Reward Forcing: Efficient Streaming Video Generation with Rewarded Distribution Matching Distillation

Supplementary Material

S1: More Video Results

Please check the videos in the project page https://reward-forcing.github.io/. These videos are compressed to approximately 40% of their original file size without significant quality degradation.

Comparison with state-of-the-art methods. In addition to Fig. 1, Fig. 4, and Fig. 5 in the main paper, we provide additional video results in our supplementary materials for a more comprehensive evaluation of long videos (approximately 1 minute) generated by different methods. This page includes comparative studies with state-of-the-art methods. We randomly sample prompts from MovieGenBench [58], focusing on Scene Navigation and Object Motion. As demonstrated in the videos, our Reward Forcing preserves high visual fidelity while exhibiting superior motion dynamics over ultra-long horizon, which is crucial for simulating dynamic environments.

Interactive videos. In addition to Fig. 7 in the main paper, we include more interactive video results in the "Reward Forcing.html" page, demonstrating that our Reward Forcing enables user interaction during streaming generation. Specifically, by switching prompts and resetting the cross-attention cache, the model can introduce new events into the ongoing video.

S2: User Studies

Experimental setup. To comprehensively evaluate the performance of our proposed method in long video generation, we conducted a user study with 20 participants. Each participant was presented with 20 video groups, where each group contained four videos generated by different methods: CausVid [93], Self-Forcing [31], LongLive [86], and Reward Forcing (ours). The videos were randomly labeled as A, B, C, and D to avoid bias. In total, we collected 1,600 evaluations (20 participants $\times$ 20 video groups $\times$ 4 videos).

Evaluation protocol. Participants are asked to evaluate each video for three key criteria using a 4-point Likert scale (1-4):

- Long-Range Temporal Consistency: This metric assesses whether the video maintains visual quality and coherence throughout its entire duration without experiencing visual drift, artifacts, or inconsistencies. Participants evaluated how well each video preserved semantic and structural consistency from start to finish.
- Dynamic Complexity: This metric measures the naturalness, richness, and engagement of motions and changes in the video. Participants assessed whether the generated content exhibited realistic and diverse dynamics rather than static or repetitive patterns.
- Overall Preference: This metric captures the holistic quality and appeal of each video, combining factors such as visual fidelity, coherence, motion quality, and subjective viewing experience.

For each criterion, participants assigned scores ranging from:

- 4 (Good): High quality with no noticeable issues.
- 3 (Borderline Accept): Acceptable quality with minor issues.
- 2 (Borderline Reject): Below acceptable quality with noticeable issues
- 1 (Poor): Unacceptable quality with major issues.

Results and analysis. The user study results unequivocally demonstrate the superiority of our proposed Reward Forcing method over the baseline models across all evaluation criteria (Tab. 4). Our method achieved the highest scores, nearing the "Good" (4) benchmark on the Likert scale, with 3.60 for Temporal Consistency, 3.72 for Dynamic Complexity, and 3.75 for Overall Preference. This indicates that participants consistently rated our videos as high-quality with no noticeable issues. These results validate that Reward Forcing sets a new state-of-the-art for coherent and engaging long video generation.

S3: More Quantitative Results and Details

Quality drift. We report the quality drift evaluation results in Tab. 2 and Tab. 3 in the main paper. To quantify the variability in long video imaging quality, we calculate the quality score drift along the temporal horizon using standard deviation,

