

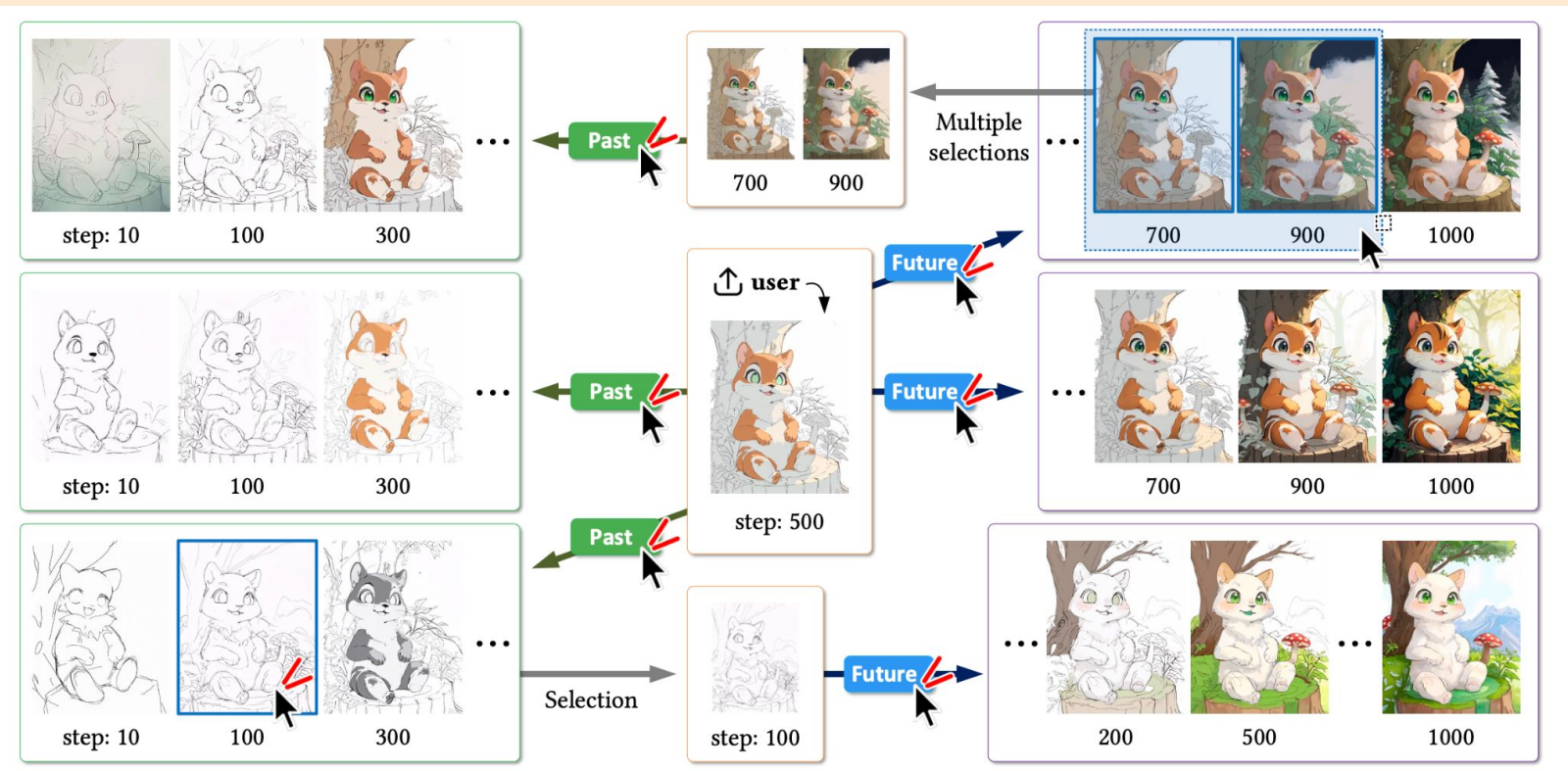


Generating Past and Future in Digital Painting Process

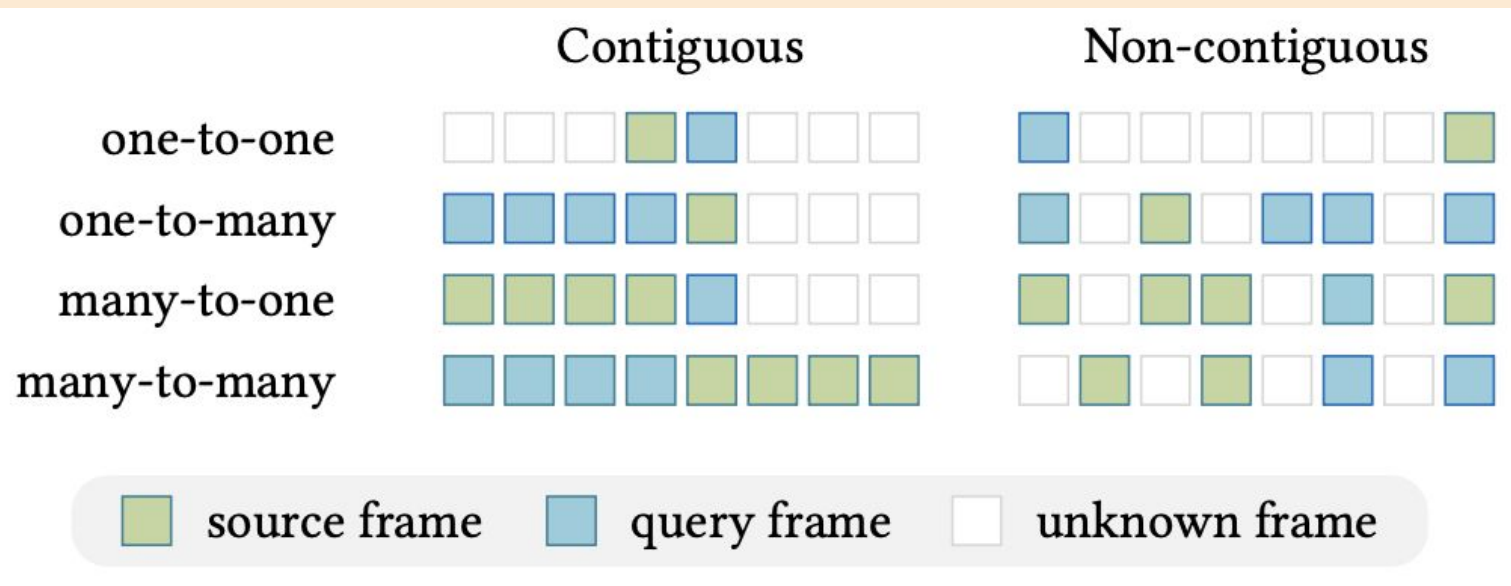
