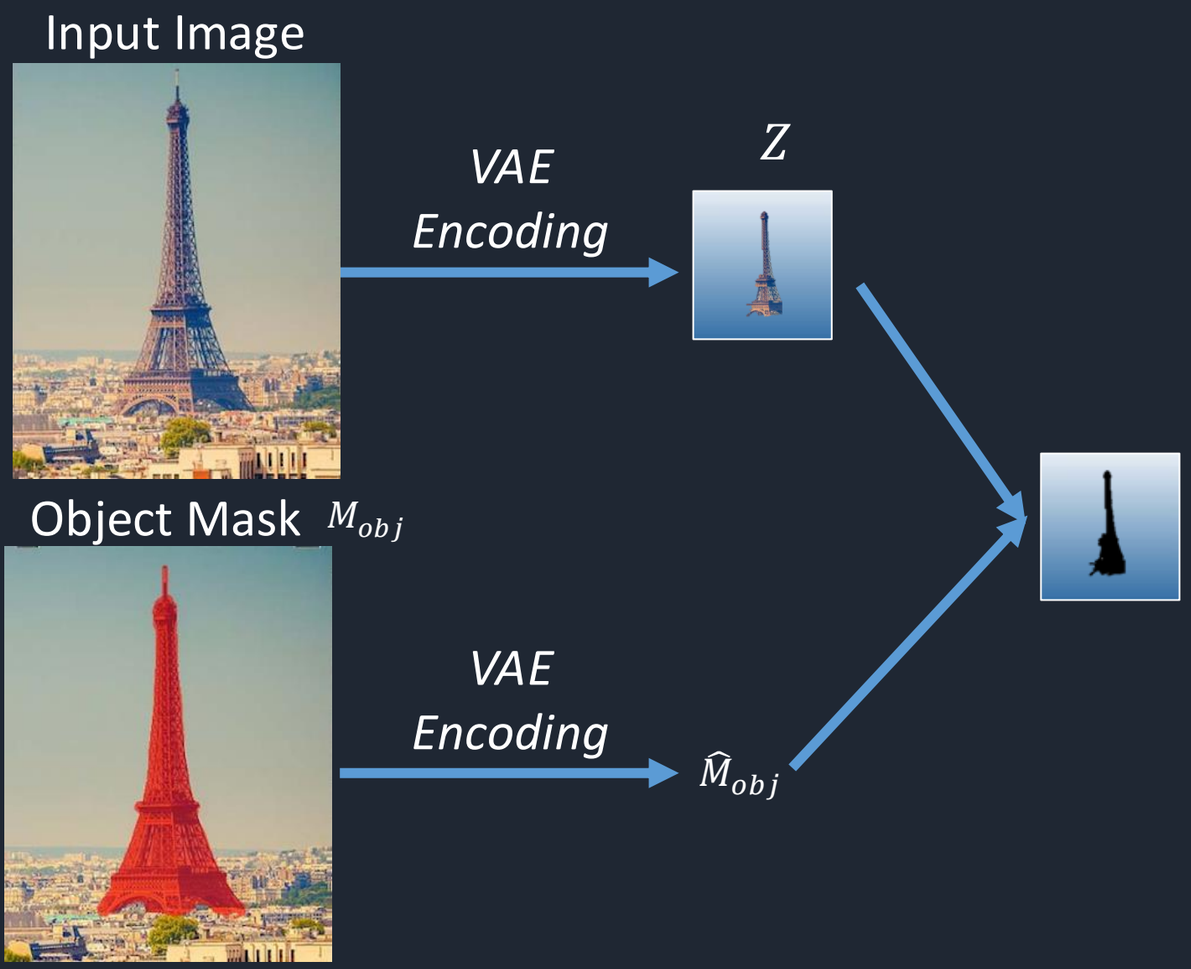
Really Fast!!

