

NaRCan: Natural Refined Canonical Image with Integration of Diffusion Prior for Video Editing

Ting-Hsuan Chen, Jiewen Chan, Hau-Shiang Shiu, Shih-Han Yen, Chang-Han Yeh, Yu-Lun Liu

National Yang Ming Chiao Tung University



國立陽明交通大學

NATIONAL YANG MING CHIAO TUNG UNIVERSITY

Outline

- 📌 Video editing with canonical image
- 📌 Introduction
- 📌 Framework
- 📌 Noise and diffusion prior scheduling
- 📌 Separated NaRCan
- 📌 Experimental results
- 📌 Ablation studies

Video editing with canonical image

- **Video representation with canonical image:** Canonical-based methods simplify video editing by consolidating content into a single image, allowing precise spatial control and maintaining temporal consistency.
- **Main challenge:** Current canonical-based methods DO NOT stipulate that the canonical image must be a natural image, which complicates downstream video editing tasks.

Video editing with canonical image

- **Current method:** CoDeF fails to generate the natural canonical image, resulting in artifacts in the video.



Canonical image

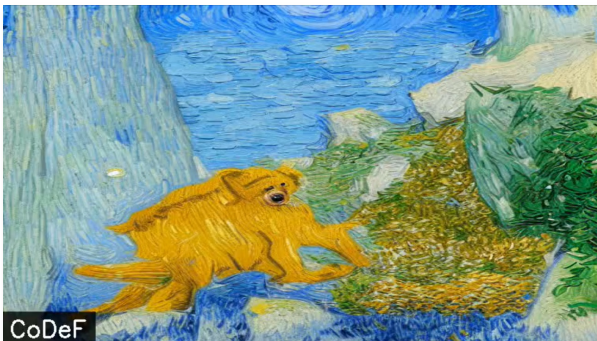


Canonical image



Canonical image

Reconstruct after editing



Style transfer



Video editing



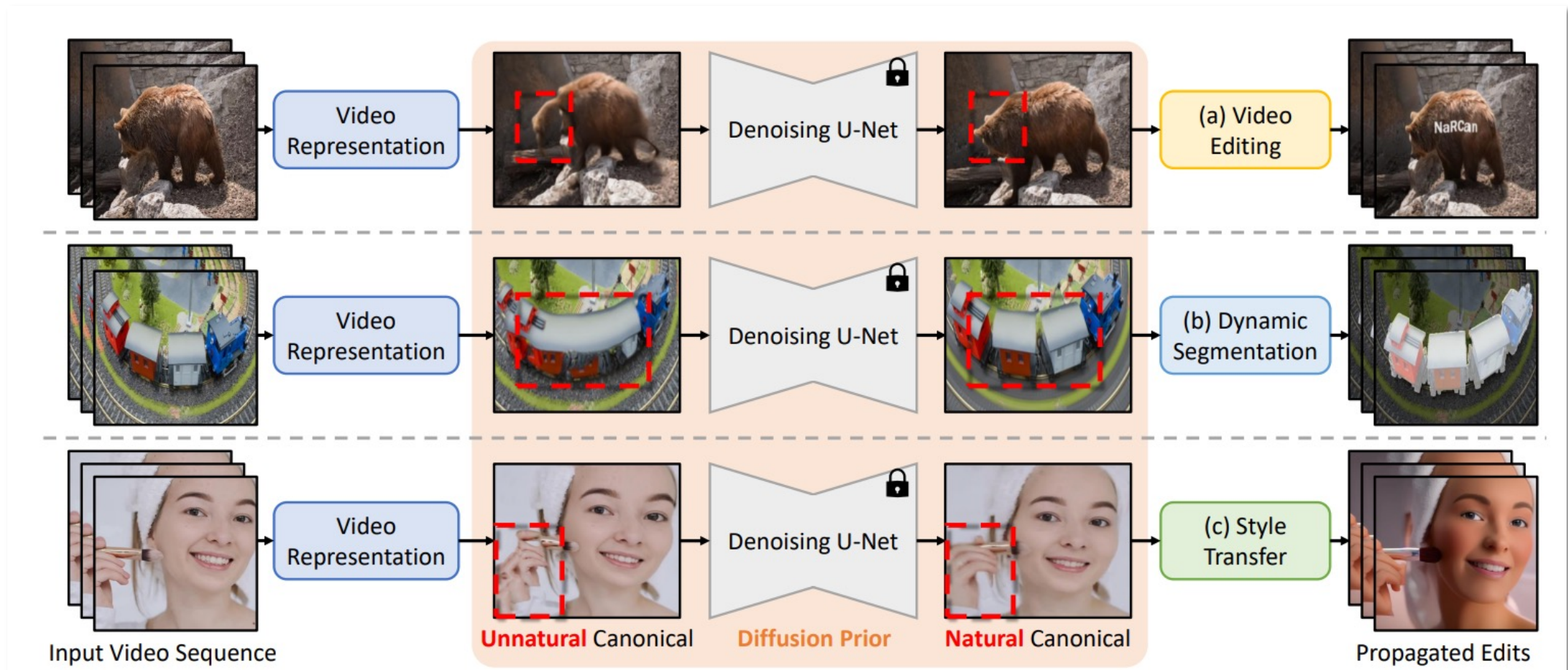
Dynamic segmentation

Outline

- 🍯 Video editing with canonical image
- 🍯 Introduction**
- 🍯 Framework
- 🍯 Noise and diffusion prior scheduling
- 🍯 Separated NaRCan
- 🍯 Experimental results
- 🍯 Ablation studies

Introduction

- Video representation with diffusion prior.

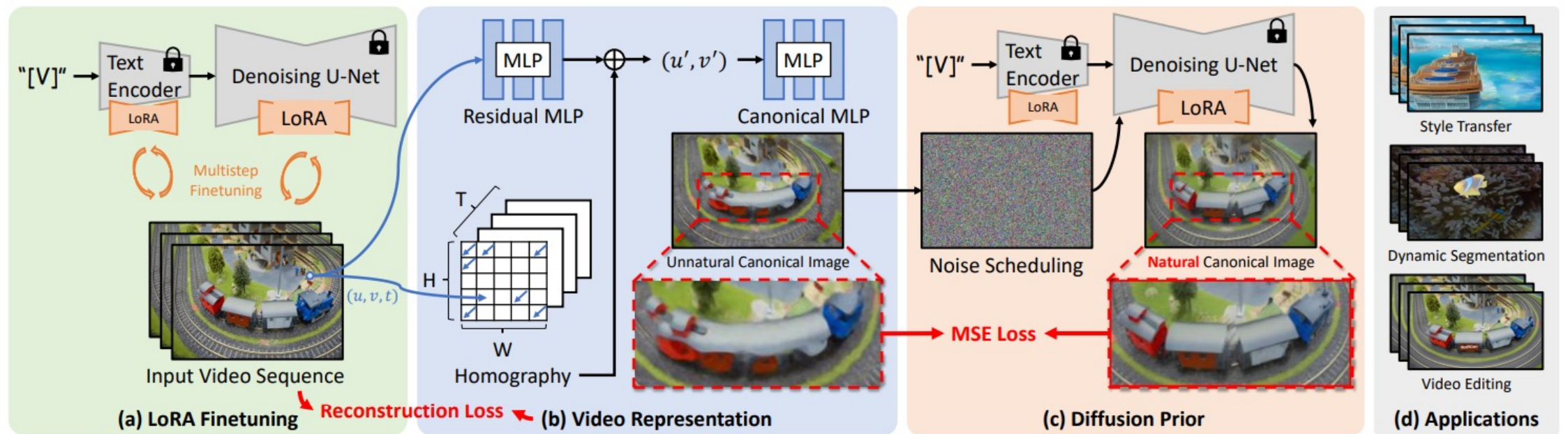


Outline

- 📌 Video editing with canonical image
- 📌 Introduction
- 📌 **Framework**
- 📌 Noise and diffusion prior scheduling
- 📌 Separated NaRCan
- 📌 Experimental results
- 📌 Ablation studies