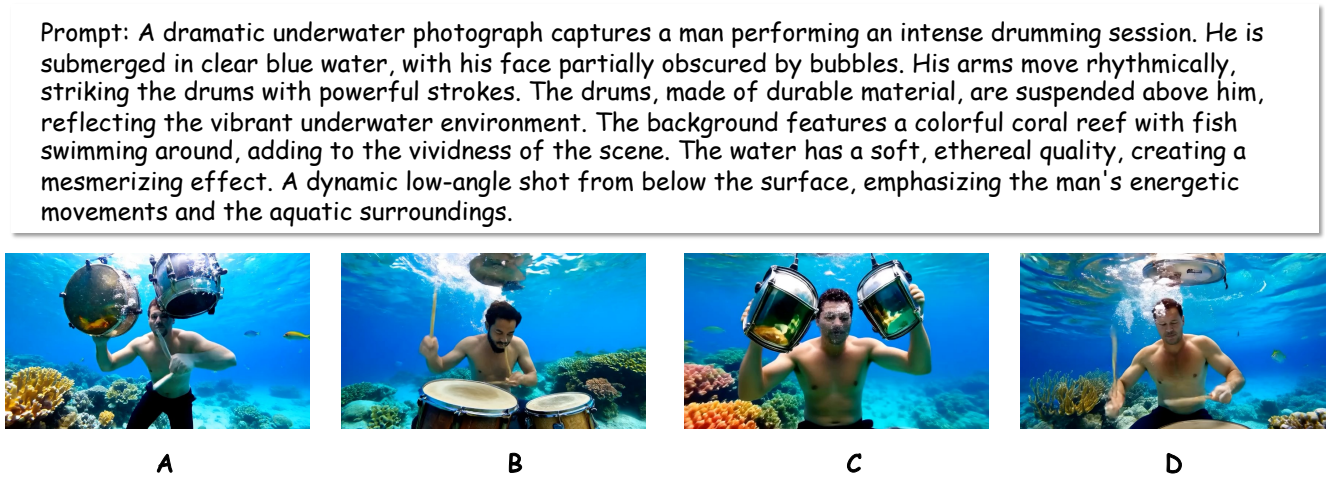

Figure 8. User study instruction screenshots.

Table 4. **Average User Rating.**

Models	Temporal Consistency	Dynamic Complexity	Overall Preference
CausVid [93]	1.81408	1.72676	1.87324
Self Forcing [31]	1.19437	1.75493	1.27042
LongLive [86]	2.78873	2.38310	2.74648
Reward Forcing	**3.60282**	**3.72113**	**3.75493**

inspired by Zhang et al. [98]. Each one-minute video is divided into M clips where $M = 30$, each lasting 2 seconds. For any given long video clip V_i, we compute the drift as follows:

$$\text{Drift}(V_i) = \sqrt{\frac{1}{M-1}\sum_{j=1}^{M}(s_{i,j} - \bar{s}_i)}, \tag{12}$$

where $C_{i,j}$ represent clip j from video i, $s_{i,j}$ be the imaging quality score of clip $C_{i,j}$. The overall drift across all videos is the mean of individual video drifts:

$$\text{Drift} = \frac{1}{N}\sum_{i=1}^{N}\text{Drift}(V_i), \tag{13}$$

where N is the total number of videos. Our results show that this metric effectively reflects video quality over long horizons, demonstrating a strong correlation between lower drift scores and more consistent visual fidelity throughout the sequences.

Qwen3-VL evaluation details. We use a powerful vision-language model, Qwen3-VL-235B-A22B-Instruct [85], for a more comprehensive evaluation and report the results in Tab. 2 in the main paper. We include the evaluation template and detailed results for different methods as follows.

Full Input Template

[VIDEO] """Please act as a video quality evaluation expert and rate the given video and text prompt on a scale of 1-5 across the following three dimensions:

Evaluation Dimensions:
1. **Text Alignment**:
Measures the consistency between the video content and the text description.
- 1: Completely Irrelevant - Content is unrelated or severely contradicts the description.
- 2: Mostly Mismatched - Only a few minor elements are relevant; the core concept is missing or incorrect.
- 3: Partially Matched - The core idea is present but with significant deviations or missing key elements.
- 4: Largely Consistent - Faithfully represents the description with only minor omissions or discrepancies.
- 5: Perfectly Aligned - Comprehensive and accurate representation of the entire text description.

2. **Dynamics**: Evaluates the dynamism and fluidity of the entire scene, including camera movement, object motion, and scene transitions.
- 1: Static / Disjointed - Little to no dynamic elements; or motion is severely broken and incoherent.
- 2: Mostly Static - Limited, simple motion; dynamics feel stiff, mechanical, or poorly executed.
- 3: Moderately Dynamic - Basic movement is present but lacks fluidity and natural flow; may appear robotic.
- 4: Largely Dynamic - Generally fluid and engaging motion with a good sense of flow; minor imperfections may exist.
- 5: Highly Dynamic - Exceptionally smooth, natural, and purposeful motion that enhances the visual narrative.