Zero-shot Image Inpainting

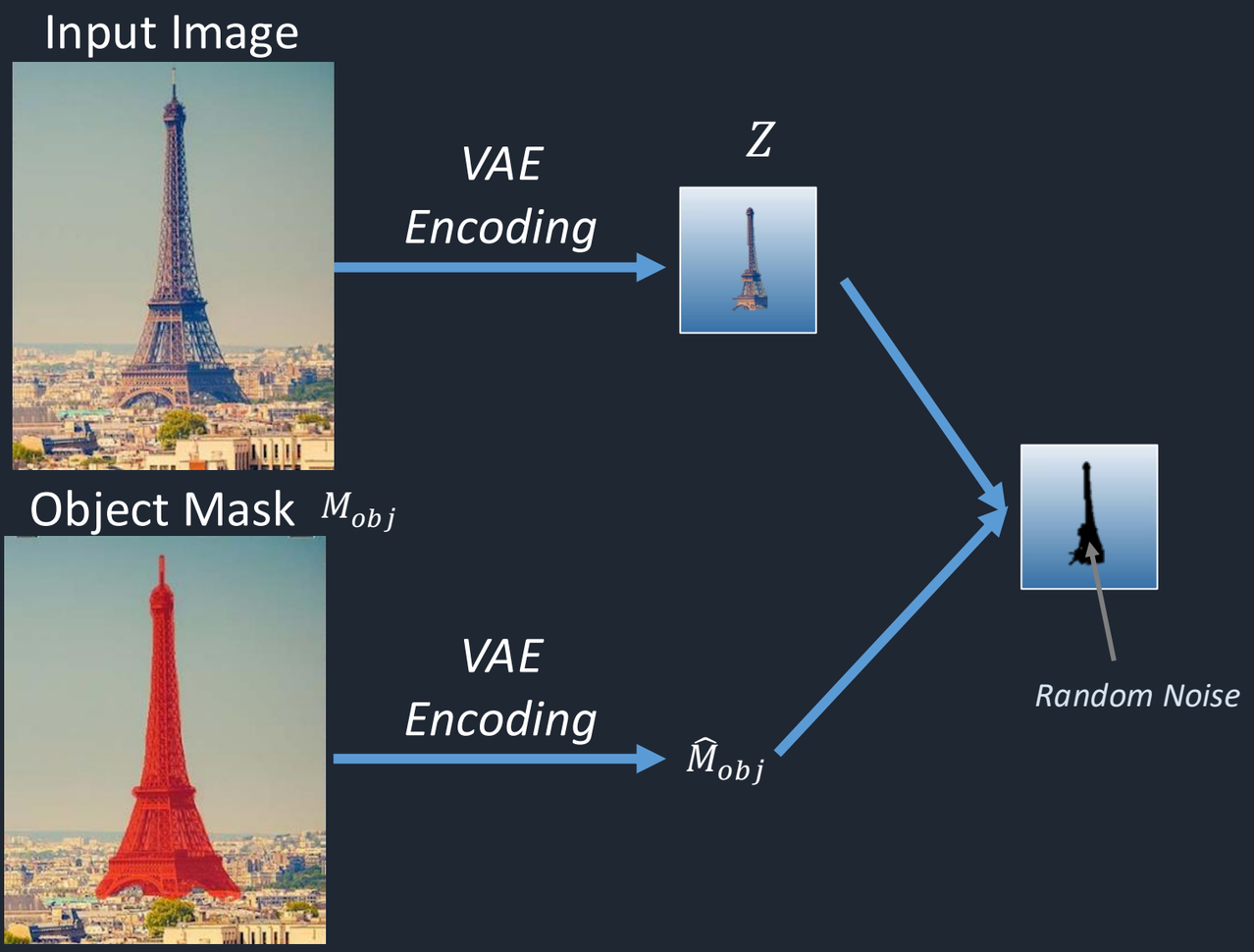
Input Image

Object Mask M_{obj}

Zero-shot Image Inpainting



Zero-shot Image Inpainting



Zero-shot Image Inpainting

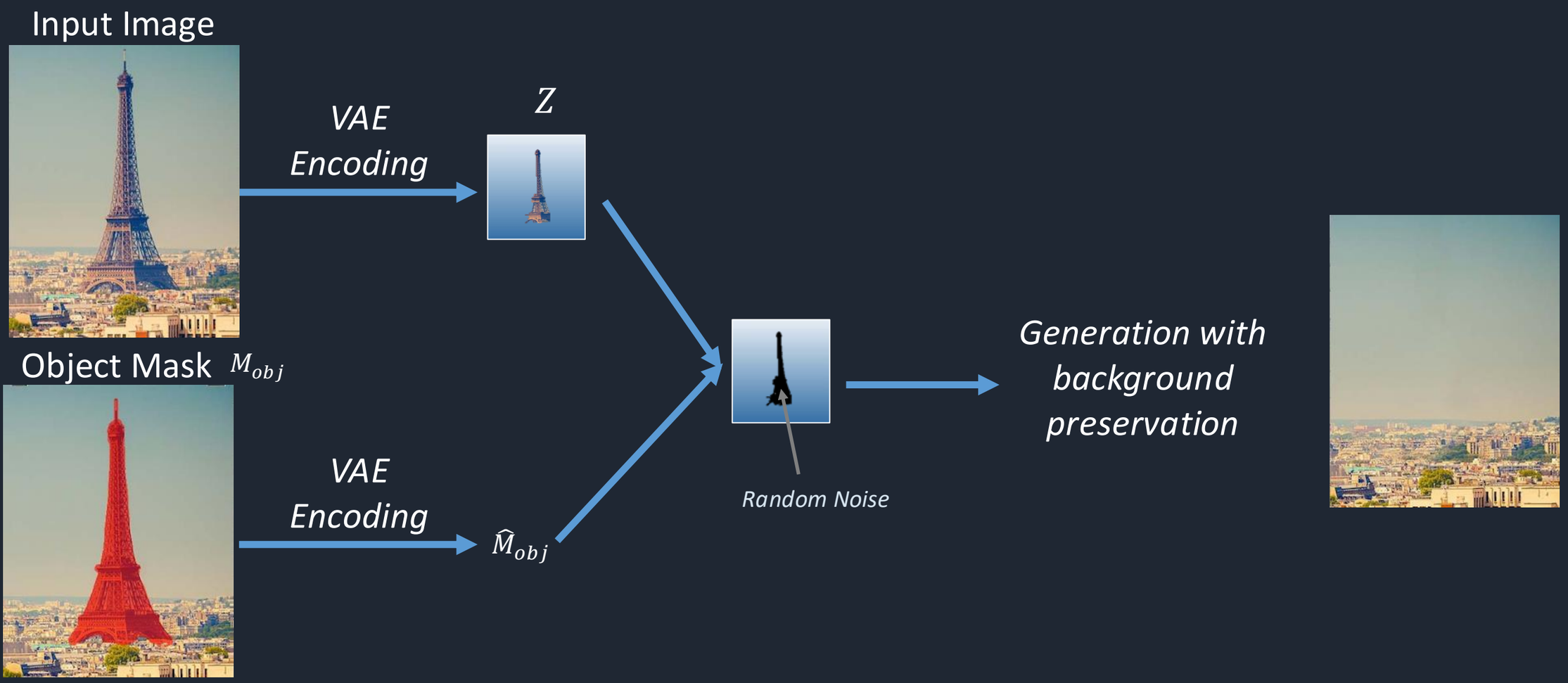


Image Inpainting Fails for Video

Core Differences

- Image inpainting ensures spatial consistency within single frame
- Video inpainting must maintain temporal consistency
- Applying frame-by-frame image-inpainting fails