Framework

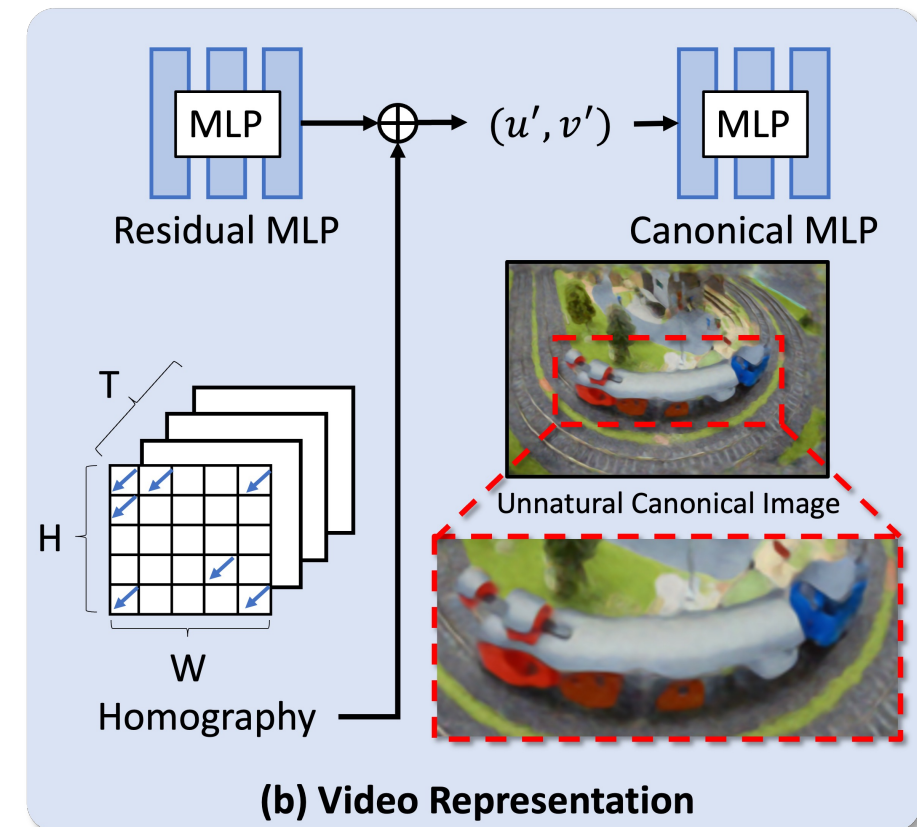
- **Overview:** Our method aims to represent an input video sequence with a **natural** canonical image, crucial for versatile downstream applications.



