

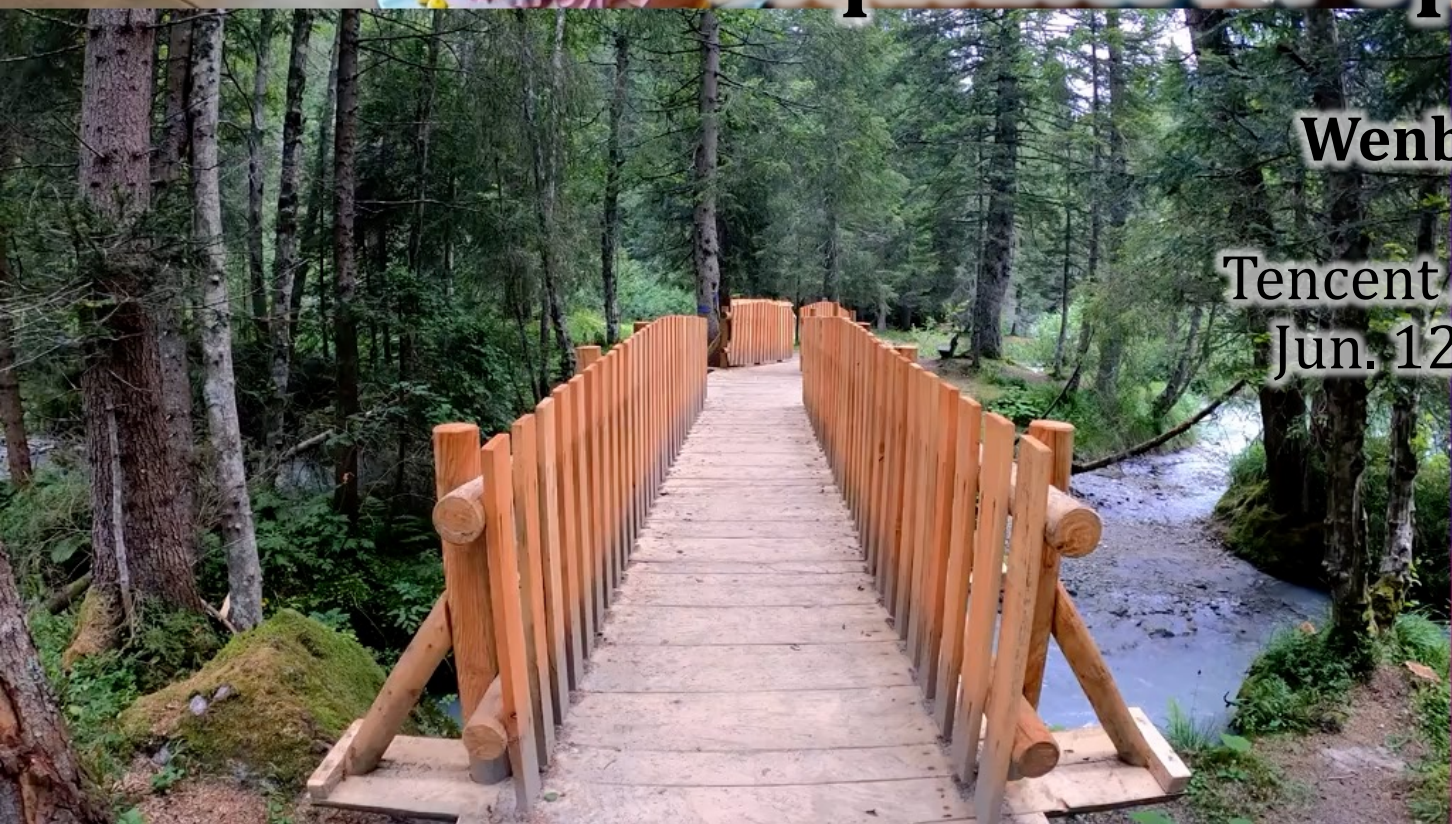


# DepthCrafter: Generating Consistent Long Depth Sequences for Open-world Videos

Wenbo Hu

Tencent ARC Lab

Jun. 12, 2025





# Background

- **Depth estimation**, serving as the **bridge**, links **2D observations** and the **3D world**



- Challenges in open-world video depth estimation:
  - ❑ Generalize to open-world videos
  - ❑ Accurate and consistent depth
  - ❑ Long videos