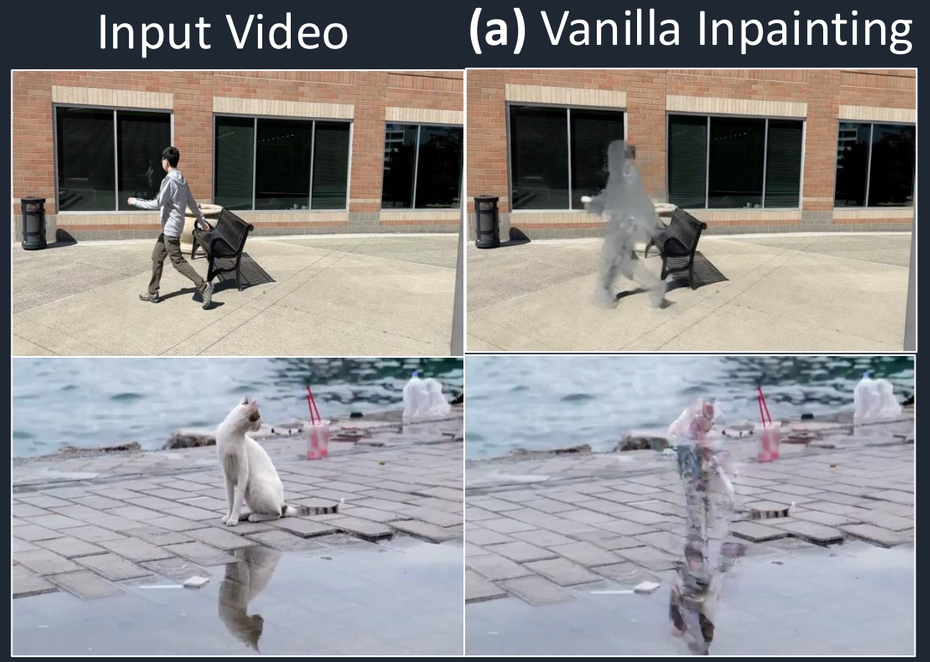
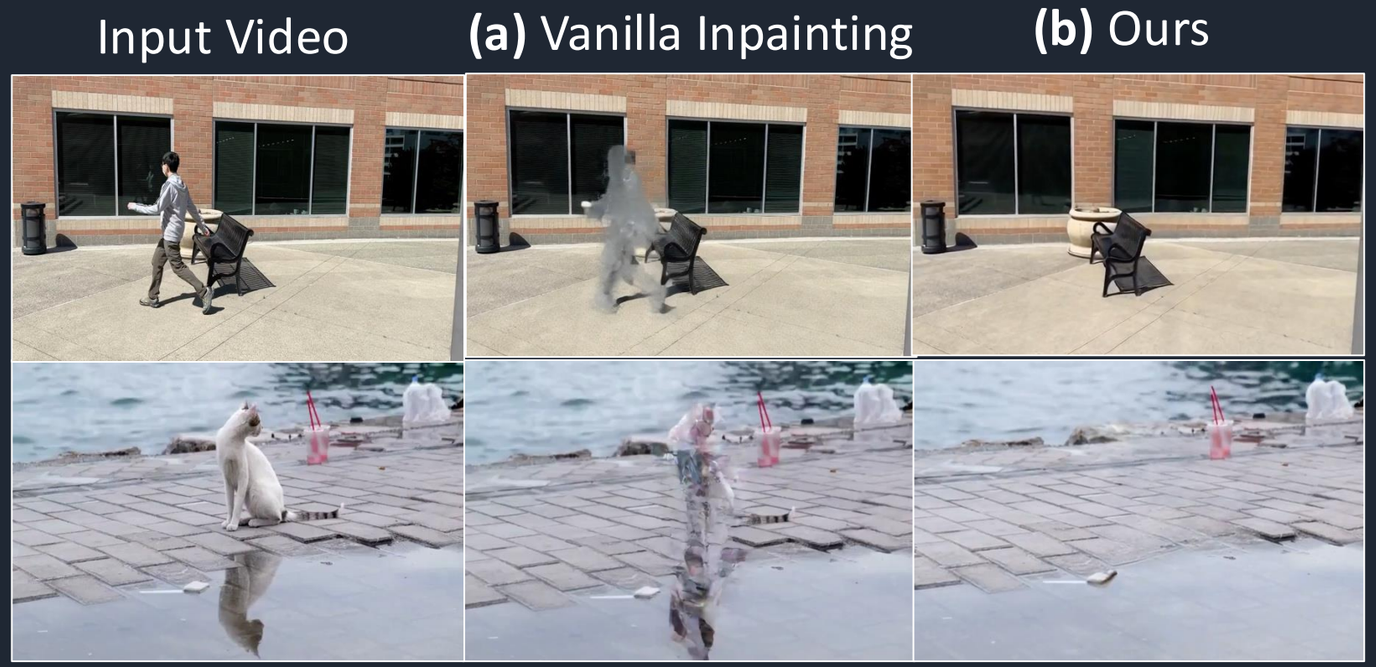


Image Inpainting Fails for Video

Core Differences

- Image inpainting ensures spatial consistency within single frame
- Video inpainting must maintain temporal consistency

