

1 **Videos with Tracks** The attached video file of zero-shot predicted point tracks demonstrates the
2 effectiveness of our method, KL-LRAS, on challenging, real-world scenes. These videos, sampled
3 from YouTube and TAP-Vid DAVIS, feature rapid and complex object and camera motion. We show
4 that with the right model properties and extraction method, large, self-supervised generative world
5 models can be tamed into a flow prediction procedure that generalizes to videos in the wild.

6 A Counterfactual World Model (CWM)

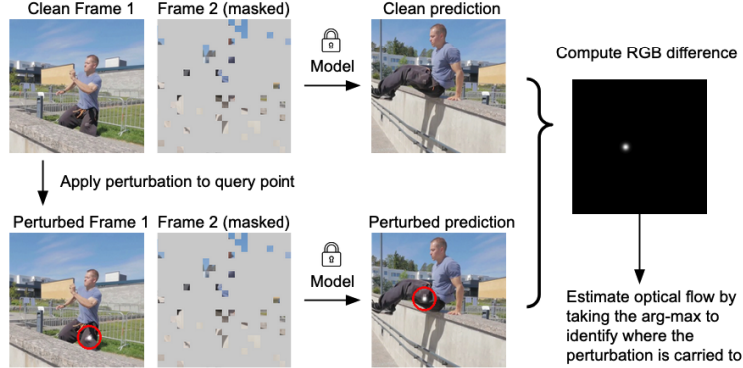


Figure 1: Flow extraction procedure for the Counterfactual World Model [1]

7 B Cosmos World Foundation Model

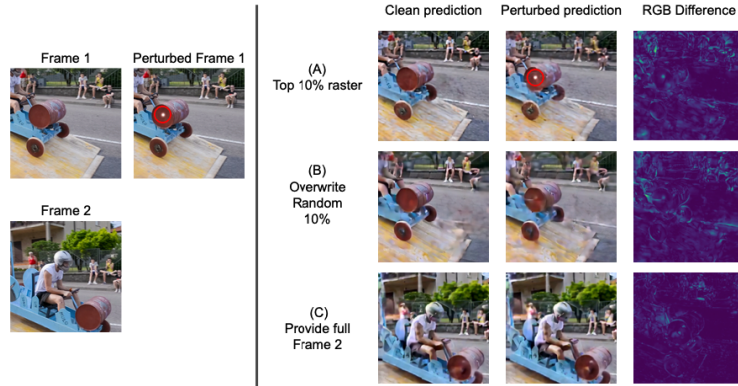


Figure 2: All evaluation settings for Cosmos [2] (Section 3.4) result in poor flow extractions. See Section 4.2 in the main text for a more detailed explanation of each result.

Model	TAP-Vid DAVIS Subset (3%) Endpoint Error (EPE)
LRAS KL (5MM, 8MS, 2STD) (ours)	5.0762
Cosmos (top 10% raster) (5MM, 2STD, 512×512) [2]	35.4338
Cosmos (overwrite 10% during rollout) (5MM, 2STD, 512×512) [2]	37.7552
Cosmos (provide full second frame) (5MM, 2STD, 512×512) [2]	66.2190

Table 1: Cosmos achieves poor optical flow results on benchmarks. Evaluations here are reported as endpoint error (EPE) on a TAP-Vid DAVIS subset. MM = multi-mask, MS = multi-scale.

References

- [1] Daniel M. Bear, Kevin Feigels, Honglin Chen, Wanhee Lee, Rahul Venkatesh, Klemen Kotar, Alex Durango, and Daniel L. K. Yamins. Unifying (Machine) Vision via Counterfactual World Modeling, June 2023. URL <http://arxiv.org/abs/2306.01828>. arXiv:2306.01828 [cs].
- [2] NVIDIA, Niket Agarwal, Arslan Ali, Maciej Bala, Yogesh Balaji, Erik Barker, Tiffany Cai, Prithvijit Chattopadhyay, Yongxin Chen, Yin Cui, Yifan Ding, Daniel Dworakowski, Jiaojiao Fan, Michele Fenzi, Francesco Ferroni, Sanja Fidler, Dieter Fox, Songwei Ge, Yunhao Ge, Jinwei Gu, Siddharth Gururani, Ethan He, Jiahui Huang, Jacob Huffman, Pooya Jannaty, Jingyi Jin, Seung Wook Kim, Gergely Klár, Grace Lam, Shiyi Lan, Laura Leal-Taixe, Anqi Li, Zhaoshuo Li, Chen-Hsuan Lin, Tsung-Yi Lin, Huan Ling, Ming-Yu Liu, Xian Liu, Alice Luo, Qianli Ma, Hanzi Mao, Kaichun Mo, Arsalan Mousavian, Seungjun Nah, Sriharsha Niverty, David Page, Despoina Paschalidou, Zeeshan Patel, Lindsey Pavao, Morteza Ramezanali, Fitsum Reda, Xiaowei Ren, Vasanth Rao Naik Sabavat, Ed Schmerling, Stella Shi, Bartosz Stefaniak, Shitao Tang, Lyne Tchapmi, Przemek Tredak, Wei-Cheng Tseng, Jibin Varghese, Hao Wang, Haoxiang Wang, Heng Wang, Ting-Chun Wang, Fangyin Wei, Xinyue Wei, Jay Zhangjie Wu, Jiashu Xu, Wei Yang, Lin Yen-Chen, Xiaohui Zeng, Yu Zeng, Jing Zhang, Qinsheng Zhang, Yuxuan Zhang, Qingqing Zhao, and Artur Zolkowski. Cosmos World Foundation Model Platform for Physical AI, March 2025. URL <http://arxiv.org/abs/2501.03575>. arXiv:2501.03575 [cs].