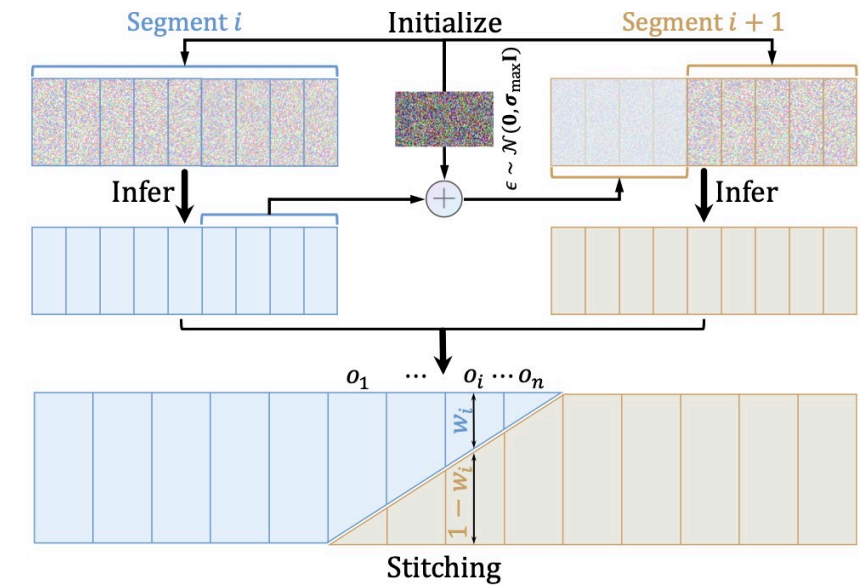
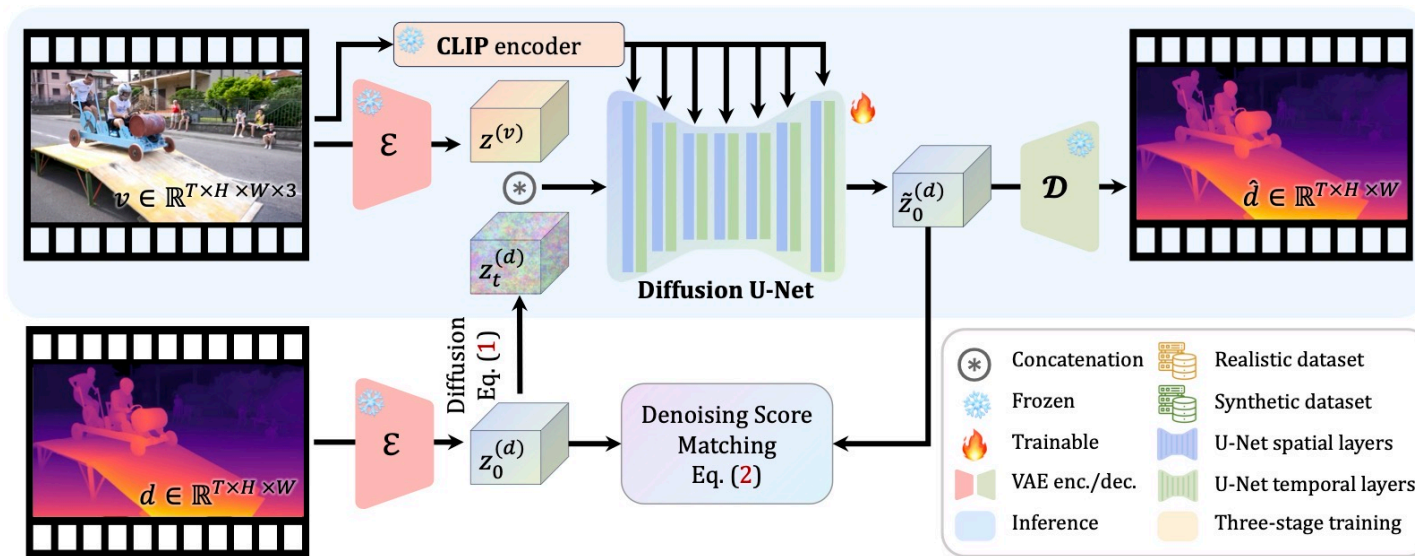
# Method

- Challenges:

- Generalize to open-world videos
- Accurate and consistent depth
- Long videos

- Solution

- Leverage video diffusion model for zero-shot generalization
- Large-scale video depth dataset, three-stage training
- Inference strategy for segment-wise processing and stitching



