← Back to **Author Console** (/group?id=thecvf.com/CVPR/2026/Conference/Authors#your-submissions)

MapReduce LoRA: Advancing the Pareto Front in Multi-Preference Optimization for Generative Models

(/pdf?id=UoYNZmraMA)

Chieh-Yun Chen (/profile?id=~Chieh-Yun_Chen1), Zhonghao Wang (/profile?id=~Zhonghao_Wang6), Qi Chen (/profile?id=~Qi_Chen3), Zhifan Ye (/profile?id=~Zhifan_Ye1), Min Shi (/profile?id=~Min_Shi2), Yue Zhao (/profile?id=~Yue_Zhao24), Yinan Zhao (/profile?id=~Yinan_Zhao1), Hui Qu (/profile?id=~Hui_Qu1), Wei-An Lin (/profile?id=~Wei-An_Lin1), Yiru Shen (/profile?id=~Yiru_Shen2), Ajinkya Kale (/profile?id=~Ajinkya_Kale1), Irfan Essa (/profile?id=~Irfan_Essa1), Humphrey Shi (/profile?id=~Humphrey_Shi1)

08 Dec 2025 (modified: 11 Dec 2025) CVPR 2026 Conference Submission
Conference, Senior Area Chairs, Area Chairs, Reviewers, Authors
Revisions (/revisions?id=UoYNZmraMA)

Supplementary Material: pdf (/attachment?id=UoYNZmraMA&name=supplementary_material)
Subject Area: Image and video synthesis and generation
Keywords: Multi-preference optimization, generative models, LoRA
Student Paper: Yes
External Links: I confirm that the paper submission and supplementary material contain no external links intended to expand content and circumvent review limitations, and no prompt injection or similar attempts to manipulate reviews.

Abstract:
Reinforcement learning from human feedback (RLHF) with reward models has advanced alignment of generative models to human aesthetic and perceptual preferences. However, jointly optimizing multiple rewards often incurs an alignment tax—improving one dimension while degrading others. To address this, we introduce two complementary methods: MapReduce LoRA and Reward-aware Token Embedding (RaTE). MapReduce LoRA trains preference-specific LoRA experts in parallel and iteratively merges them to refine a shared base model; RaTE learns reward-specific token embeddings that compose at inference for flexible preference control. Experiments on Text-to-Image generation (Stable Diffusion 3.5 Medium and FLUX.1-dev) show improvements of 36.1%, 4.6%, and 55.7%, and 32.7%, 4.3%, and 67.1% on GenEval, PickScore, and OCR, respectively. On Text-to-Video generation (HunyuanVideo), visual and motion quality improve by 48.1% and 90.0%, respectively. Our framework sets a new state-of-the-art multi-preference alignment recipe across modalities.

Compute Report: pdf (/attachment?id=UoYNZmraMA&name=compute_report)
Original Submission Link: /revisions?id=5zD8ZbctK7 (/revisions?id=5zD8ZbctK7)
Submission Number: 39389

Filter by reply type... Filter by author... Search keywords... Sort: Newest First

Everyone | Program Chairs | Submission39389... | Submission39389... | Submission39389... | Submission39389... | Submission39389... | Submission39389... | Submission39389... | ✖

3 / 3 replies shown

Add: **Withdrawal** **Rebuttal**

Review - MapReduce LoRA: Advancing the Pareto Front in Multi-Preference Optimization for Generative Models

Official Review by Reviewer E5EG 13 Jan 2026, 13:27 (modified: 22 Jan 2026, 10:55)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Reviewer E5EG, Authors

Revisions (/revisions?id=YFaSDNJAFS)

Paper Summary:
The paper tackles the alignment issue with multiple objectives in generative models when using RL, where improving one objective often degrades performance on others. To solve this, the authors propose MapReduce LoRA, a framework that trains specialized expert models in parallel (Map) and iteratively merges them into the base model (Reduce).

They also introduce RaTE, a method that allows users to control these preferences at inference time using special tokens.

