ICReward: Learning Image-to-Video Consistency Rewards

Agnes Liang (agliang@stanford.edu), Renee Zbizika (rzbizika@stanford.edu)

Project Overview

Problem: Image-to-video (I2V) models often fail to preserve visual consistency with the input image. Small mismatches (like altered facial features, missing objects, or style drift) may degrade quality. Automated metrics like CLIP similarity [5] and FVD [3] do not reliably detect these inconsistencies and often fail to reflect human preferences.

We propose **ICReward**, a learned reward model trained on human preferences from the VBench++ dataset.

- Adapts VideoReward [2]to use a reference image instead of a text prompt.
- Adds an attention-based **Image Consistency head** to improve alignment.

Goal: Use DPO-style fine-tuning to train a generative I2V model (e.g., Open-Sora) to prefer videos selected by ICReward over those selected by the baseline.

Dataset

We use a curated 300-sample subset of the VBench++ dataset, designed to evaluate image-to-video (I2V) consistency in generative models. Each sample consists of:

- **1 source image** (used as the visual generation prompt)
- **5 generated videos** from OpenSora's I2V model
- **2 candidate videos** selected for evaluation per sample
- **Image Consistency Score** (real-valued, in [0, 1])

To improve label diversity, we manually added ∼30 examples with **IC scores** < 0.4, ensuring a more even distribution of supervision signals.

Method

Below, we show how we get I2V consistency metrics using VBench++ (a), adapt them for ICReward training with an image consistency head (b), and fine-tune a generative model using a DPO-inspired pairwise loss (c). Unlike the baseline MLP reward model, ICReward uses VLM features for better alignment. While DPO is typically used with human-labeled preferences, we instead treat the outputs of the two reward models as implicit preferences—using ICReward as the preferred signal over the MLP baseline.

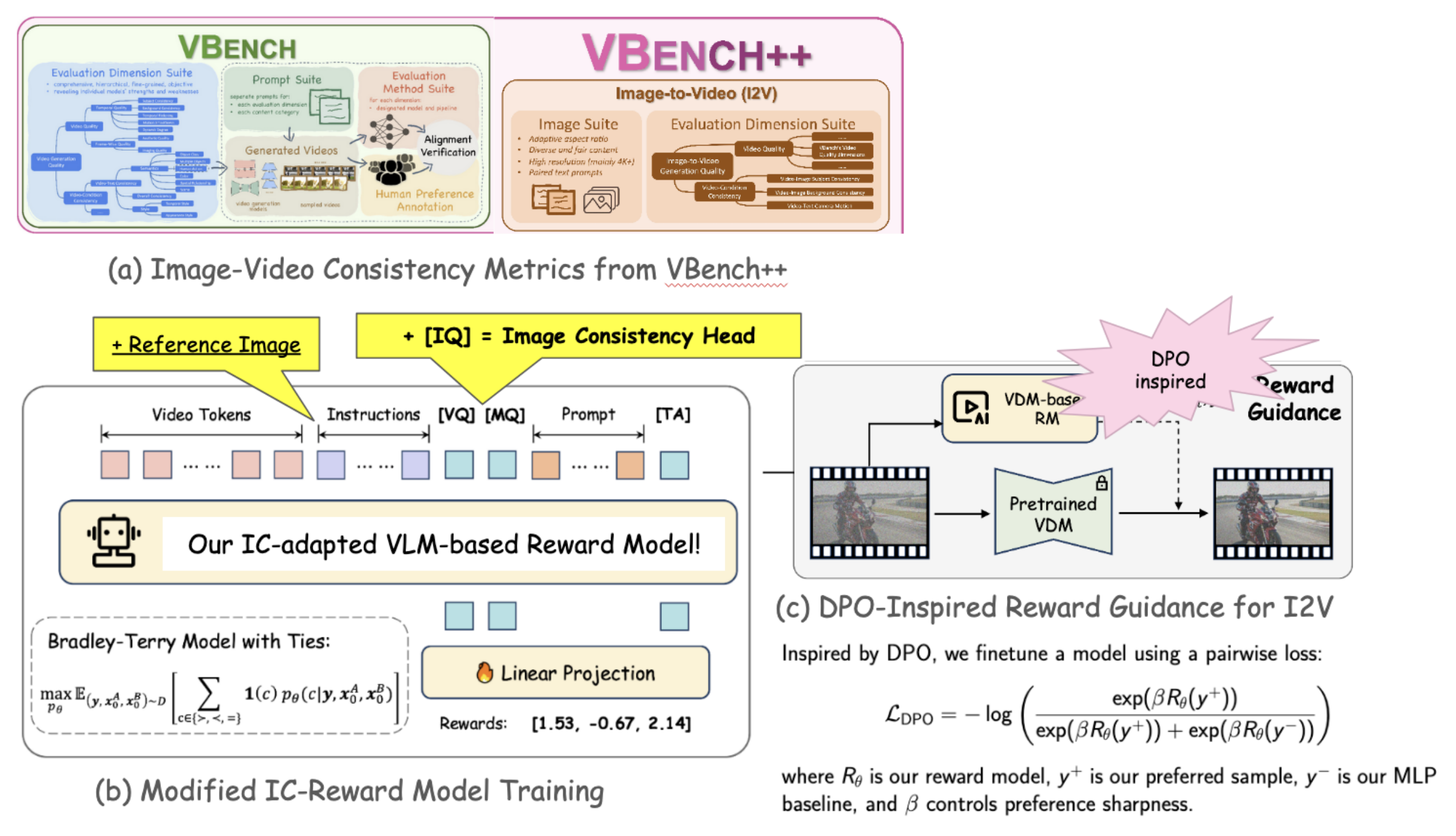
VBench
Evaluation Dimension Suite
Prompt Suite
Evaluation Method Suite
Generated Videos
Alignment Verification
Human Preference Annotation
VBench++
Image-to-Video (I2V)
Image Suite
Evaluation Dimension Suite
(a) Image-Video Consistency Metrics from VBench++
+ Reference Image
+ [IQ] = Image Consistency Head
Video Tokens
Instructions
[VQ]
[MQ]
Prompt
[TA]
Our IC-adapted VLM-based Reward Model!
Bradley-Terry Model with Ties:
$\max_{p_\theta} \mathbb{E}_{(y, x_0^A, x_0^B)\sim D}\left[\sum_{c\in\{>,<,=\}} \mathbb{1}(c)\, p_\theta(c|y, x_0^A, x_0^B)\right]$
Linear Projection
Rewards: [1.53, -0.67, 2.14]
(b) Modified IC-Reward Model Training
DPO inspired
VDM-base RM
Reward Guidance
Pretrained VDM
(c) DPO-Inspired Reward Guidance for I2V
Inspired by DPO, we finetune a model using a pairwise loss:
$\mathcal{L}_{\text{DPO}} = -\log\left(\frac{\exp(\beta R_\theta(y^+))}{\exp(\beta R_\theta(y^+)) + \exp(\beta R_\theta(y^-))}\right)$
where R_θ is our reward model, y^+ is our preferred sample, y^- is our MLP baseline, and β controls preference sharpness.