Framework

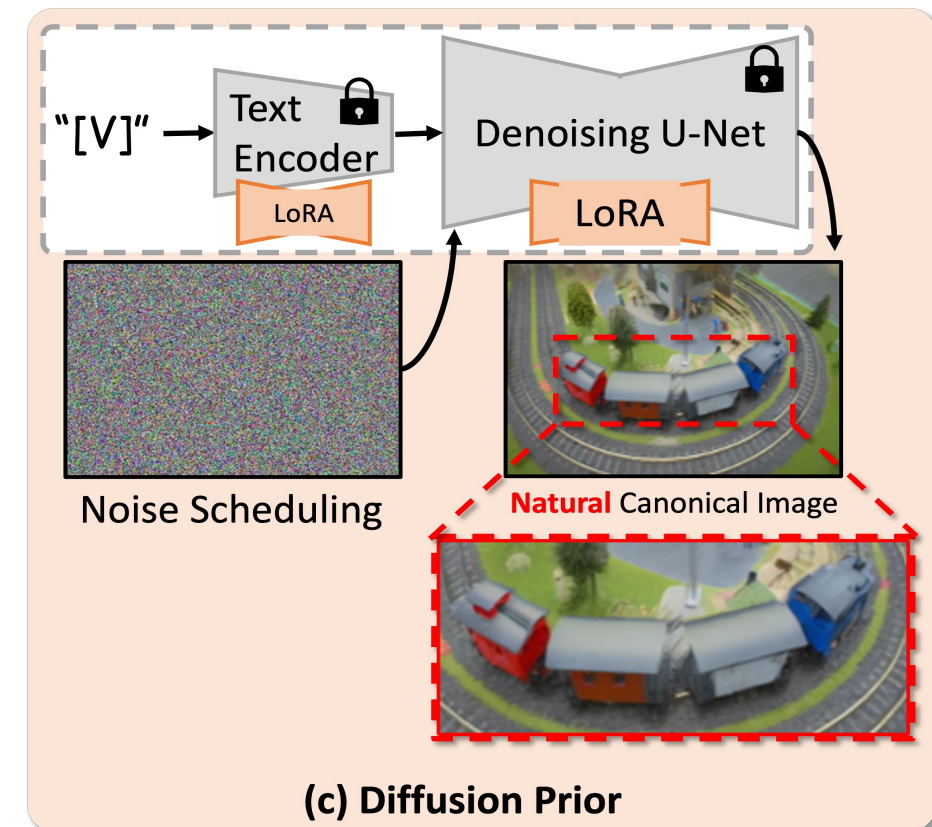
- **Video representation with hybrid deformation field and canonical field.**

- Deformation field: Formed by the combination of **Homography** and **Residual MLP**.
- Canonical field: The unnatural canonical image will be regularized and corrected by **diffusion prior**.



Framework

- **Diffusion prior for canonical image refinement.**
 - Latent diffusion model: We introduce diffusion priors, which successfully generate **Natural** Canonical image.
 - LoRA fine-tune: Ensures that the diffusion model generates high-quality **natural** canonical images **tailored to** the testing sequence.

