ByteDance 字节跳动

中国科学院自动化研究所 Institute of Automation Chinese Academy of Sciences

VlogReward: Learning Multi-Dimensional Evaluation for Vlog Editing

Yexiang Liu, Wen Zhong, Sijie Zhu, Xin Gu, Fan Chen,
Junxian Duan, Jie Cao, Longyin Wen, Zhenfang Chen

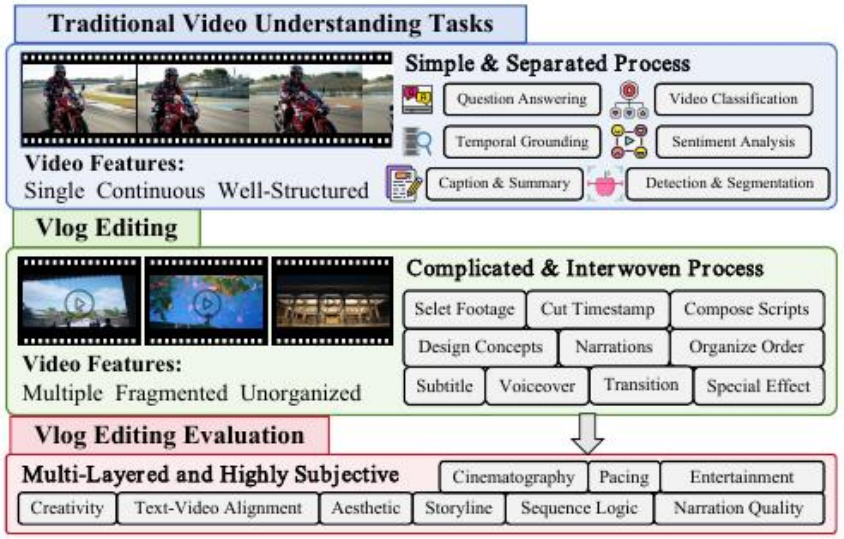
Key question: can multimodal models evaluate and improve vlog edit plans?

Motivation: Why Vlog Editing Needs a Reward Model

Core challenge

Vlog quality is not just video understanding; it is an editing judgment.

- Raw clips are fragmented, redundant, and weakly organized.
- An edit plan combines shot order, cut timestamps, subtitles, voice-over, and pacing.
- The key criteria include both factual alignment and subjective aesthetics.



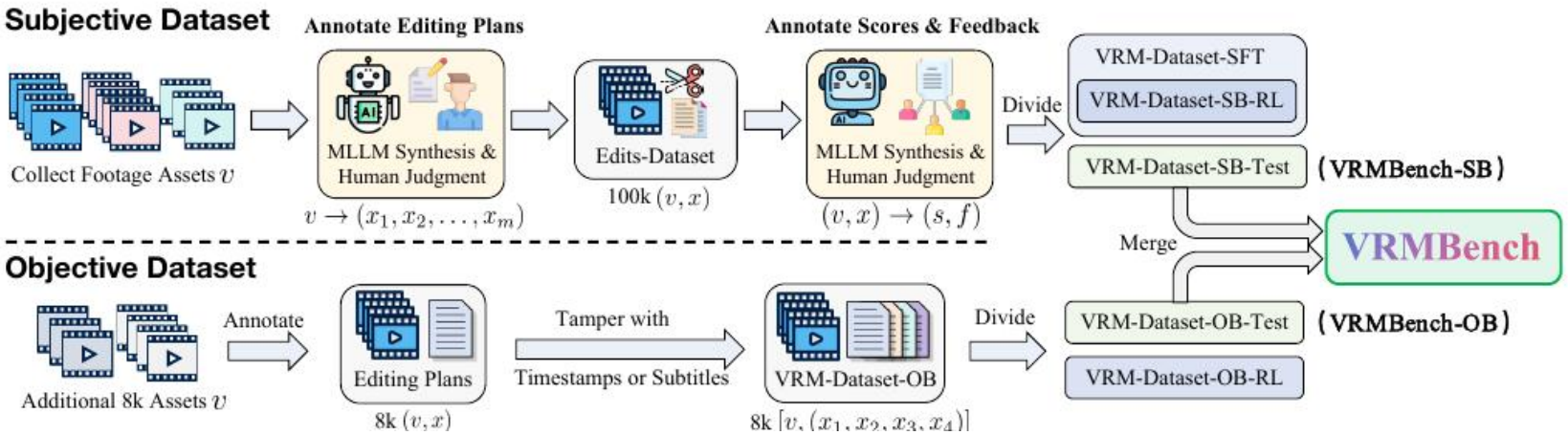
Vlog editing is more interwoven and subjective than standard video understanding.

Six-dimensional evaluation rubric

Creativity | Consistency | Concept Design | Cinematography | Narration | Pacing

Goal: produce fine-grained scores and actionable feedback for each candidate edit plan.

Framework & Data: From Raw Footage to VRMBench



Data collection pipeline: subjective edits, objective tampering, and benchmark construction.