3. **Visual Quality**: Assesses the technical execution, including clarity, color grading, composition, and the absence of artifacts.
- 1: Very Poor - Severely blurry, heavy visual artifacts (e.g., distortion, tearing), and/or extreme color issues (e.g., over-saturation, color banding).
- 2: Poor - Consistently blurry, noticeable noise, unnatural color palette, or frequent minor artifacts.
- 3: Fair - Passable clarity and color, but with visible technical flaws; composition may be unremarkable.
- 4: Good - Clear and mostly sharp, with natural and balanced colors; good composition and only minor, infrequent issues.
- 5: Excellent - Technically superior: sharp, well-composed, with vivid yet natural colors, and free from visible artifacts or distortions.

Scoring Requirements:
- Please output strictly in the following format, only numbers and brief reasons:
Text Alignment: [1-5]
Reason: [brief explanation]
Dynamics: [1-5]
Reason: [brief explanation]
Visual Quality: [1-5]
Reason: [brief explanation]

Now please evaluate the following content:
Text Prompt: ""
Video Content: Please carefully watch the provided video
"""

CausVid Full Results

Average Scores by Dimension:
Text Alignment: 3.32
Dynamics: 3.16
Visual Quality: 4.66

Detailed Scores by Dimension:
Text Alignment: [5, 1, 2, 4, 5, 5, 5, 5, 5, 1, 3, 4, 1, 1, 4, 2, 4, 4, 2, 2, 3, 3, 2, 1, 5, 3, 2, 5, 3, 4, 5, 4, 2, 3, 2, 1, 3, 3, 2, 3, 5, 3, 1, 4, 5, 2, 2, 3, 2, 4, 5, 5, 3, 5, 3, 5, 3, 5, 5, 5, 4, 2, 5, 5, 5, 5, 5, 2, 5, 5, 5, 3, 2, 2, 5, 1, 3, 2, 4, 2, 2, 3, 2, 3, 3, 2, 1, 3, 2, 1, 2, 5, 3, 4, 4, 5, 3, 3, 5, 5, 1, 2, 4, 5, 5, 4, 3, 2, 5, 2, 5, 5, 3, 5, 4, 2, 5, 4, 2, 5, 5, 2, 1, 2, 5, 1, 3, 1, 5]
Dynamics: [4, 3, 3, 5, 4, 3, 4, 4, 3, 1, 4, 4, 2, 2, 4, 2, 4, 3, 4, 3, 4, 4, 3, 4, 3, 3, 2, 3, 4, 2, 3, 4, 3, 3, 4, 2, 3, 4, 2, 2, 3, 4, 3, 3, 4, 3, 2, 3, 2, 4, 3, 3, 4, 4, 3, 4, 4, 3, 3, 4, 3, 4, 4, 4, 1, 3, 4, 1, 5, 3, 4, 4, 2, 3, 4, 3, 2, 3, 2, 4, 4, 5, 2, 3, 2, 2, 1, 4, 3, 4, 3, 3, 3, 2, 3, 4, 3, 2, 4, 4, 1, 1, 3, 4, 4, 4, 2, 4, 3, 3, 3, 3, 2, 4, 5, 3, 4, 4, 3, 3, 3, 3, 4, 2, 4, 2, 2, 3, 3]
Visual Quality: [5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 4, 5, 4, 5, 5, 4, 3, 5, 5, 4, 5, 5, 4, 4, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 4, 5, 4, 5, 5, 5, 5, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 4, 5, 5, 5, 5, 4, 4, 5, 4, 4, 4, 5, 5, 4, 4, 4, 4, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 3, 3, 4, 5, 5, 5, 4, 4, 4, 4, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 4, 5]

SkyReels Full Results

Average Scores by Dimension:
Text Alignment: 2.70
Dynamics: 3.05
Visual Quality: 3.30