Inferencing 110 frames in a pass, and arbitrary length with inference strategy

# Results

- Quantitative results

Method	Sintel (~50 frames)		Scannet (90 frames)		KITTI (110 frames)		Bonn (110 frames)		NYU-v2 (1 frame)	
	AbsRel ↓	$\delta_1$ ↑	AbsRel ↓	$\delta_1$ ↑	AbsRel ↓	$\delta_1$ ↑	AbsRel ↓	$\delta_1$ ↑	AbsRel ↓	$\delta_1$ ↑
NVDS [63]	0.408	0.483	0.187	0.677	0.253	0.588	0.167	0.766	0.151	0.780
ChronoDepth [54]	0.587	0.486	0.159	0.783	0.167	0.759	0.100	0.911	0.073	0.941
Marigold [32]	0.532	0.515	0.166	0.769	0.149	0.796	0.091	0.931	0.070	0.946
Depth-Anything [67]	<u>0.325</u>	<u>0.564</u>	<u>0.130</u>	<u>0.838</u>	0.142	0.803	<u>0.078</u>	<u>0.939</u>	<b>0.042</b>	<b>0.981</b>
Depth-Anything-V2 [68]	0.367	0.554	0.135	0.822	<u>0.140</u>	<u>0.804</u>	0.106	0.921	<u>0.043</u>	<u>0.978</u>
DepthCrafter (Ours)	<b>0.270</b>	<b>0.697</b>	<b>0.123</b>	<b>0.856</b>	<b>0.104</b>	<b>0.896</b>	<b>0.071</b>	<b>0.972</b>	0.072	0.948





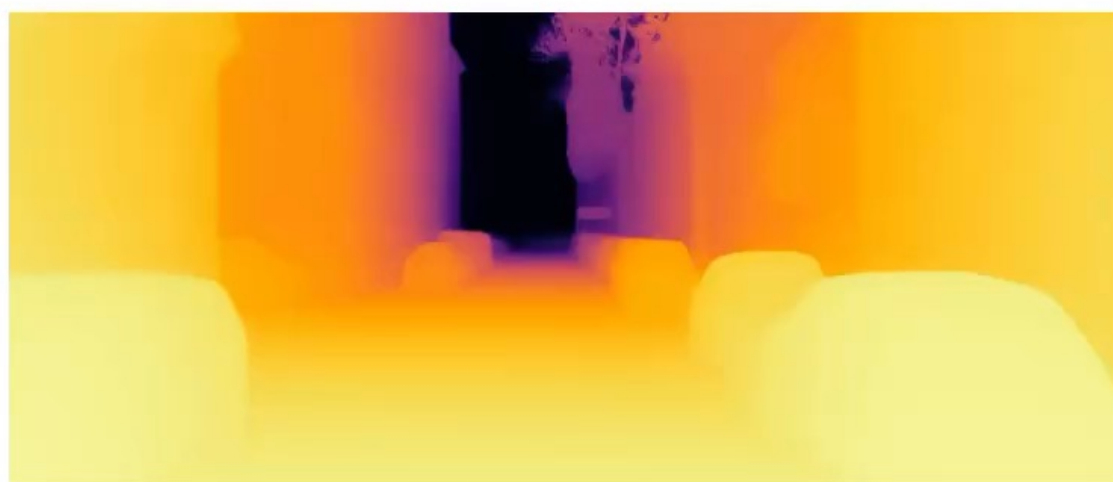
Input Video



GT



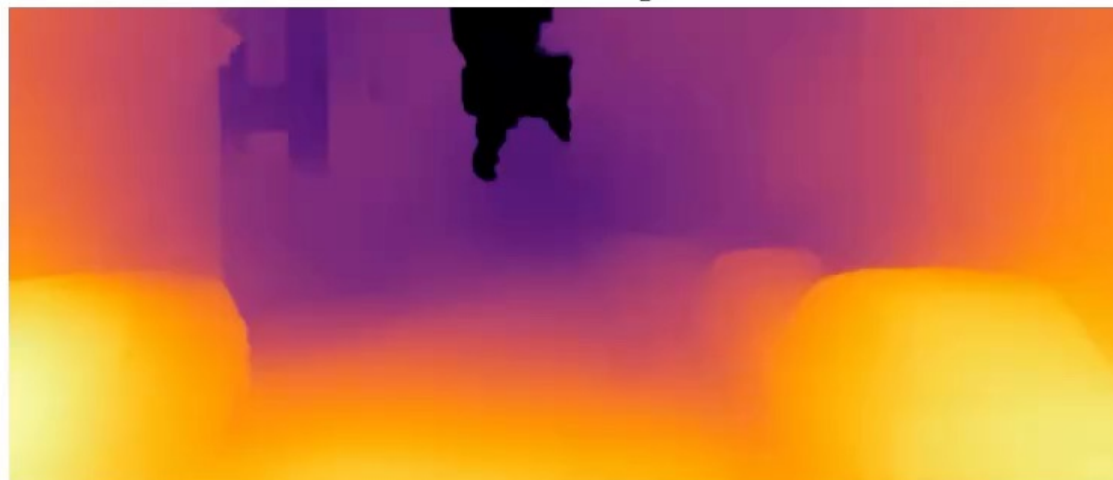
DepthCrafter (Ours)