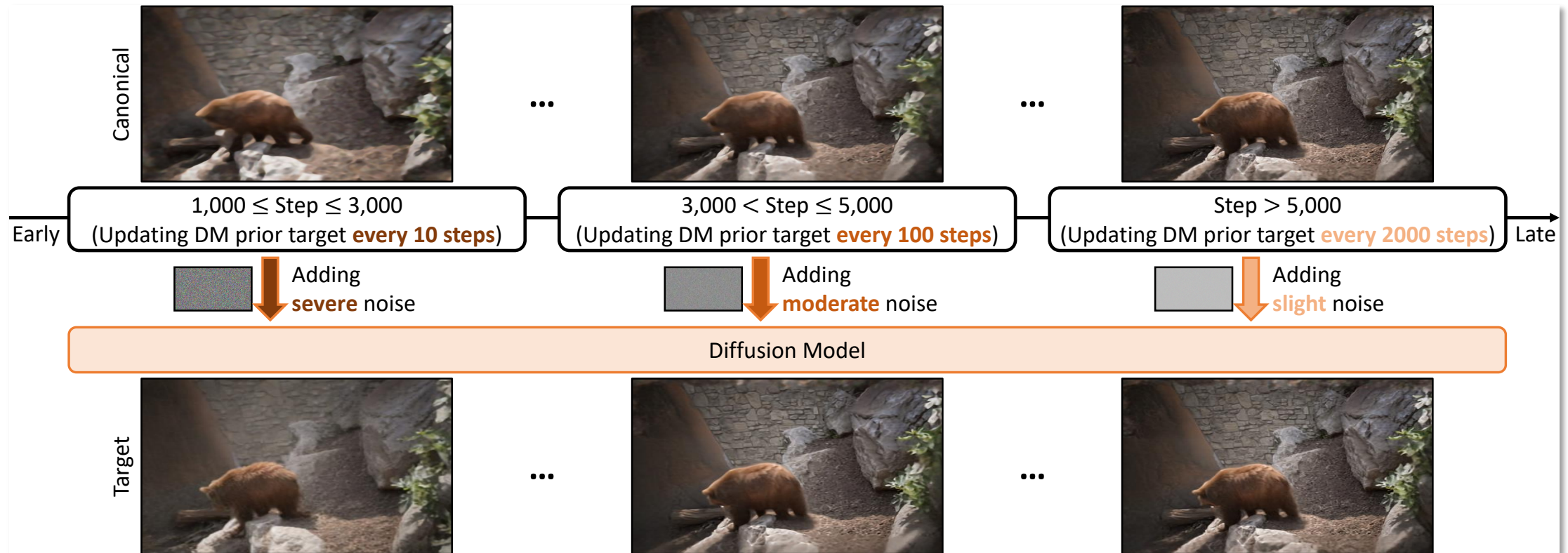


Outline

- 📌 Video editing with canonical image
- 📌 Introduction
- 📌 Framework
- 📌 Noise and diffusion prior scheduling**
- 📌 Separated NaRCan
- 📌 Experimental results
- 📌 Ablation studies

Noise and diffusion prior update scheduling

- **Hierarchical scheduling:** Ensures the final canonical image matches per-step update quality, accelerating training by **14 times** (from 4.8 hours to 20 minutes).

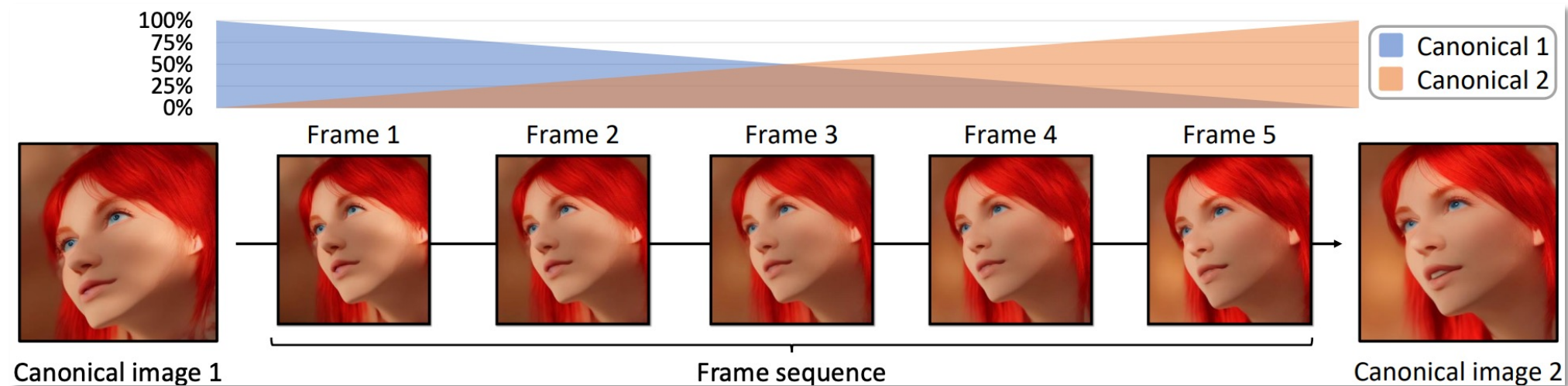


Outline

- 📌 Video editing with canonical image
- 📌 Introduction
- 📌 Framework
- 📌 Noise and diffusion prior scheduling
- 📌 **Separated NaRCan**
- 📌 Experimental results
- 📌 Ablation studies