P76134901 林懋



Introduction



Introduction



Related work

Paint undo

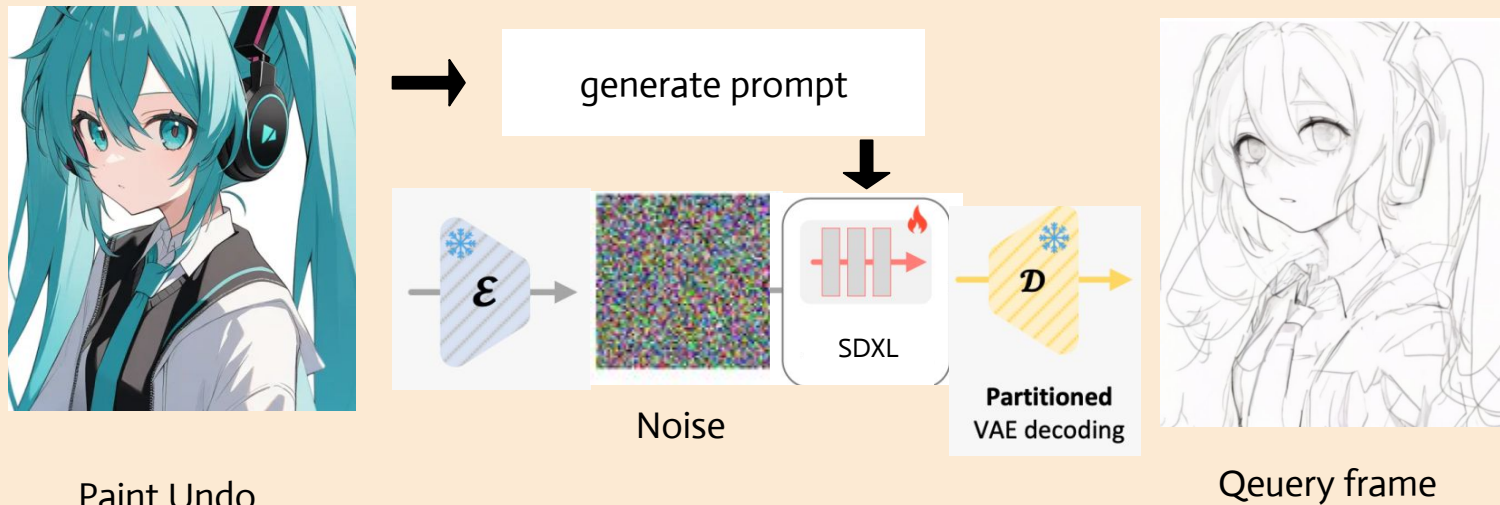


[Demo Link](#)