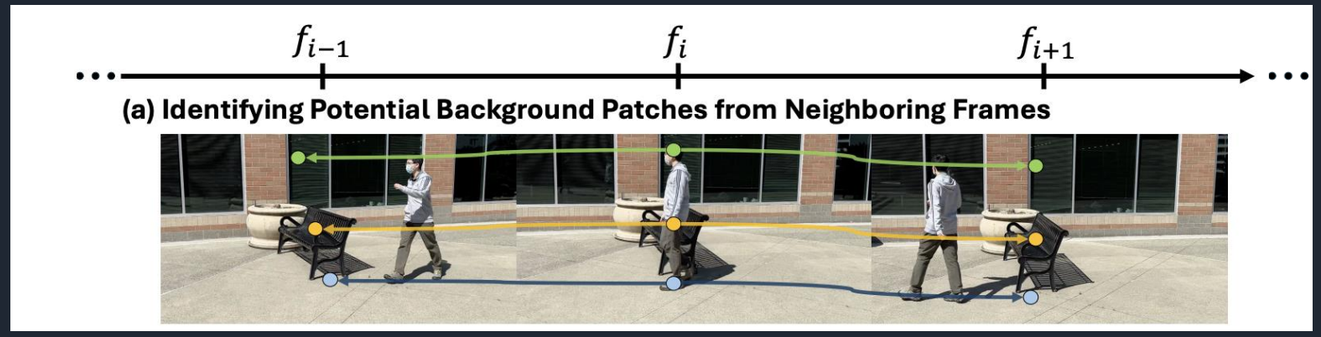
  Applying frame-by-frame image-inpainting fails



Attention Guidance

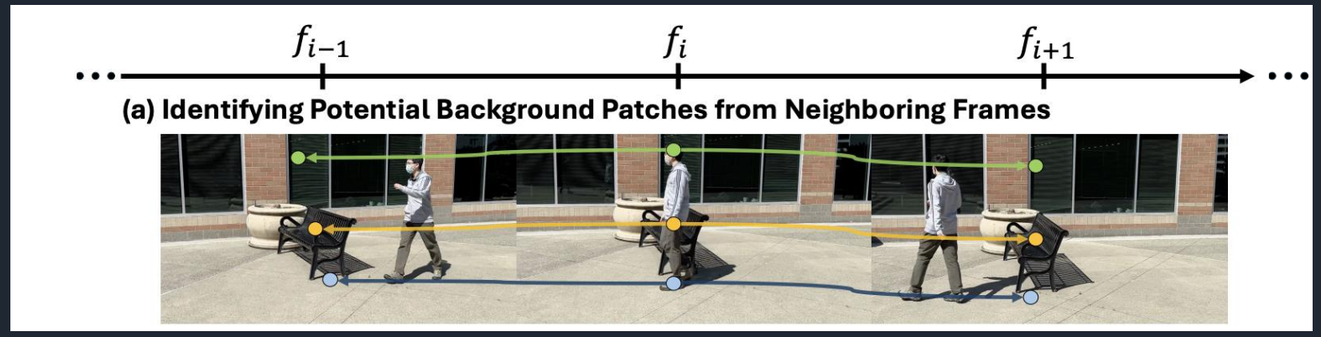
Attention Guidance

(a) Identifies background correspondences across frames using Track-Any-Point (TAP-Net).