Separated NaRCan

- **Challenge:** Relying on a single natural canonical image to represent overly complex scenes is impractical and unrealistic.
- **Video segmentation and canonical image interpolation.**

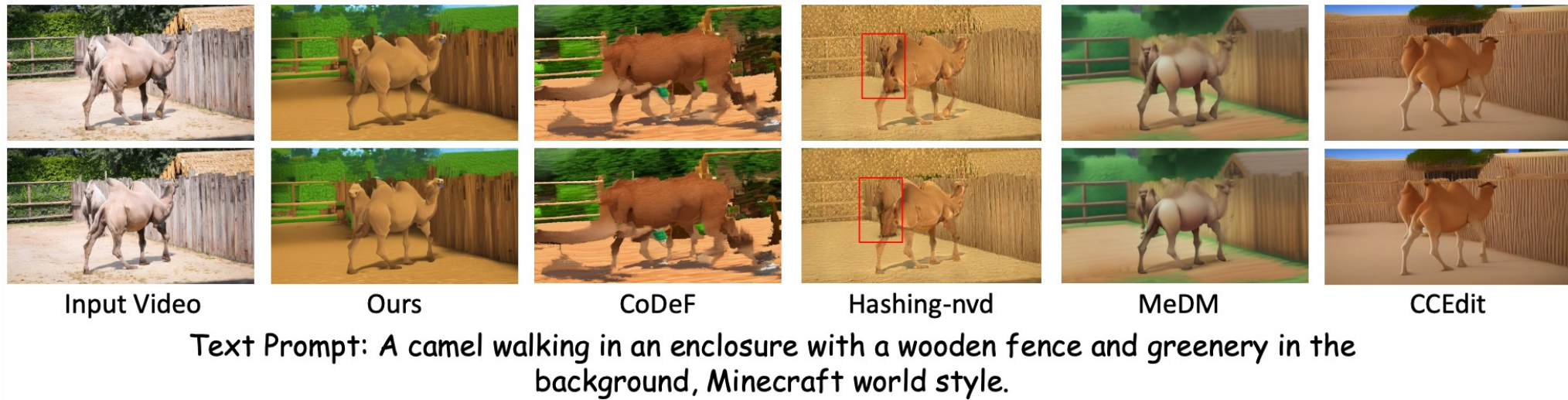


Outline

- 🍯 Video editing with canonical image
- 🍯 Introduction
- 🍯 Framework
- 🍯 Noise and diffusion prior scheduling
- 🍯 Separated NaRCan
- 🍯 **Experimental results**
- 🍯 Ablation studies