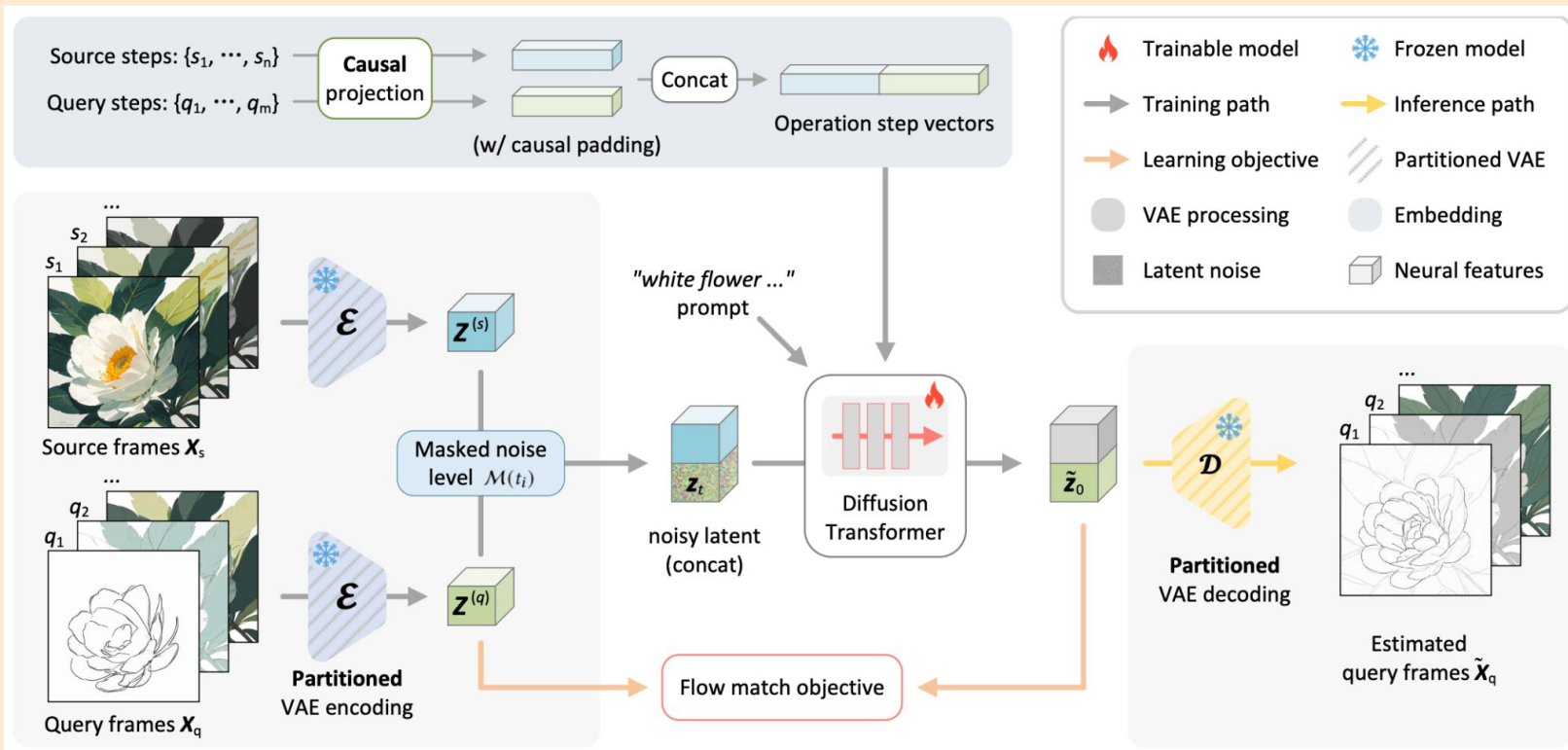
#

Related work

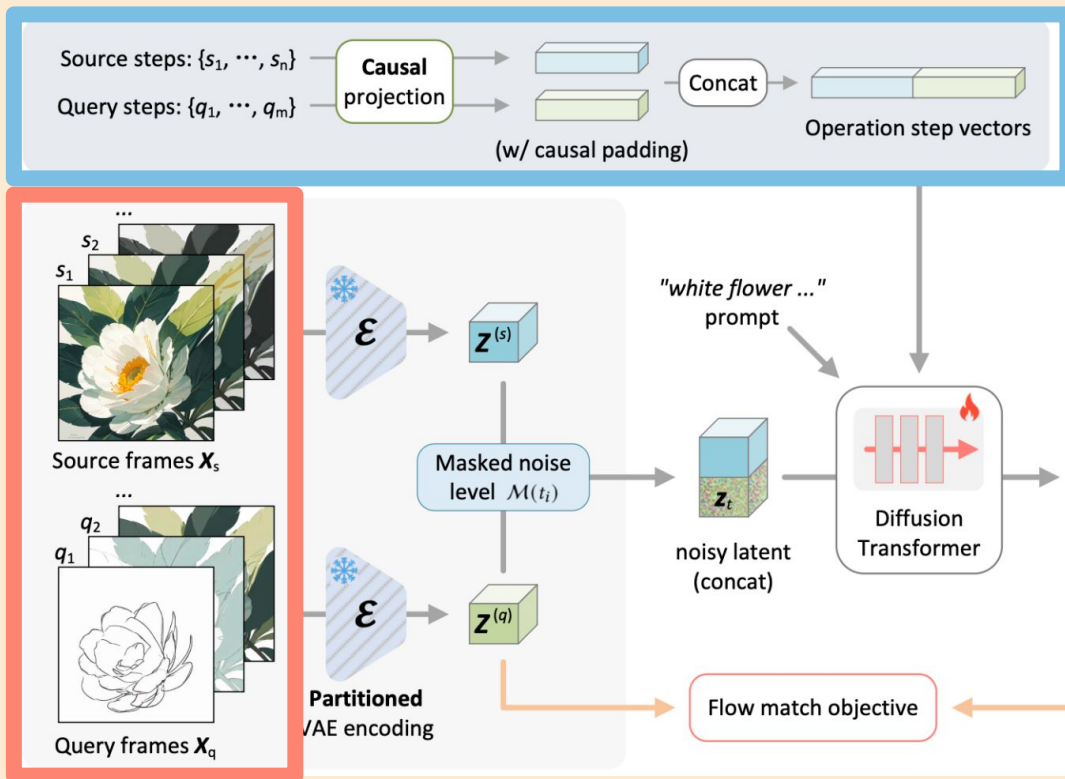