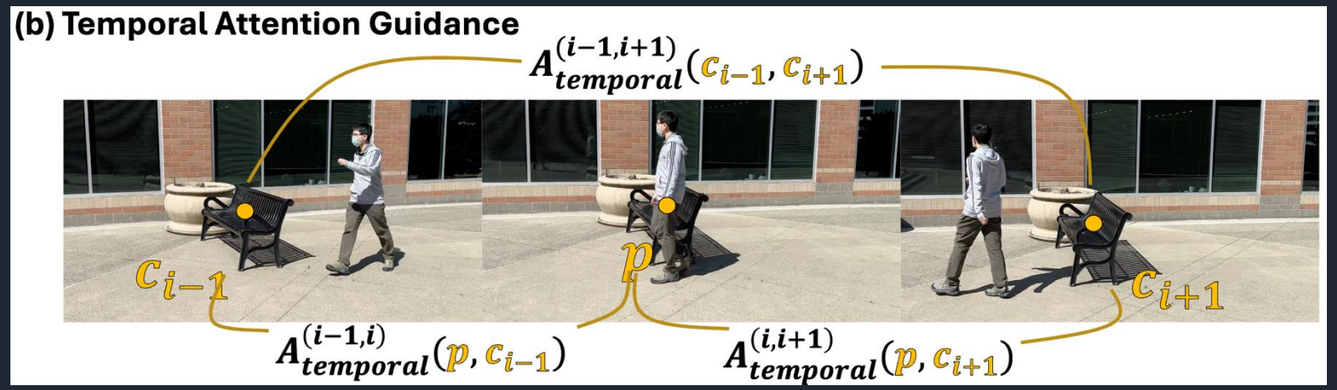


Attention Guidance

(a) Identifies background correspondences across frames using Track-Any-Point (TAP-Net).



(b) **Temporal Attention Guidance** Replaces temporal attention scores with mean attention between background pairs.