Figure 1. DPO Inspired Paired Loss. Original Figure by VideoReward [2]

Baseline Comparison

- **Baseline:** A simple MLP trained to predict human preferences from concatenated CLIP features of the reference image and video frames.

Experimental Results

Sanity Check: Does ICReward's image-conditioned feedback *actually* improve alignment?

I. Reward Model Development

Compare ICReward vs. MLP on held-out VBench++ pairs (vid, score) to confirm ICReward better reflects consistency.

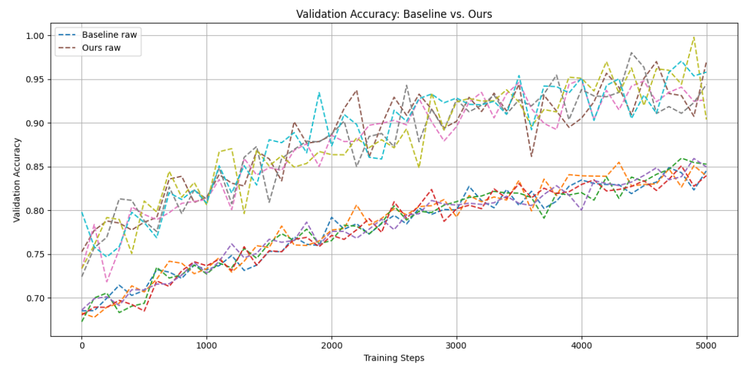

Figure 2. Validation MSE comparison between the baseline MLP reward model and our proposed ICReward model.

★ We evaluate both models on held-out video pairs from VBench++, where ICReward — trained on image-to-video consistency scores — consistently outperforms MLPBaseline, supporting our hypothesis.

II. Naive Sampling Evaluation

For each input image, we generate 5 videos using Open-Sora, rank them by average frame reward from each reward model, and compare the CLIP-Sim [5] scores of three selections: random, MLP-selected, and ICReward-selected.