Paint undo



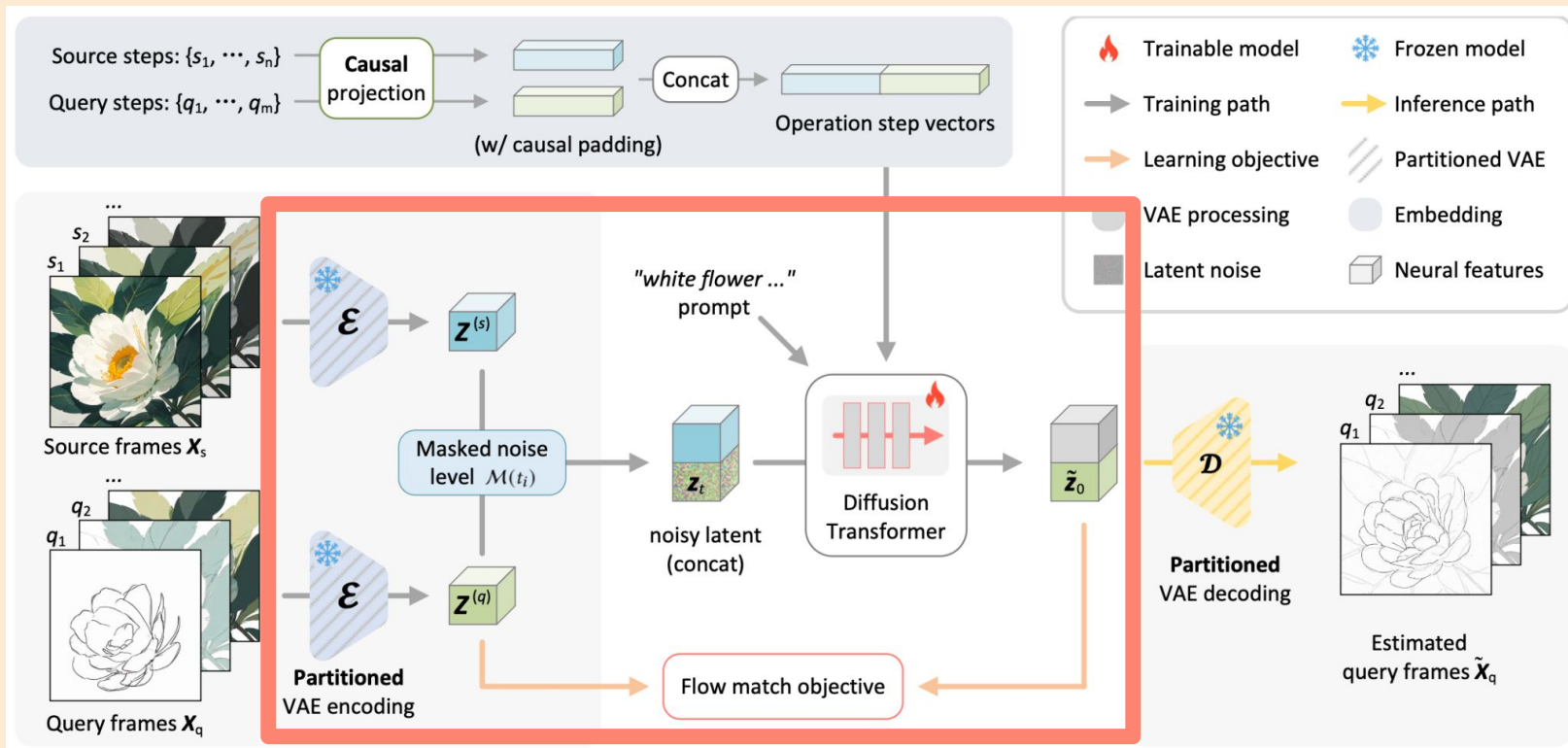
Method



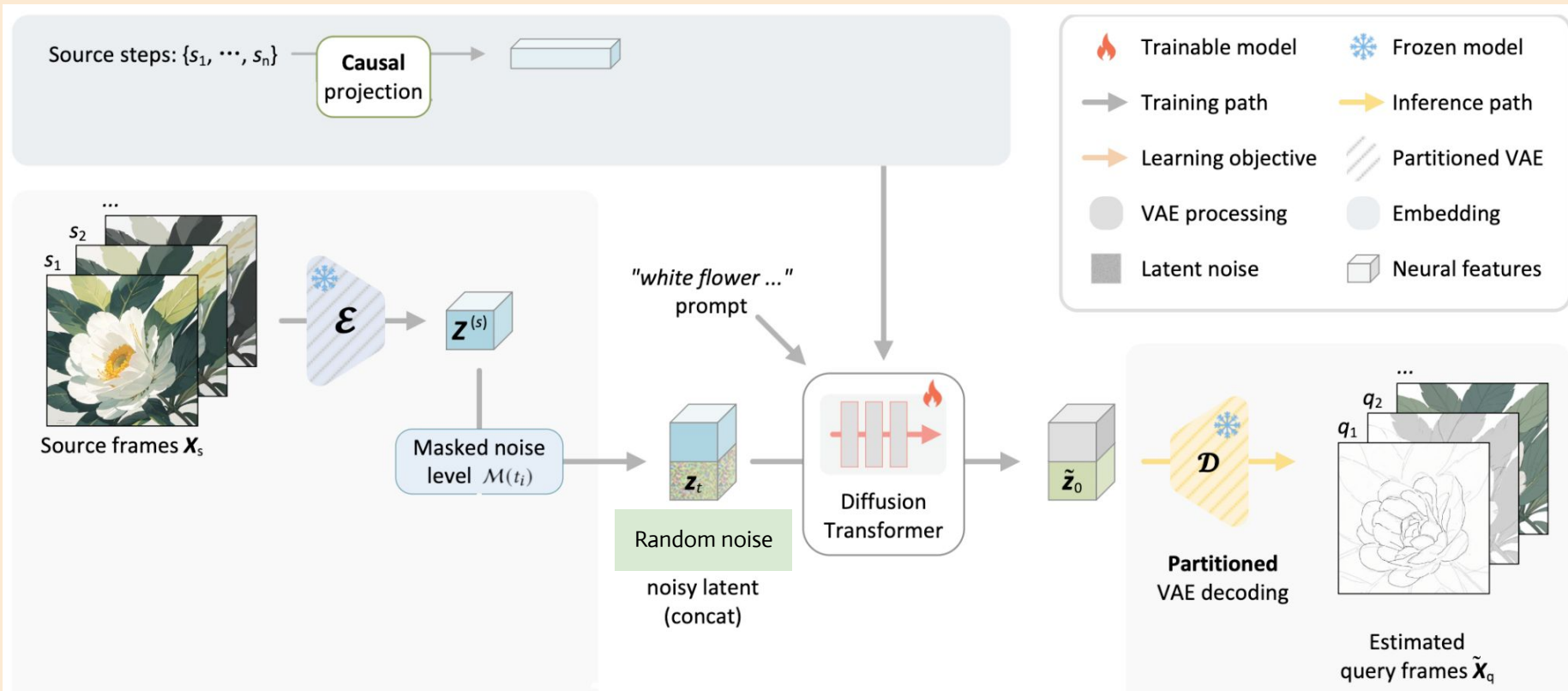