ChronoDepth



Depth-Anything-V2



NVDS



Input Video



GT



DepthCrafter (Ours)



ChronoDepth



Depth-Anything-V2



NVDS



# Results

- Comparison with **Depth-Anything-V2** on in-the-wild videos



Input

Depth-Anything-V2

DepthCrafter (Ours)

# Results

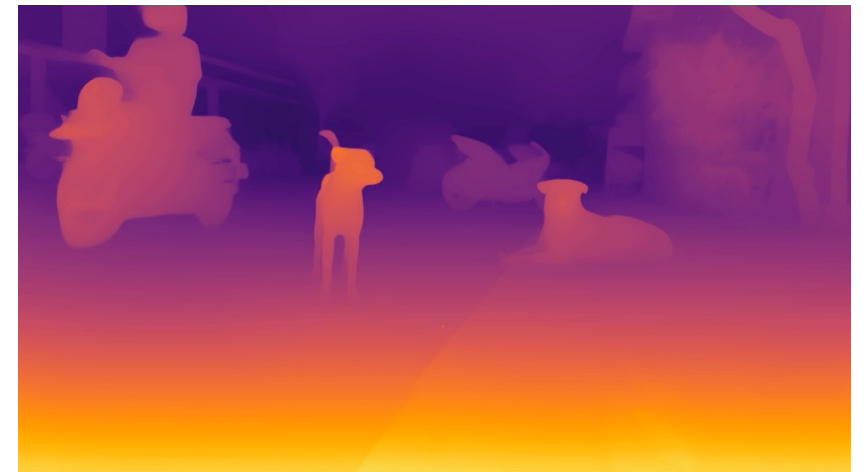
- Comparison with **DepthPro** on in-the-wild videos



Input



DepthPro



DepthCrafter (Ours)



# Results

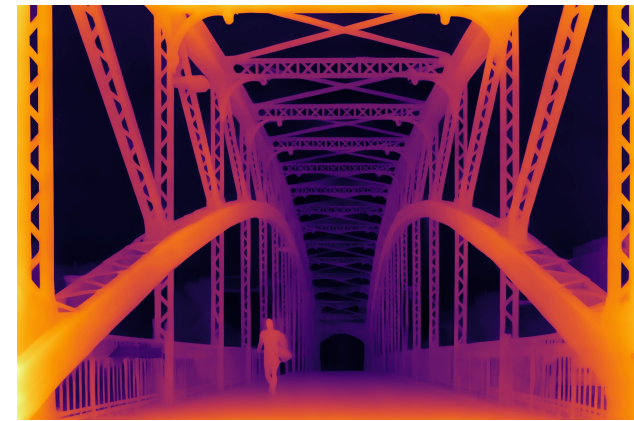
- More results for open-world videos





# Results

- Support variable video lengths, from single images to extremely long formats



Input

Depth-Anything-V2

DepthCrafter (Ours)