Detailed Scores by Dimension:
Text Alignment: [4, 2, 3, 2, 5, 5, 4, 2, 4, 1, 5, 3, 1, 2, 3, 1, 5, 2, 3, 1, 1, 1, 4, 1, 1, 2, 2, 1, 3, 4, 4, 1, 3, 3, 3, 1, 2, 5, 2, 2, 2, 2, 1, 4, 1, 1, 1, 2, 2, 2, 4, 3, 2, 5, 4, 5, 2, 5, 1, 5, 2, 2, 5, 2, 1, 5, 2, 4, 5, 3, 4, 2, 3, 3, 2, 4, 3, 1, 2, 2, 1, 3, 1, 3, 1, 5, 5, 1, 1, 3, 4, 4, 2, 1, 5, 4, 2, 3, 1, 4, 3, 1, 3, 2, 5, 3, 1, 1, 1, 5, 5, 1, 3, 3, 2, 3, 4, 1, 3, 5, 3, 2, 4, 1, 1, 5, 1, 5]
Dynamics: [5, 3, 1, 2, 3, 4, 3, 3, 3, 3, 4, 3, 3, 3, 4, 3, 5, 4, 4, 2, 3, 2, 4, 1, 3, 3, 4, 1, 2, 3, 3, 3, 2, 4, 4, 3, 3, 4, 3, 3, 2, 1, 3, 4, 4, 1, 2, 4, 3, 3, 5, 4, 2, 4, 3, 4, 3, 3, 2, 3, 3, 3, 3, 4, 1, 2, 4, 3, 5, 3, 2, 4, 4, 4, 3, 3, 4, 3, 2, 2, 3, 1, 4, 3, 2, 2, 5, 4, 2, 1, 3, 3, 3, 2, 1, 3, 3, 3, 3, 4, 3, 4, 3, 1, 4, 3, 5, 2, 1, 3, 1, 3, 4, 3, 4, 4, 2, 3, 4, 4, 2, 4, 4, 4, 5, 2, 4, 4, 3, 3]
Visual Quality: [4, 4, 2, 1, 4, 5, 3, 4, 3, 4, 5, 4, 4, 3, 4, 3, 5, 5, 3, 1, 3, 1, 4, 1, 3, 5, 4, 1, 4, 5, 4, 3, 5, 2, 4, 4, 3, 5, 4, 2, 3, 2, 3, 3, 4, 1, 3, 4, 3, 4, 4, 3, 2, 4, 3, 4, 3, 3, 3, 3, 2, 4, 3, 1, 3, 5, 3, 4, 3, 1, 4, 4, 4, 2, 3, 3, 4, 1, 3, 1, 3, 3, 2, 4, 4, 5, 3, 1, 2, 4, 4, 5, 3, 1, 4, 4, 4, 4, 2, 4, 3, 2, 4, 4, 5, 2, 3, 3, 1, 4, 4, 4, 3, 2, 2, 4, 3, 4, 2, 4, 3, 5, 4, 4, 5, 5, 4, 5]

Self Forcing Full Results

Average Scores by Dimension:
Text Alignment: 3.11
Dynamics: 3.44
Visual Quality: 3.89