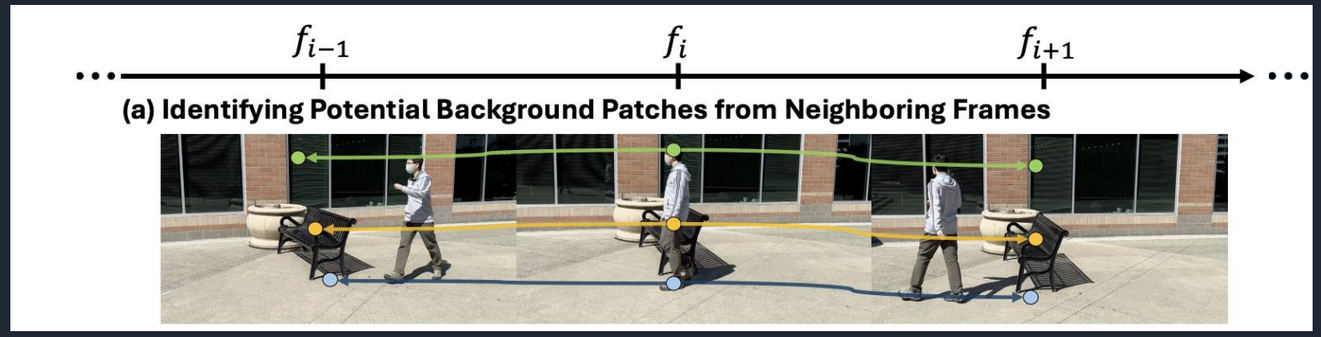


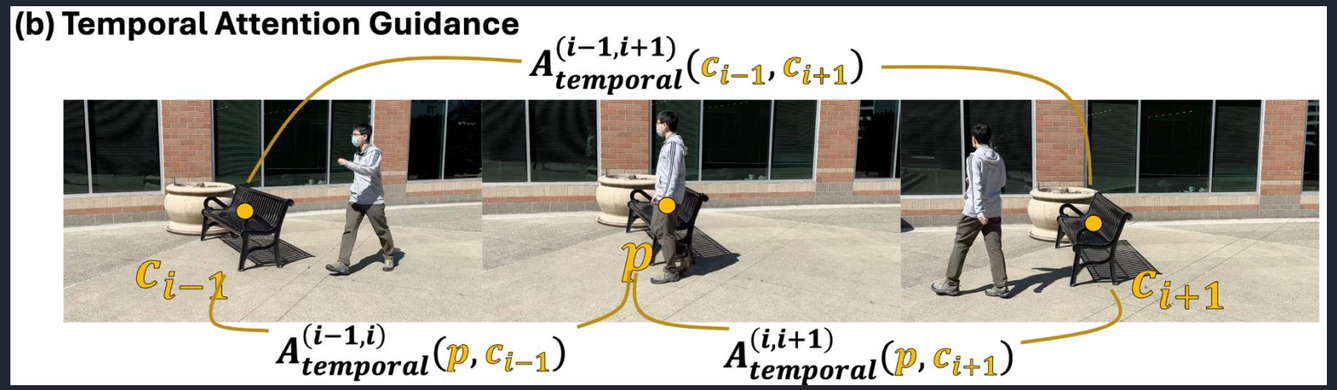
Attention Guidance

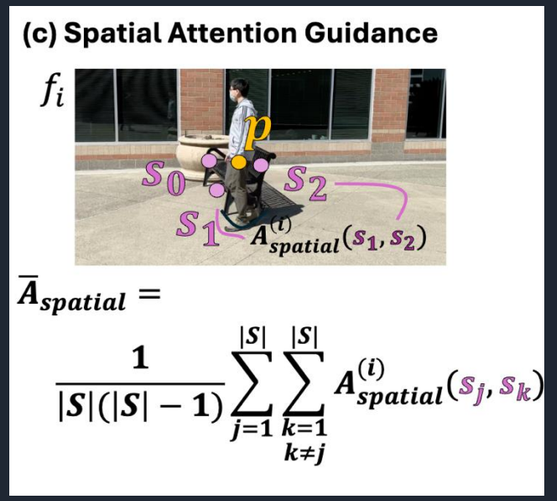
(a) Identifies background correspondences across frames using Track-Any-Point (TAP-Net).

(b) **Temporal Attention Guidance**
Replaces temporal attention scores with mean attention between background pairs.