They show experimentally that this iterative approach works significantly better than standard "one-shot" merging (Rewarded soup), and achieves large gains on multiple benchmarks.

Paper Strengths:
1. The idea of framing the merging process as an iterative "MapReduce" cycle is very clever. They show mathematically that this "progressive souping" is theoretically better than standard merging (Rewarded Soups).
2. Strong performance: The quantitative results are really strong, especially on the video tasks. Getting a +90% improvement on Motion Quality is huge and shows the method is very effective.
3. Generalization: It is interesting to see that the model improves even on metrics that were not used during training (untargeted rewards like VQAScore). This suggests the model is robust and not just overfitting to the rewards.

Major Weaknesses:
1. The authors should mention similar approaches in a different context (DiLoCo, FedAvg), which are similar in spirit (averaging models every N steps), but not in the main goal (they target federated learning, i.e. learning on multiple GPU clusters).
2. Experiments without Lora: Would the method still work while fine-tuning the whole network?

Minor Weaknesses:
The method and experiments on RATE are a small portion of the paper; this method could have been its own work.

Preliminary Recommendation: 5: Weak Accept
Justification For Recommendation And Suggestions For Rebuttal:
The strengths of the paper justify the ratings: the performances are strong, the method is quite simple (a generalization of rewarded soups, with more steps). The authors justify the results with many experiments.

Confidence Level: 4: High Confidence - The reviewer has strong expertise in the area. They are highly familiar with the relevant literature and can critically evaluate the paper.

Comments

Official Review by Reviewer k4GD 08 Jan 2026, 05:11 (modified: 22 Jan 2026, 10:55)

Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Reviewer k4GD, Authors

Revisions (/revisions?id=4K8Lhwzrkx)

Paper Summary:
This paper addresses the critical challenge of "alignment tax" in multi-preference post-training for generative models—where optimizing one objective (e.g., text-image alignment) often degrades others (e.g., aesthetic quality). To tackle this, the authors propose two complementary methods: MapReduce LoRA and Reward-aware Token Embedding (RaTE).

Paper Strengths:
1. The paper's greatest strength is its simple yet powerful approach to multi-preference optimization.
2. RaTE fills a critical need for flexible preference customization. By distilling expert knowledge into task-specific tokens, the method enables users to adjust reward trade-offs (e.g., prioritizing text alignment over aesthetics) by appending tokens to prompts.
3. The experiments are rigorous and cover multiple modalities (image, video) and base model.

Major Weaknesses:
1. MapReduce LoRA: Iterative merging of expert models echoes ideas from "skill merging" (e.g., SELMA) and weight interpolation (e.g., Rewarded Soup). The authors frame it as a "progressive souping" process, but the core mechanism—training experts in parallel and merging iteratively—has been explored in multi-objective RL for generative models.
2. RaTE: Distilling expert knowledge into token embeddings is closely related to Textual Inversion and LoRA-based style control. The paper acknowledges Textual Inversion as inspiration but does not clearly articulate how RaTE differs from prior token-based customization methods.
3. Scalability to More Rewards: The experiments use 2–3 rewards. Can MapReduce LoRA scale to 5+ diverse rewards without diminishing returns or increased training cost?
4. SELMA also trains skill-specific experts and merges them for text-to-image generation. The paper does not compare against SELMA, despite overlapping goals (multi-skill alignment) and methods (expert merging).

Minor Weaknesses:
N/A

Preliminary Recommendation: 4: Borderline Accept
Justification For Recommendation And Suggestions For Rebuttal:
See Weaknesses.

Confidence Level: 4: High Confidence - The reviewer has strong expertise in the area. They are highly familiar with the relevant literature and can critically evaluate the paper.