Detailed Scores by Dimension:
Text Alignment: [4, 1, 2, 2, 5, 3, 5, 3, 5, 1, 4, 3, 1, 1, 5, 3, 2, 2, 3, 3, 1, 2, 2, 1, 5, 4, 2, 5, 4, 5, 3, 2, 3, 3, 2, 5, 1, 3, 3, 1, 4, 3, 1, 2, 1, 2, 1, 4, 4, 3, 1, 5, 5, 5, 1, 5, 4, 3, 3, 2, 4, 2, 5, 5, 2, 5, 3, 5, 5, 3, 4, 1, 3, 3, 2, 3, 3, 3, 3, 3, 3, 2, 2, 2, 5, 5, 4, 2, 1, 2, 5, 4, 2, 5, 5, 3, 2, 5, 1, 1, 3, 1, 5, 5, 3, 4, 2, 5, 2, 5, 4, 3, 3, 3, 3, 5, 2, 2, 5, 5, 3, 3, 4, 5, 3, 2, 1, 5]
Dynamics: [4, 3, 3, 2, 3, 3, 3, 3, 4, 3, 3, 3, 3, 3, 4, 3, 4, 4, 4, 4, 3, 4, 3, 1, 4, 3, 3, 4, 3, 4, 3, 4, 2, 4, 4, 5, 3, 3, 3, 3, 4, 2, 3, 2, 4, 4, 3, 4, 3, 3, 4, 4, 2, 3, 2, 4, 4, 3, 4, 3, 4, 4, 5, 3, 4, 4, 4, 5, 3, 3, 5, 3, 4, 4, 4, 4, 4, 2, 4, 4, 5, 3, 3, 2, 5, 5, 4, 4, 4, 4, 3, 3, 2, 4, 3, 3, 3, 4, 2, 2, 4, 1, 4, 4, 4, 3, 5, 3, 4, 3, 3, 4, 4, 4, 4, 3, 4, 3, 3, 3, 4, 4, 4, 4, 4, 3, 2, 3]
Visual Quality: [5, 4, 4, 2, 4, 4, 5, 4, 4, 4, 5, 3, 2, 3, 5, 4, 5, 4, 3, 4, 4, 3, 4, 2, 5, 4, 4, 4, 5, 5, 4, 4, 3, 3, 3, 5, 4, 4, 3, 2, 5, 4, 4, 3, 5, 3, 4, 4, 4, 3, 4, 5, 4, 4, 3, 3, 5, 4, 4, 3, 4, 5, 5, 4, 3, 4, 3, 5, 4, 4, 5, 4, 3, 4, 5, 5, 4, 1, 4, 3, 4, 2, 3, 4, 5, 5, 5, 4, 4, 5, 4, 5, 4, 4, 3, 3, 4, 4, 3, 4, 3, 1, 4, 4, 5, 5, 3, 3, 4, 3, 3, 4, 4, 5, 4, 5, 3, 3, 5, 4, 3, 5, 5, 4, 5, 4, 4, 4]

LongLive Full Results

Average Scores by Dimension:
Text Alignment: 3.98
Dynamics: 3.81
Visual Quality: 4.79

Detailed Scores by Dimension:
Text Alignment: [5, 2, 4, 5, 5, 5, 5, 5, 5, 1, 5, 3, 1, 2, 5, 4, 4, 4, 3, 5, 3, 3, 3, 4, 5, 5, 3, 5, 3, 5, 5, 4, 4, 3, 3, 2, 5, 5, 2, 3, 5, 3, 1, 5, 5, 5, 4, 5, 4, 1, 5, 5, 5, 5, 3, 5, 4, 5, 5, 5, 5, 2, 5, 5, 5, 3, 2, 5, 5, 5, 4, 1, 5, 5, 2, 5, 5, 5, 4, 4, 1, 4, 4, 2, 3, 5, 5, 3, 1, 5, 5, 3, 3, 5, 5, 4, 2, 3, 5, 3, 3, 5, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 5, 4, 5, 4, 3, 5, 5, 5, 2, 4, 5, 1, 3, 1, 5]
Dynamics: [5, 4, 3, 4, 4, 4, 3, 4, 3, 4, 4, 3, 4, 3, 4, 3, 5, 5, 4, 5, 3, 4, 4, 3, 4, 4, 4, 4, 2, 4, 4, 4, 4, 4, 4, 3, 4, 4, 3, 4, 4, 2, 3, 5, 4, 4, 5, 5, 4, 3, 4, 4, 1, 3, 4, 4, 4, 3, 4, 3, 5, 4, 4, 1, 4, 4, 4, 4, 3, 3, 4, 3, 5, 5, 4, 5, 4, 4, 5, 5, 4, 4, 3, 4, 2, 5, 4, 4, 4, 5, 3, 3, 2, 3, 3, 3, 2, 3, 4, 3, 4, 4, 4, 4, 4, 5, 5, 5, 3, 4, 3, 4, 5, 5, 5, 5, 3, 5, 4, 4, 4, 5, 4, 5, 3, 4, 2, 4, 3]
Visual Quality: [5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 5, 4, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 4, 3, 5, 5, 5, 5, 5, 4, 4, 4, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5, 4, 5, 5, 4, 4, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 5, 5, 5, 5, 5, 5]