# Results

- Support variable video lengths, from single images to extremely long formats



# Applications

- Convert 2D videos to 3D stereoscopic ones



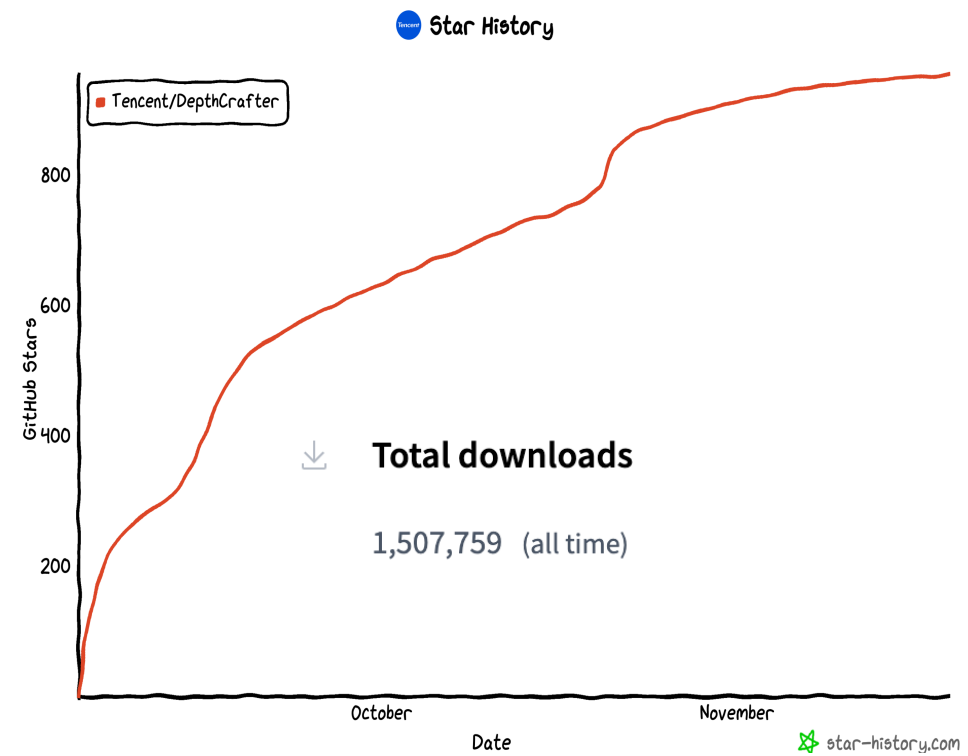


# Applications

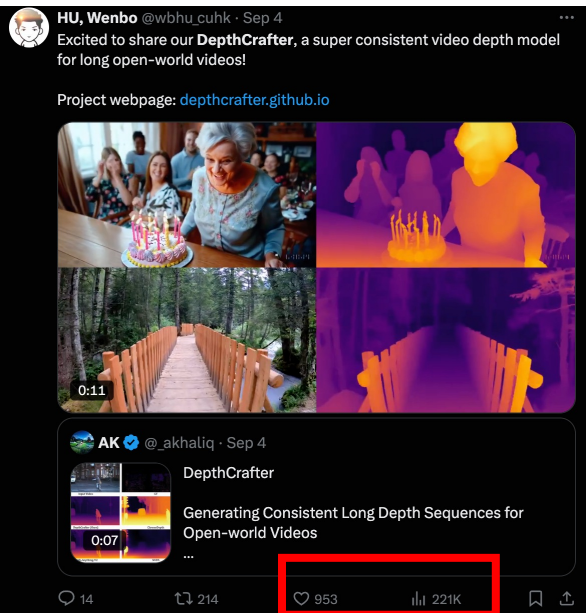
- Editing, visual effects



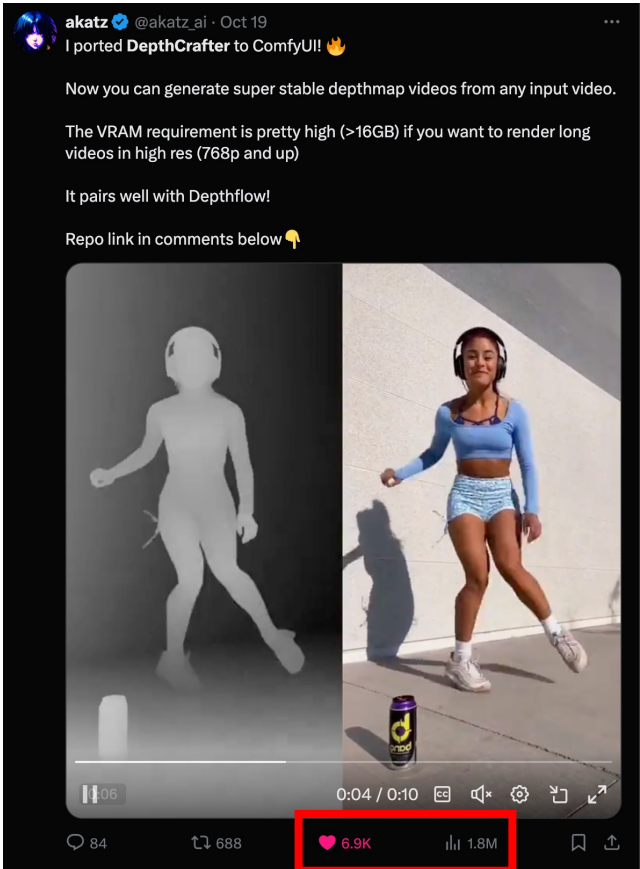
# Community Influence



GitHub code > **1.3K** Stars, HF model downloading > **1.5M**



Demo on X receives > **2M** Views, > **7K** Likes





# Thank You!

DepthCrafter: Generating Consistent Long Depth Sequences for Open-world Videos