(c) **Spatial Attention Guidance**
Sets attention scores to mean spatial attention among background points.







Self-Attention Maps Reveal Object Effects

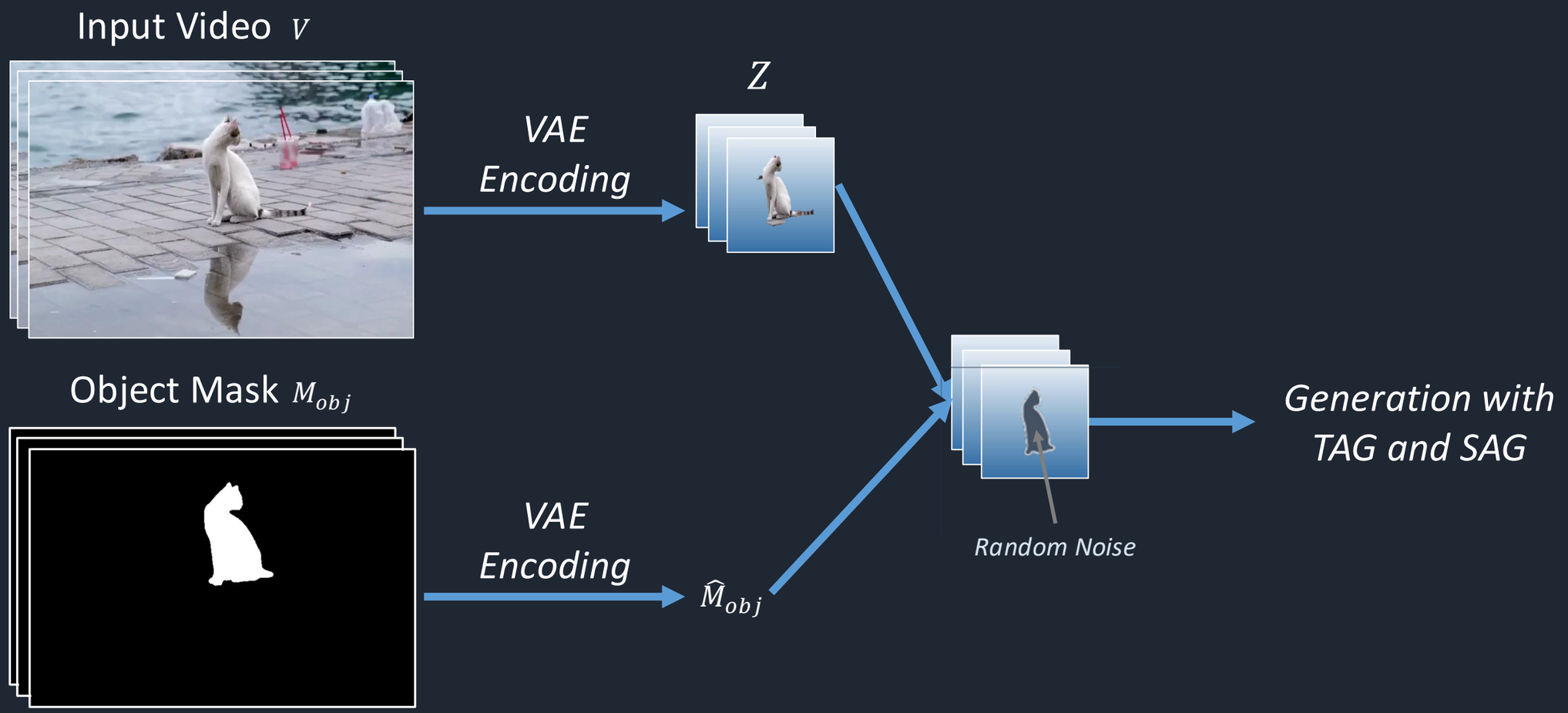
Video diffusion models naturally capture object-effect associations through motion, following **Gestalt psychology's "common fate”.**