Reward Forcing Full Results

Average Scores by Dimension:
Text Alignment: 4.04
Dynamics: 4.18
Visual Quality: 4.82

Detailed Scores by Dimension:
Text Alignment: [5, 2, 4, 2, 5, 5, 5, 5, 3, 1, 5, 5, 1, 1, 5, 5, 5, 4, 5, 5, 3, 5, 5, 4, 5, 5, 4, 5, 3, 5, 5, 4, 5, 4, 3, 2, 4, 4, 1, 3, 5, 3, 1, 5, 5, 5, 2, 5, 3, 2, 5, 5, 5, 5, 5, 5, 3, 5, 5, 5, 5, 2, 5, 5, 5, 2, 2, 5, 5, 5, 4, 1, 5, 5, 3, 5, 5, 5, 3, 4, 2, 4, 3, 4, 4, 5, 4, 5, 1, 5, 5, 4, 3, 5, 5, 4, 3, 5, 5, 2, 5, 4, 5, 5, 5, 4, 4, 5, 3, 5, 4, 5, 5, 2, 4, 5, 5, 3, 4, 5, 4, 2, 5, 5, 2, 5, 2, 5]
Dynamics: [5, 4, 3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 5, 5, 4, 5, 3, 5, 4, 5, 4, 4, 4, 4, 4, 4, 4, 4, 4, 5, 4, 4, 4, 4, 3, 4, 4, 3, 4, 5, 5, 4, 4, 5, 4, 4, 4, 5, 4, 4, 4, 4, 4, 4, 4, 4, 5, 4, 4, 4, 4, 4, 3, 5, 4, 4, 5, 4, 5, 5, 4, 4, 5, 5, 4, 5, 4, 4, 4, 4, 3, 5, 4, 5, 4, 5, 4, 3, 3, 4, 4, 3, 4, 4, 4, 3, 5, 4, 4, 5, 5, 5, 5, 4, 4, 4, 3, 5, 5, 4, 5, 4, 5, 4, 4, 4, 5, 4, 5, 4, 4, 4, 4, 4]
Visual Quality: [5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 5, 4, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 4, 5, 5, 5, 5, 4, 5, 5, 3, 5, 5, 5, 5, 5, 4, 4, 4, 5, 4, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 5, 4, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 4, 5, 5, 4, 4, 5, 5, 5, 5, 5, 5]

Detailed results on VBench [32]. We report quantitative evaluation on VBench [32] using the extended prompts in Tab. 1 in the main paper. Specifically, the Quality Score is a weighted average of the following dimensions: subject consistency, background consistency, temporal flickering, motion smoothness, aesthetic quality, imaging quality, and dynamic degree. The Semantic Score is a weighted average of the following dimensions: object class, multiple objects, human action, color, spatial relationship, scene, appearance style, temporal style, and overall consistency. Each dimension's results are normalized using the following formula: normalized score = (score - min) / (max - min). The normalization range (minimum and maximum) for each dimension and the assigned weights used to compute the weighted average are provided in the Tab. 5.

Table 5. **Normalization ranges and weighting coefficients of VBench score.**

	Subject Consistency	Background Consistency	Temporal Flickering	Motion Smoothness	Dynamic Degree	Aesthetic Quality	Imaging Quality	Overall Consistency
min	0.1462	0.2615	0.6293	0.706	0.0	0.0	0.0	0.0
max	1.0	1.0	1.0	0.9975	1.0	1.0	1.0	0.364
weighting coefficients	1	1	1	1	0.5	1	1	1
	Object Class	**Multiple Objects**	**Human Action**	**Color**	**Spatial Relationship**	**Scene**	**Temporal Style**	**Appearance Style**
min	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0009
max	1.0	1.0	1.0	1.0	1.0	0.8222	0.364	0.2855
weighting coefficients	1	1	1	1	1	1	1	1

In addition, we provide detailed evaluation in Tabs. 6 and 7. Our method achieves a total score of 84.13 on the short video generation task using the extended VBench prompts, consistently surpassing current state-of-the-art baselines and demonstrating its effectiveness.

