Making Video Models Adhere to User Intent with Minor Adjustments

Daniel Ajisafe[1] Eric Hedlin[1] Helge Rhodin[2,1] Kwang Moo Yi[1]
[1]The University of British Columbia, [2]Bielefeld University

RBC BOREALIS
CVPR JUNE 3-7, 2026 DENVER COLORADO

Main idea 💡

Small adjustments improve control adherence

Background

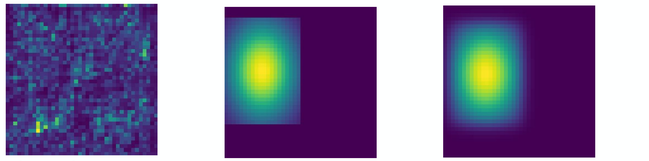
- Existing methods still struggle to reliably follow user controls
- User-provided control signals may not align with the model's internal representations

Our approach

- We adjust control boxes to better align with the internal attention mechanisms of the model
- We use smooth masking to allow optimization-based adjustments

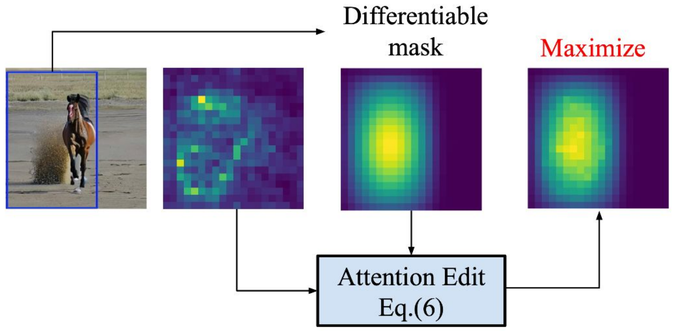
Method Overview

- **Smooth editing**

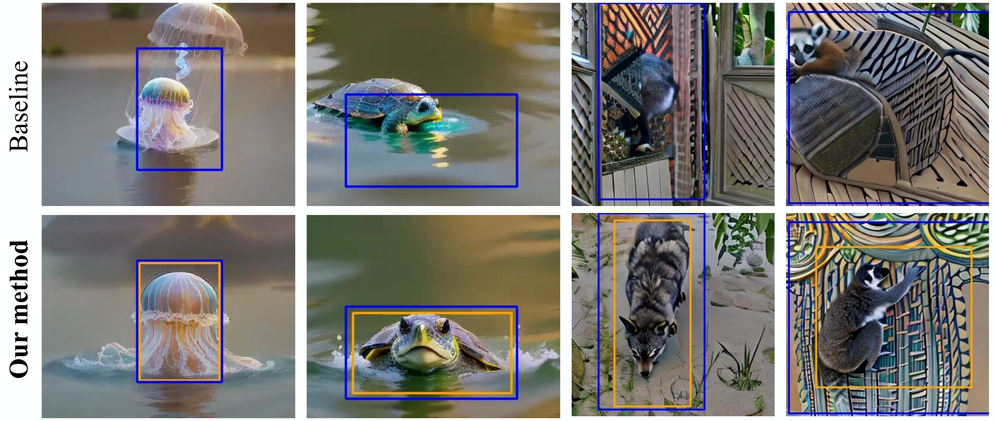


Attention map | Attention map edited by baseline | Attention map edited by our method

- **Forward and backward guidance**



Qualitative results



The jellyfish is swimming | The turtle is swimming | The wolf is exploring | The ring-tailed lemur is exploring

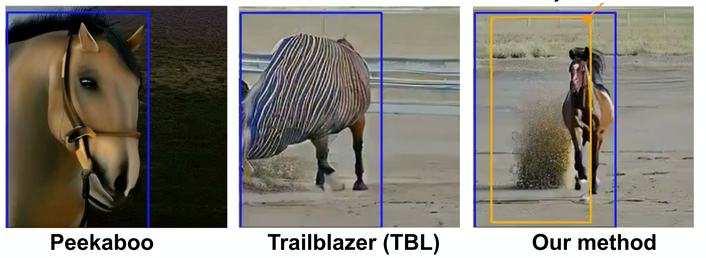
Teaser: Original boxes (top), adjusted boxes (bottom); Ours deliver better quality within control

Comparisons against baselines

Model	PickScore ↑	HPSv2 ↑	mIOU ↑
Trailblazer Ma et al. (2024b)	0.244	0.222	0.37
Our boxes + Trailblazer backbone	0.257	0.223	0.36
Our method w/o Box Opt.	0.243	0.221	0.37
Our method (full)	0.257	0.225	0.37
Peekaboo Jain et al. (2024)	0.125	0.189	0.30
Trailblazer Ma et al. (2024b)	0.146	0.222	0.37
Freetraj Qiu et al. (2024)	0.178	0.223	0.34
Trailblazer + T2V-Turbo backbone	0.234	0.253	0.41
Our method using T2V-Turbo backbone	0.317	0.263	0.41

Our full method outperforms baseline; demonstrates consistent preference across different architectures, while achieving competitive control

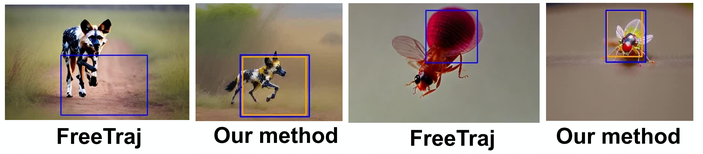
Additional qualitative results



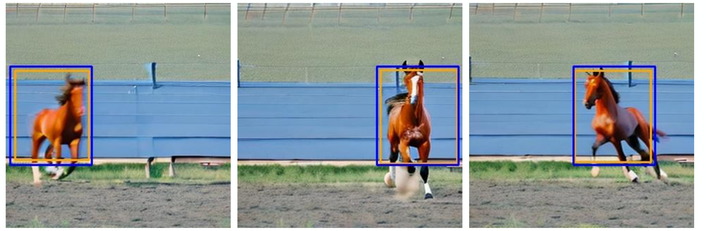
Peekaboo | Trailblazer (TBL) | Our method

TBL | Our boxes + TBL | TBL | Our boxes + TBL

Morphing task (Parrot → Penguin): Our boxes can provide better control



FreeTraj | Our method | FreeTraj | Our method



Our method can preserve complex trajectories such as U-turn

Future work

- Optimizing other conditional inputs, such as trajectories, sketches, or depth cues