Training input



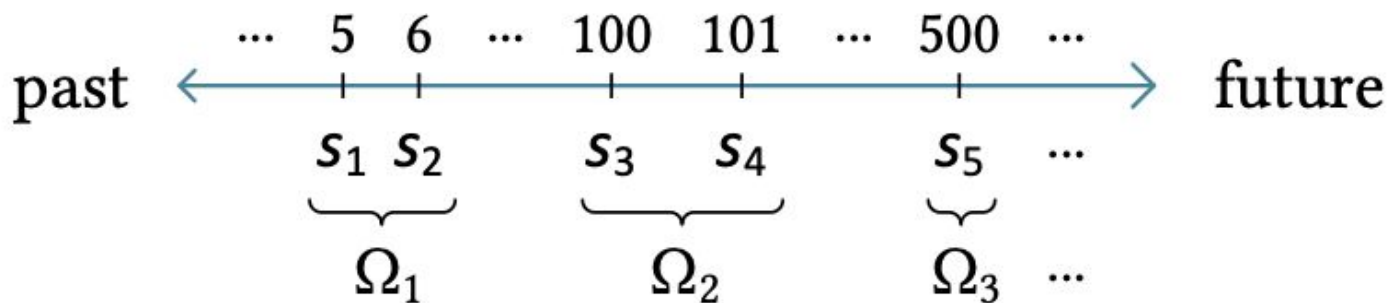
Training



Inference