All Effect

Automatically detected and included in object masks

Missing effects

OmnimatteZero Overview



Qualitative Results

Quantitative Results

Object Removal

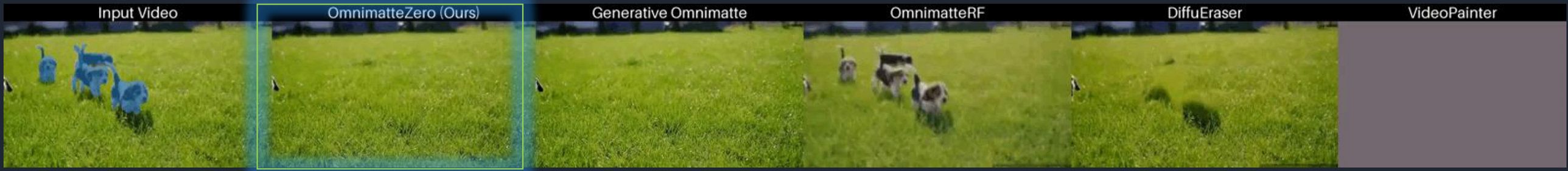
OmnimatteZero outperforms all omnimatte and video inpainting methods.

	Scene	Movie		Kubric		Average			
	Metric	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	T_{train} (hours)	T_{run} (s/frame)
Image Inpainting	ObjectDrop [Winter et al. 2024]	28.05	0.124	34.22	0.083	31.14	0.104	-	-
Video Inpainting	Lumiere inpainting [Bar-Tal et al. 2024]	26.62	0.148	31.46	0.157	29.04	0.153	-	9
	Propainter [Zhou et al. 2023]	27.44	0.114	34.67	0.056	31.06	0.085	-	0.083
	DiffuEraser [Zhou et al. 2023]	29.51	0.105	35.19	0.048	32.35	0.077	-	0.8
Omnimatte	Ominmatte [Lu et al. 2021]	21.76	0.239	26.81	0.207	24.29	0.223	3	2.5
	D2NeRF [Wu et al. 2024]	-	-	34.99	0.113	-	-	3	2.2
	LNA [Kasten et al. 2021]	23.10	0.129	-	-	-	-	8.5	0.4
	OmnimatteRF [Lin et al. 2023]	33.86	0.017	40.91	0.028	37.38	0.023	6	3.5
	Generative Omnimatte [Lee et al. 2024]	32.69	0.030	44.07	0.010	38.38	0.020	-	9
	OmnimatteZero [LTXVideo] (Ours)	**35.11**	**0.014**	44.97	0.010	**40.04**	**0.012**	**0**	**0.04**
	OmnimatteZero [Wan2.1] (Ours)	**34.12**	0.017	**45.55**	**0.006**	**39.84**	**0.011**	**0**	3.2

Real Time!!

Qualitative Results

Object Removal



Foreground Extraction

To isolate an object's layer:

(1) Remove the target object. Results with background video: V_{obj} and Z_{obj}.

(2) Foreground layer can be extracted by subtracting the two latents:

$$Z_{obj} = Z - Z_{bg}$$

Compositions

To seamlessly compose layers, one can simply add the latents to make compositions:

$$Z'_{obj+bg} = Z'_{bg} + Z_{obj}$$

Refinement: To further smooth transitions between composed foreground and background, we apply 3 noising-denoising steps.

Thank You for Listening!

Project Page
& Code