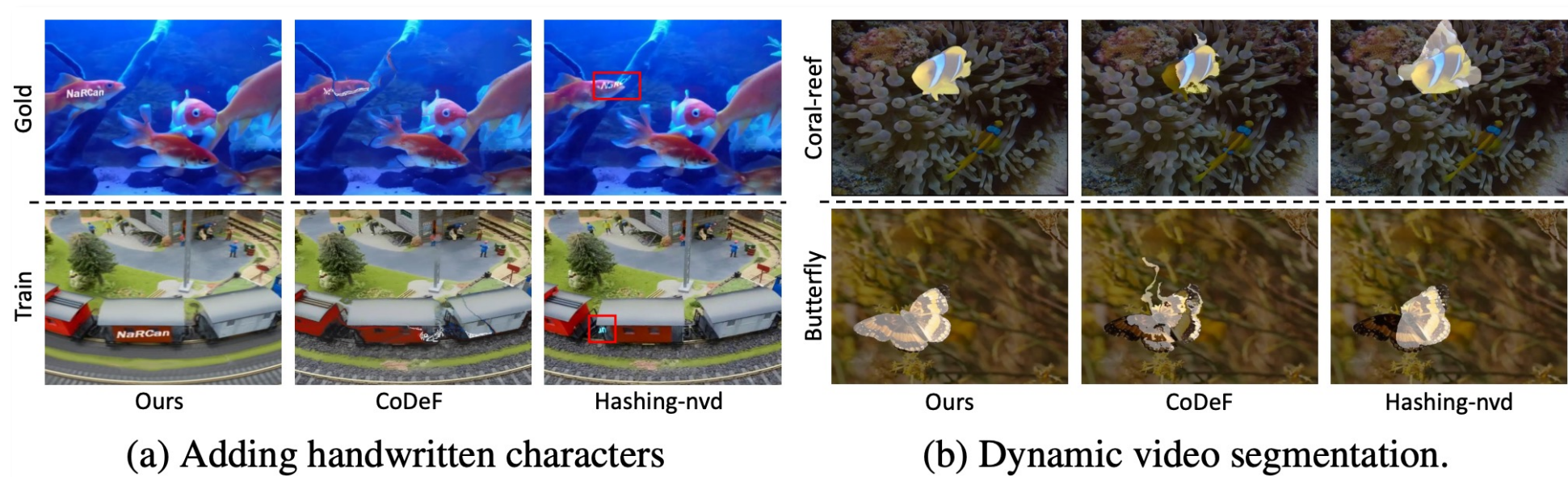
Experimental results

- **Video style transformation.**



Experimental results

- Video editing (a) and segmentation (b).