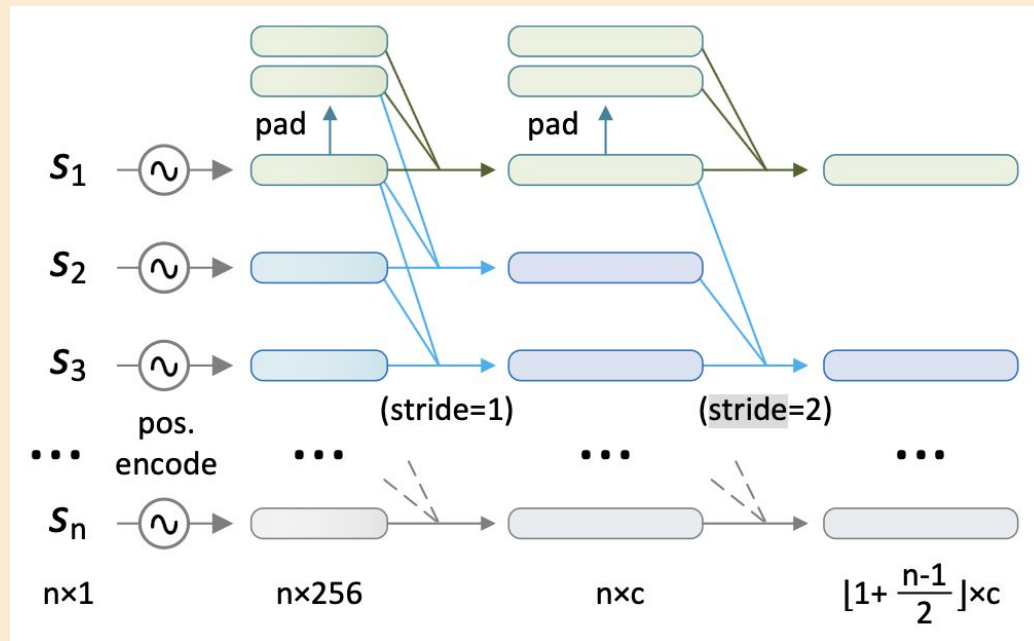
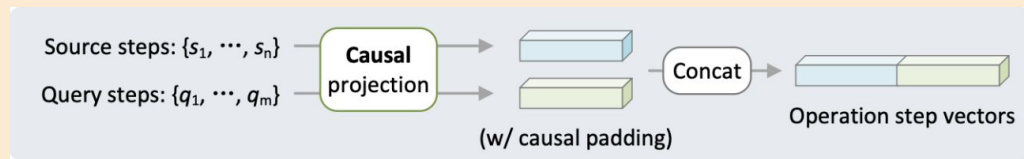
Partitioned 3D VAE