Reward model output

M(v, x) -> (s, f)

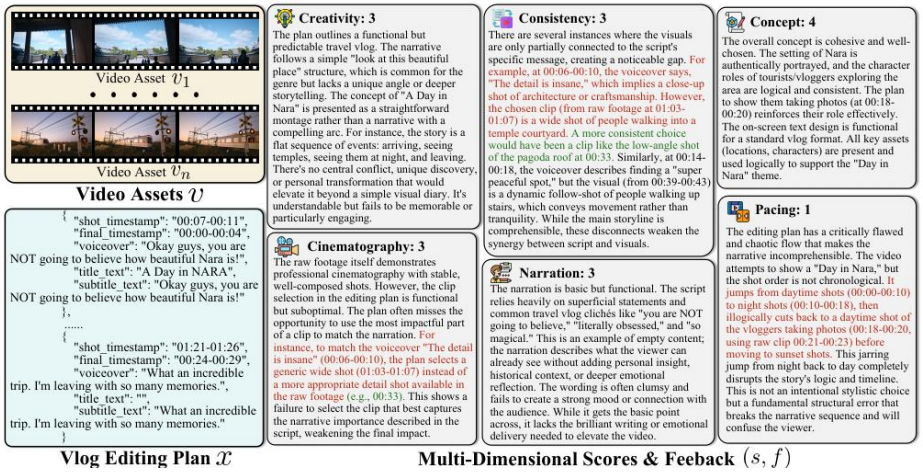
- s: six integer scores from 1 to 5
- f: diagnostic feedback per dimension
- Used for selection, RL reward signals, and iterative refinement

Dataset scale

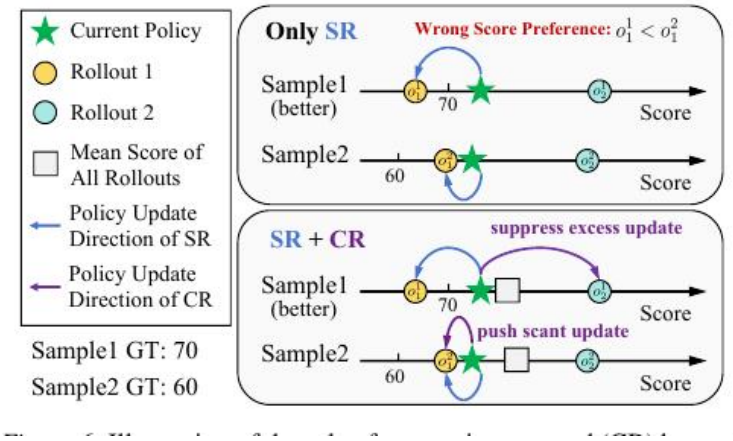
100k SFT edits
40k RL edits
400 benchmark groups / 1.6k plans

VRMBench contains subjective cases and objective factual-error cases.

Method: SFT + GRPO with Inter-Group Comparison Reward



Input-output example: raw assets and a JSON plan are mapped to scores and feedback.



Comparison reward adds preference direction across plans.

Training pipeline

- Cold start with supervised fine-tuning.
- Then train with GRPO on groups of four plans.
- Base model: Qwen2.5-VL-7B-Instruct.

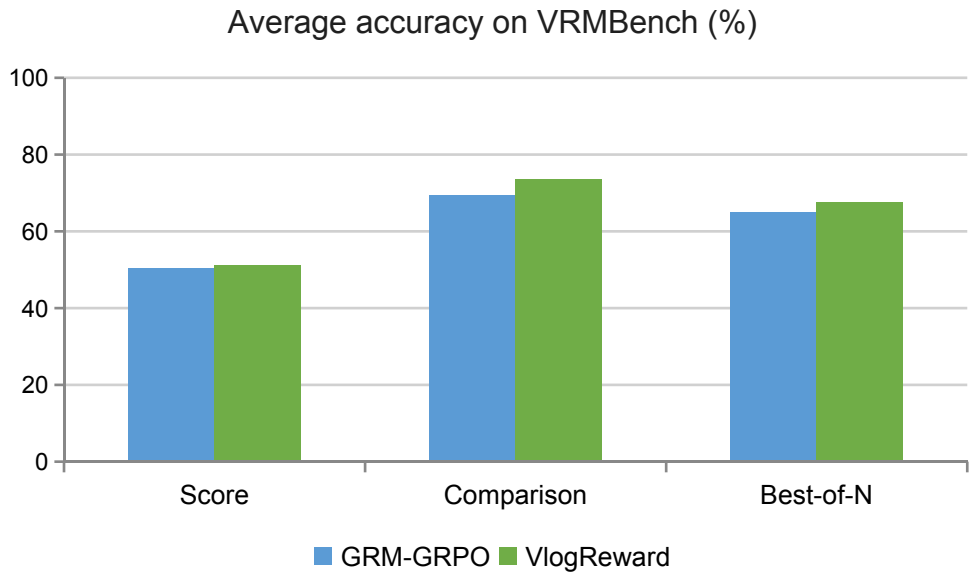
Total reward

```
r_total = r_format * [(1 - lambda) * r_score
+ lambda * r_comp]
```

- Score reward aligns each rubric score with the label.
- Comparison reward enforces better plans receiving higher total scores.
- This mitigates direction blindness in score-only GRPO.

Results: State-of-the-Art Vlog Rewarding

Qualitative example: VlogReward detects chronological disorder while the base model gives identical scores.



Key numbers

VlogReward average on VRMBench:
Score accuracy: 51.3%
Comparison accuracy: 73.5%
Best-of-N accuracy: 67.6%

Effect of comparison reward:
+4.1 points in comparison accuracy
+2.6 points in Best-of-N accuracy

Takeaways and Implications

Three main takeaways

1. Vlog editing evaluation is a distinct multimodal reward modeling task.
2. A six-dimensional rubric enables interpretable scoring and feedback.
3. Inter-group comparison reward improves preference discrimination.

Downstream value

Best-of-8 selection improves perceived edit quality in the user study.
Textual feedback supports iterative refinement of editing plans.
The comparison reward also transfers to general preference learning.

Thanks