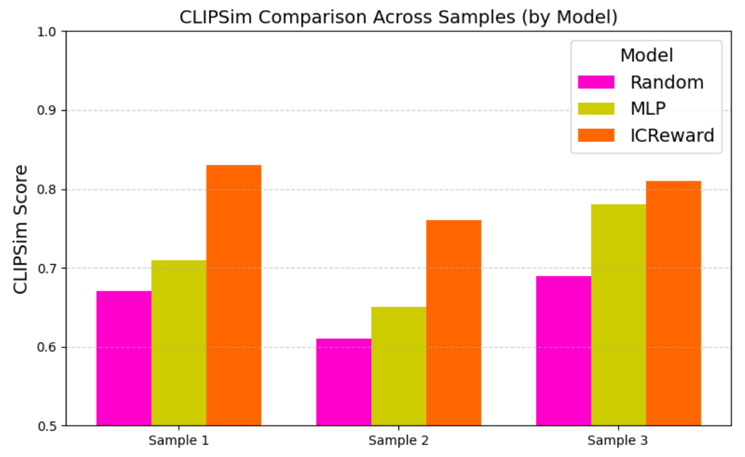

Figure 3. CLIPSim scores for top-1 selected videos across three input samples, grouped by reward model. Higher score means better semantic alignment with reference image.

★ ICReward consistently selects videos with higher semantic similarity and lower visual degradation compared to the MLP baseline.

Table 1. Average metrics across all test samples. ICReward outperforms MLP on both semantic alignment (CLIPSim) and perceptual realism (FVD).

Model	Avg. CLIPSim (↑)	Avg. FVD (↓)
MLP Reward	0.72	110.3
ICReward (ours)	0.81	89.6

On average, ICReward improves semantic alignment by +0.09 CLIPSim and reduces video degradation by over 20 FVD points.

Experimental Results (cont.)

III. Human Preference Validation Although ICReward is trained using model-generated labels from VBench++, we use human ratings to validate whether its preferences reflect human judgment. Written feedback from 20 participants highlights the types of visual consistency they value, reinforcing the usefulness of ICReward and MLP baseline as training-time proxies.

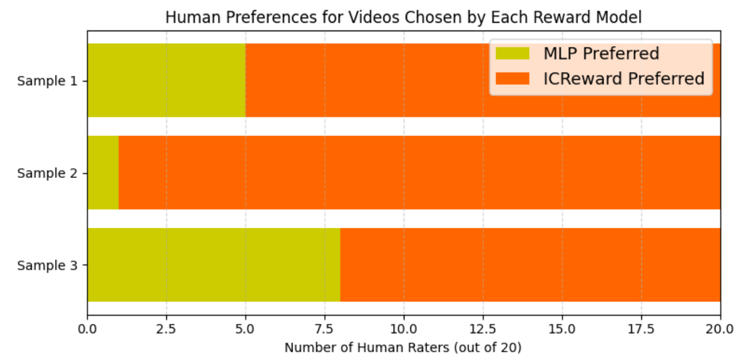

Figure 4. Human preference distribution across three samples. Each bar shows how many of 20 human raters preferred the video selected by the MLP reward model versus the ICReward model. ICReward was consistently favored.

Qualitative Feedback from Human Raters.

Since videos couldn't be shown on the poster, we observe key trends:

- **Sample 1** (person walking with scarf): ICReward preserved the subject better; MLP was more balanced but less consistent overall.
- **Sample 2** (autumnal foliage): MLP background drifted noticeably; ICReward was more faithful to the reference.
- **Sample 3** (coffee pouring): Both videos had unrealistic water physics.

DPO-style Fine-tuning on Open-Sora w/ ICReward vs. Baseline MLP

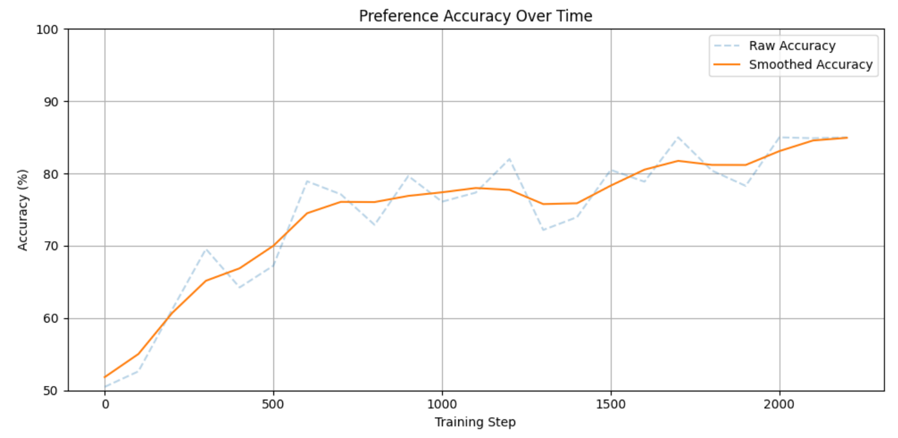

Figure 5. Preference accuracy over test steps. The model learns to prefer ICReward outputs over MLP. Accuracy plateaus 83% on test set.

The model rapidly improves in preference accuracy during early training, then stabilizes around 83% as it learns to align with ICReward. Cumulative gain steadily increases, showing consistent improvement in alignment with ICReward preferences over the baseline.