影片擴散框架通常使用具有因果卷積的 3D VAE來對多個幀進行編碼。這種編碼允許在重建多個幀時實現時間一致性

但非連續幀時會出現不必要的重影偽影

Causaul Padding



Datas

- **200 videos** : 483 hours, 35k objects
- Total **20k** Training set
- Use Ltx standard ignore inconspicuous frames
- Use Training rate $1e-5$ Adafactor [Shazeer and Stern 2018]

Devices

- **8x H100 80G**
- LTXVideo : 4 days
- CogVideoX : 9 days
- SDXL : 3 days

EXPERIMENTS

Method	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
[Zhao et al. 2020]	12.27	0.4847	0.6851
[Song et al. 2024]	13.52	0.4901	0.6143
Ours (LTXVideo)	16.31	0.6510	0.4259
Ours (CogVideo)	17.04	0.6712	0.4022
Ours (image model, SDXL)	15.98	0.5995	0.4515

EXPERIMENTS - Random seeds



5% ($s = 5$)



20% ($s = 200$)



40% ($s = 400$)



60% ($s = 600$)



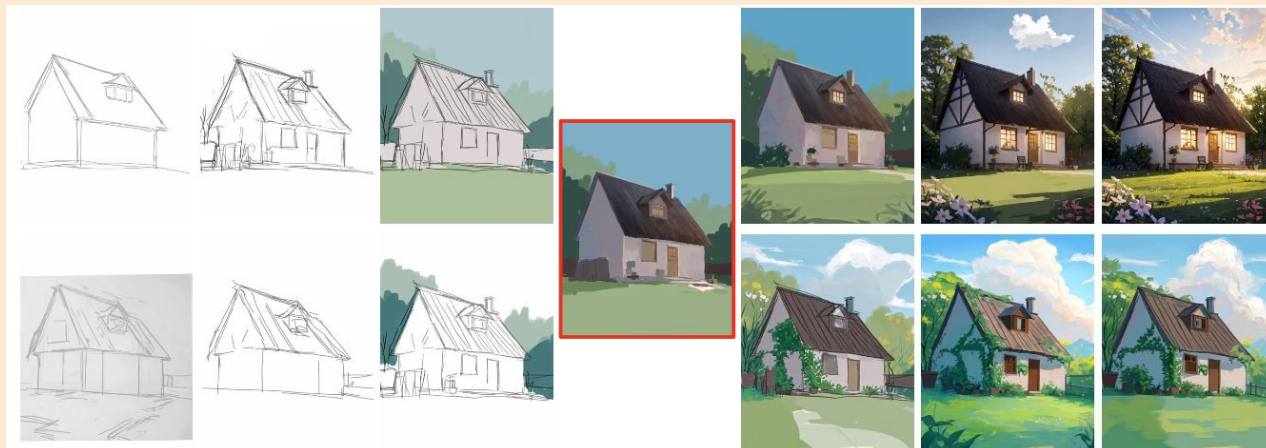
75% ($s = 750$)



90% ($s = 900$)



100% ($s = 1000$)



EXPERIMENTS

ProcessPainter:



InversePainting:



Ours:



EXPERIMENTS

Candidate	Quality $\uparrow$	Responsiveness $\uparrow$
[Zhao et al. 2020]	$1.20 \pm 0.0\%$	/
[Song et al. 2024]	$3.61 \pm 1.21\%$	/
Ours (LTXVideo)	$18.07 \pm 4.3\%$	$27.49 \pm 9.7\%$
Ours (CogVideo)	$20.48 \pm 5.2\%$	$8.46 \pm 6.5\%$
Ours (LTXVideo + image model)	$26.51 \pm 6.1\%$	$47.43 \pm 9.1\%$
Ours (CogVideo + image model)	$30.12 \pm 5.8\%$	$16.62 \pm 5.2\%$