Simple and Effective Multi-Preference Alignment with Questions on RaTE Novelty

Official Review by Reviewer d7Kb 07 Jan 2026, 05:24 (modified: 22 Jan 2026, 10:55)
Program Chairs, Senior Area Chairs, Area Chairs, Reviewers Submitted, Reviewer d7Kb, Authors
Revisions (/revisions?id=qRzyPa0kSe)

Paper Summary:
This paper proposes MapReduce LoRA and Reward-aware Token Embedding (RaTE) to address the multi-objective optimization problem in generative models, where improving one reward often degrades others. MapReduce LoRA trains reward-specific LoRA experts, each optimized for a single reward, in parallel and iteratively merges them to refine a shared base model, which can be interpreted as a progressive weight-souping process. In addition, RaTE distills each reward-specific expert into a lightweight learned token embedding, enabling flexible and composable preference control at inference time without retraining the model. Together, these components provide a simple and scalable alternative to reward scalarization and one-shot model merging, and show consistent improvements across multiple preferences and modalities.

Paper Strengths:
1. Consistent improvements across modalities and base models. The proposed approach outperforms baseline methods across image generation, video generation, and language tasks, including weighted-reward scalarization methods, and one-shot merging baselines. The breadth of evaluation suggests that the approach

generalizes beyond a single task or modality.

2. Simple yet effective approach to multi-objective alignment. A key strength of this paper is the simplicity of the proposed MapReduce LoRA framework. Rather than introducing complex optimization machinery, the method alternates between training reward-specific LoRA experts and merging them through simple averaging. This design is easy to implement on top of existing RLHF pipelines and provides a clear and intuitive alternative to reward scalarization and one-shot model merging.

Major Weaknesses:

- **Limited novelty of RaTE.**
  RaTE is inspired by textual inversion and follows a similar paradigm of learning trainable token embeddings with a frozen backbone. While the method applies this idea to encode reward-specific preferences, the paper does not clearly explain what fundamentally new insight or capability is introduced beyond this adaptation, making the novelty of RaTE appear incremental.
- **Insufficient justification for RaTE over LoRA-based control.**
  Since reward-specific LoRA experts are already trained, it is unclear why learning additional token embeddings is preferable to directly using or composing LoRA adapters at inference time. A direct comparison in terms of efficiency or performance would strengthen the motivation for RaTE.
- **Potentially misleading presentation of MORL baselines.**
  The MORL baselines reported in Table 1 (e.g., MORL-D and MORL-DR) are not sufficiently explained and may be misleading given the terminology used in the related work section and Figure 3. It is unclear whether these baselines correspond to Flow-GRPO trained with scalarized multi-reward objectives, mixed datasets, or other multi-objective training strategies. Clearer alignment between the terminology, figures, and experimental setup is necessary to properly interpret these comparisons.

Minor Weaknesses:

- Important hyperparameters, such as reward-specific optimization steps and the KL regularization coefficient β, vary across tasks and models. It would be helpful to clarify whether there is any general guidance or analysis for selecting these values, and whether the method is robust to such choices or potentially sensitive to task-specific tuning.
- Eq. (1) uses the likelihood ratio term r_t^g without explicitly defining it in the main text, and a brief clarification would improve readability.

Preliminary Recommendation: 3: Borderline Reject

Justification For Recommendation And Suggestions For Rebuttal:

The paper proposes a simple and practical framework for multi-objective alignment, and MapReduce LoRA demonstrates consistent improvements over scalarized multi-reward training and one-shot merging baselines across multiple modalities. The approach is intuitive and empirically effective. My main concern lies in the **novelty and motivation of RaTE**. RaTE closely follows the paradigm of textual inversion by learning trainable token embeddings to control model behavior, and the paper does not clearly explain what fundamentally new insight or capability is introduced beyond applying this idea to reward-specific preferences. Moreover, since reward-specific LoRA experts are already trained, it remains unclear why learning additional token embeddings is preferable to directly using or composing LoRA adapters at inference time. Clearer justification of RaTE's advantages, or a simple comparison with LoRA-based control, would significantly strengthen this part of the paper. If the concerns regarding the novelty and motivation of RaTE are adequately clarified, I would be open to reconsidering my score after reviewing the rebuttal and other reviewers' comments.

Confidence Level: 3: Moderate Confidence - The reviewer is reasonably knowledgeable about the topic. They understand the paper's methodology and results but may not be a leading expert in the specific subfield.