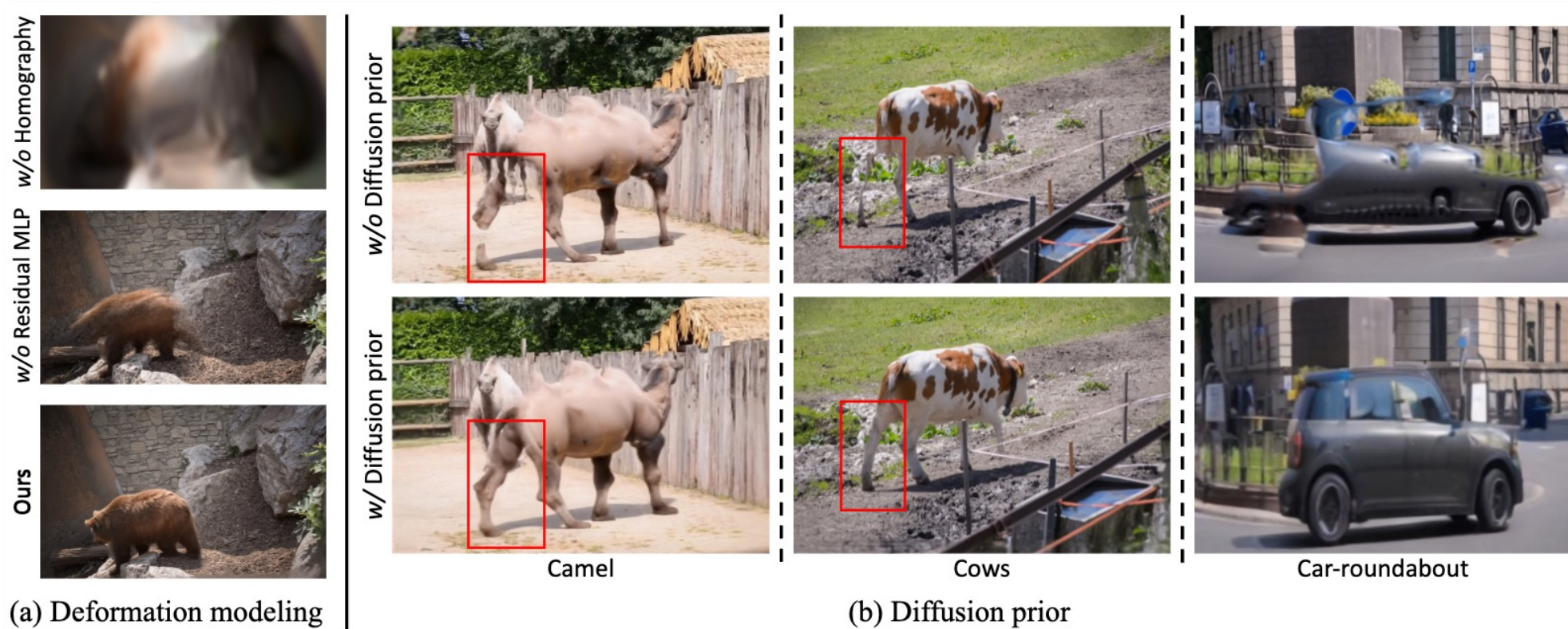


Outline

- 🧪 Video editing with canonical image
- 🧪 Introduction
- 🧪 Framework
- 🧪 Noise and diffusion prior scheduling
- 🧪 Separated NaRCan
- 🧪 Experimental results
- 🧪 Ablation studies

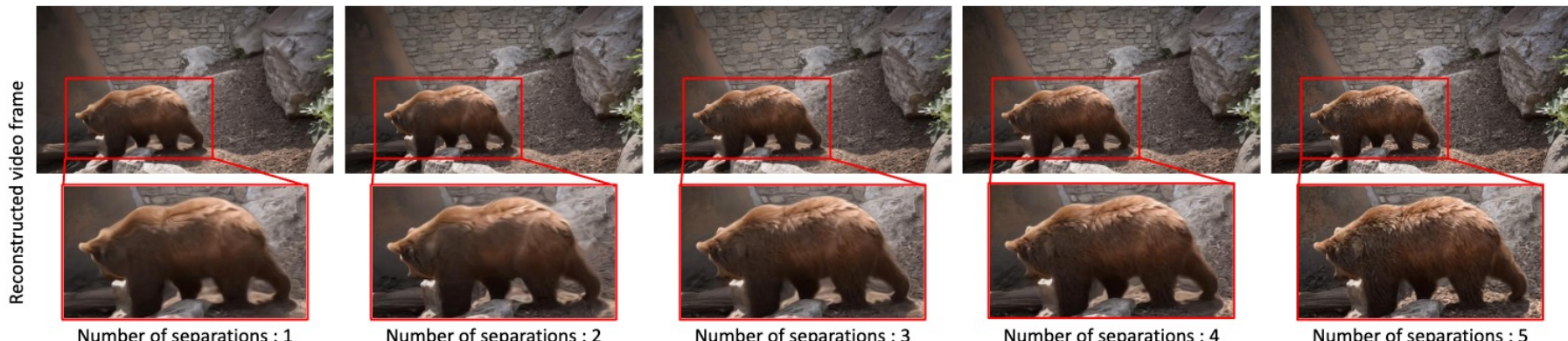
Ablation studies

- Homography, Residual Deformation MLP and Diffusion prior.



Ablation studies

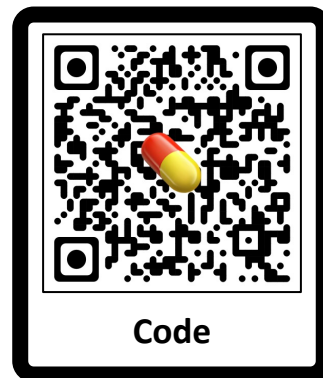
- Trade-off between reconstruction quality and temporal consistency with varying separations.



N of separations	Training time(s)	PSNR \uparrow	SSIM \uparrow	short-term $E_{warp} \downarrow$	long-term $E_{warp} \downarrow$
1	771.50	23.814	0.6369	0.0016	0.0304
2	1530.79	24.423	0.6762	0.0019	0.0310
3	2275.20	24.852	0.7017	0.0021	0.0312
4	3015.32	25.185	0.7191	0.0022	0.0326
5	3761.40	25.398	0.7301	0.0022	0.0334



Thanks for your attention!



National Yang Ming Chiao Tung University
Computational Photography Lab