Table 6. **Quality evaluation on extended VBench.**

Model	Subject Consistency	Background Consistency	Temporal Flickering	Motion Smoothness	Dynamic Degree	Aesthetic Quality	Imaging Quality	Quality Score	Total Score
CausVid [93]	96.33	95.84	99.44	97.98	61.11	64.52	67.96	83.93	82.88
Self Forcing [31]	95.09	96.10	99.01	98.24	66.38	65.79	69.71	84.59	83.80
LongLive [86]	96.98	96.92	99.35	98.79	40.83	67.03	69.18	83.68	83.22
Reward Forcing	95.43	96.59	98.97	98.32	68.05	65.66	69.38	**84.84**	**84.13**

Table 7. **Semantic evaluation on extended VBench.**

Model	Object Class	Multiple Objects	Human Action	Color	Spatial Relationship	Scene	Temporal Style	Appearance Style	Overall Consistency	Semantic Score	Total Score
CausVid [93]	92.78	88.32	96.20	86.67	74.05	51.35	23.95	20.19	25.95	78.69	82.88
Self Forcing [31]	93.16	87.19	96.40	86.83	81.77	56.13	24.45	20.34	26.85	80.64	83.80
LongLive [86]	96.28	86.49	95.80	90.79	80.56	58.79	24.16	20.42	26.61	**81.37**	83.22
Reward Forcing	94.81	86.79	96.80	89.42	82.47	57.19	24.33	20.38	26.88	81.32	**84.13**

S4: More Implementation details

Noise schedule and model parameterization. Building upon the Wan2.1 and Self Forcing, our approach utilizes the flow matching framework. We implement a time step shift defined as $t'(k,t) = (kt/1000)/(1+(k-1)(t/1000)) \cdot 1000$ with a shift factor k set to 5. In the forward process, a sample is generated according to $x_t = \frac{t'}{1000}x + \frac{1-t'}{1000}\epsilon$, where ϵ is drawn from a standard normal distribution $\mathcal{N}(0, \mathbf{I})$ and t ranges from 0 to 1000. The data prediction model is formulated as :

$$G_\theta(\boldsymbol{x}, t, c) = c_{\text{skip}} \cdot \epsilon - c_{\text{out}} \cdot v_\theta(c_{\text{in}} \cdot x_t, c_{\text{noise}}(t'), c). \tag{14}$$

The preconditioning coefficients remain consistent with the base models' settings: specifically, c_{skip}, c_{in}, c_{out} are all 1, and $c_{\text{noise}}(t) = t$. For our few-step diffusion sampling, we adopt a uniform 4-step schedule with time steps $\left[t_1, t_2, t_3, t_4\right] = \left[1000, 750, 500, 250\right]$.

S5: Further Related Works

Video diffusion models. Video diffusion models [3, 24, 27, 28, 82] have evolved from UNet [6, 64] backbones to Diffusion Transformers (DiTs) [56]. Early approaches extended image diffusion models temporally [66] but lacked scalability. DiT's Transformer blocks enhance spatio-temporal modeling, enabling models like Sora [4] and Hunyuan-Video [38] to generate realistic, coherent videos. Hunyuan-Video integrates causal 3D VAE [37] and language models for textual control. Open-Sora [41] advanced efficiency and realism, while Wan 2.1 [75] validated large-scale pre-training benefits and CogVideoX [29, 87] improved alignment via adaptive LayerNorm. For long video generation, Phenaki [74] uses discrete tokens, LVDM [24] employs hierarchical 3D latent generation, and NUWA-XL [89] uses coarse-to-fine processing. LaVie [77] integrates rotary encoding and temporal attention, SEINE [10] enables smooth transitions via stochastic masking, and LCT [22] extends to multi-shot generation. Diffusion forcing [7] combines diffusion quality with autoregressive efficiency. StreamingT2V [25] adds memory modules, History-guided video [67] uses historical context, FramePack [98] compresses frames, Lumos-1 [95] integrates LLM-like architecture, and LongVie [19] introduces multi-modal guidance and degradation-aware training. Test-time training methods [13] generate minute-long videos but incur high costs. Training-free methods—RIFLEx [100] adjusting positional embeddings, FreeNoise [61] combining noise rescheduling with windowed attention, and FreeLong [50] integrating multi-frequency information.