EXPERIMENTS : ablation experimanet

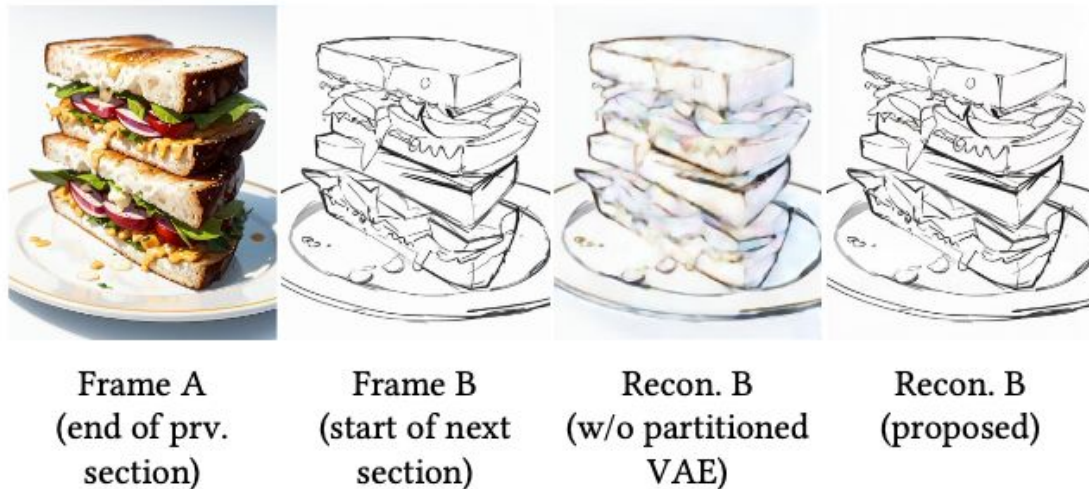


Fig. 9. **Influence of Partitioned VAE.** We show two connected sections with non-contiguous frames and the VAE reconstruction of second frame. Without the Partitioned VAE design, the second frame will have blurred reconstruction caused by latent temporal compression.

EXPERIMENTS : limitation

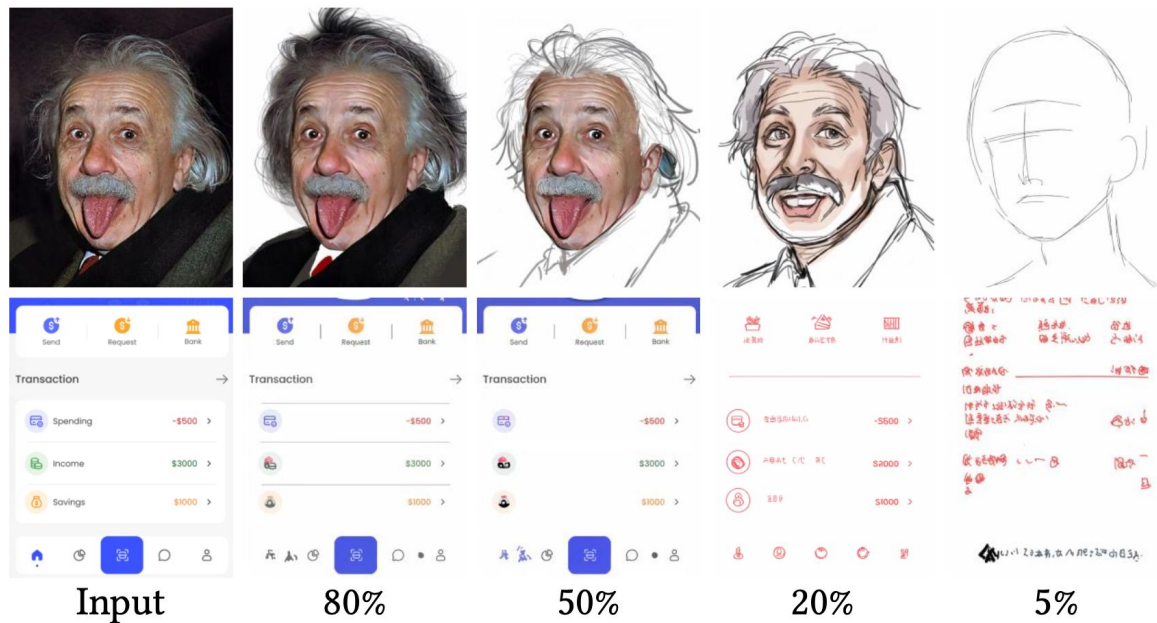


Fig. 13. **Limitation.** When processing images with styles too different to pretrained paintings, the process may have style offsets. When processing out-of-scope images like UI designs, errors will occur with text and structure.