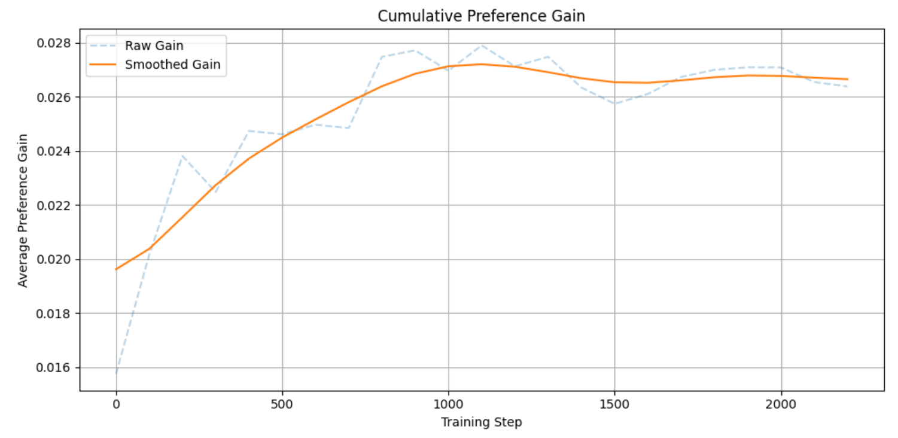

Figure 6. Running average of preference wins over the baseline MLP. The model steadily improves alignment with ICReward preferences

Future Work

- **Expand model comparisons:** Evaluate more SOTA models beyond VideoCrafter and OpenSora (ex. Gen-2, Hun Yuan, Pika)
- **Vary reward weighting:** Experiment w/ different weights for I2V subject vs. background consistency. Early feedback suggests neither dimension can be ignored without degrading overall alignment.
- **Multi-dimensional reward learning:** Extend ICReward to produce separate scores (subject, background, motion) and train models to balance multiple reward heads.
- **Broad human evaluation:** Scale up user studies to include forced-choice ranking, attribute-specific comparisons (e.g., realism vs. consistency)

Challenges and Limitations

While our ICReward model improves visual consistency in image-to-video (I2V) generation, several challenges remain:

- **Proxy Reward:** We use ICReward as a stand-in for human preferences, but real annotations would strengthen DPO fine-tuning.
- **Generalization:** Performance on unseen prompts and motion types remains underexplored.
- **Reward Hacking:** The model may exploit ICReward's biases rather than learning true consistency.

Improving robustness will require broader data, new consistency benchmarks, and more human-in-the-loop validation.

Conclusion

We present **ICReward**, a learned reward model for improving visual consistency in image-to-video (I2V) generation. Trained on VBench++ scores, it uses reference images and an Image Consistency head to better capture alignment.

We use ICReward to fine-tune Open-Sora with a DPO-style loss, improving preference accuracy over a baseline MLP and enabling reward-guided I2V generation aligned with human judgment.

References

[1] Ziqi Huang, Fan Zhang, Xiaojie Xu, Yinan He, Jiashuo Yu, Ziyue Dong, Qianli Ma, Nattapol Chanpaisit, Chenyang Si, Yuming Jiang, Yaohui Wang, Xinyuan Chen, Ying-Cong Chen, Limin Wang, Dahua Lin, Yu Qiao, and Ziwei Liu.
Vbench++: Comprehensive and versatile benchmark suite for video generative models.
2024.

[2] Jie Liu, Gongye Liu, Jiajun Liang, Ziyang Yuan, Xiaokun Liu, Mingwu Zheng, Xiele Wu, Qiulin Wang, Wenyu Qin, Menghan Xia, Xintao Wang, Xiaohong Liu, Fei Yang, Pengfei Wan, Di Zhang, Kun Gai, Yujiu Yang, and Wanli Ouyang.
Improving video generation with human feedback.
2025.

[3] Thomas Unterthiner, Sjoerd van Steenkiste, Karol Kurach, Raphaël Marinier, Marcin Michalski, and Sylvain Gelly.
Towards accurate generative models of video: A new metric & challenges.
CoRR, abs/1812.01717, 2018.

[4] Cong Wang, Jiaxi Gu, Panwen Hu, Songcen Xu, Hang Xu, and Xiaodan Liang.
Dreamvideo: High-fidelity image-to-video generation with image retention and text guidance.
2023.

[5] Tianhao Wu, Ji Cheng, Chaorui Zhang, Jianfeng Hou, Gengjian Chen, Zhongyi Huang, Weixi Zhang, Wei Han, and Bo Bai.
Clipsim: A gpu-friendly parallel framework for single-source simrank with accuracy guarantee.
Proceedings of the ACM on Management of Data, 1(1):1–26, May 2023.