Reinforcement learning for video models. Video generative models [5, 9, 36, 62, 69, 71, 73, 80] using MLE or reconstruction loss misalign with human preferences. RL enables direct optimization of preference-aligned objectives [20, 43]. Direct Preference Optimization (DPO) [17] dominates post-training alignment, including VideoDPO [48] for temporal consistency, VisionReward [83] for multi-objective preferences, and variants with physics-based generation. Group Relative Policy Optimization (GRPO), extending PPO [30, 53, 99], improves generalization as shown in DanceGRPO [84]. Reward-based approaches like InstructVideo [94] with pretrained reward feedback and VADER [59] with unified differentiable rewards bypass policy learning. Inference-time methods like InfLVG [15] incorporate GRPO for dynamic long-form optimization. Collectively, RL serves as both post-training alignment and structural component for preference-aware generation [11, 16, 18, 23, 44, 49, 78], bridging surrogate objectives and human-valued quality.

S6: Discussion and Future Work

Generalizability. Our method is designed to be general-purpose and plug-and-play, enabling seamless integration with various video generation architectures without requiring substantial modifications to existing pipelines. This flexibility represents a significant practical advantage, as it allows researchers and practitioners to adopt our approach with minimal overhead.

Misalignment between reward functions and evaluation criteria. The first factor contributing to inconsistent performance is the misalignment between our reward function's optimization direction and VBench's evaluation criteria. VBench employs comprehensive metrics including temporal consistency, motion smoothness, subject consistency, background quality, aesthetics, and semantic alignment. Our reward model may prioritize certain dimensions over others—for example, heavily weighting temporal coherence while underemphasizing aesthetic qualities. This asymmetric optimization creates scenarios where reward improvements don't translate proportionally to VBench score gains.

Video reward models. Our experiments show that current reward models can effectively guide quality improvements, as reflected in our competitive performance across multiple benchmarks. However, video reward models still face challenges in capturing certain nuanced aspects of video quality, such as long-range temporal dependencies, subtle temporal artifacts like frame jitter, and complex semantic attributes. These models are typically trained on datasets with subjective annotations that may not fully represent all quality dimensions. As video reward models continue to advance, our framework will naturally benefit from these improvements, enabling further optimization.

Future research directions. Future research should develop more sophisticated reward models capturing video quality nuances. Promising directions include: multi-objective reward modeling with separate components for different quality dimensions; hierarchical models assessing quality at multiple temporal scales; human-in-the-loop feedback mechanisms grounding models in perceptual judgments; domain-adaptive models adjusting criteria by content type; and architectures encoding physical and semantic priors about real-world dynamics. Advancing reward modeling along these dimensions could help our method achieve its full potential and demonstrate substantial, consistent improvements across comprehensive evaluation frameworks.

S7: Border Social Impact

This work on efficient streaming video generation presents both significant opportunities and risks. On the positive side, the reduced computational requirements could democratize access to video synthesis technology, benefiting educational content creators, small organizations, and researchers with limited resources. The improved efficiency also reduces energy consumption, contributing to more sustainable AI development. However, we acknowledge serious concerns regarding potential misuse. The accessibility and speed of our method lowers barriers for creating deepfakes and misleading visual content that could spread misinformation or enable identity fraud. Additionally, our reward-based prioritization of dynamic content may inadvertently amplify biases present in vision-language models, potentially marginalizing underrepresented groups or activities. Questions of copyright infringement and consent regarding training data and generated likenesses also warrant careful consideration. To mitigate these risks, we strongly advocate for implementing digital watermarking and provenance tracking in any deployment of this technology. We encourage development of detection tools for synthetic content, clear content labeling practices, and robust usage policies prohibiting malicious applications such as non-consensual deepfakes. We support collaborative efforts among researchers, policymakers, and civil society to establish ethical guidelines, legal frameworks, and media literacy initiatives. Effective governance requires not only technological safeguards but also transparent data practices, diverse evaluation metrics to reduce bias, and ongoing dialogue about responsible use of generative video technologies.