ReasonEdit: Editing Vision-Language Models Using Human Reasoning

ICML 2026

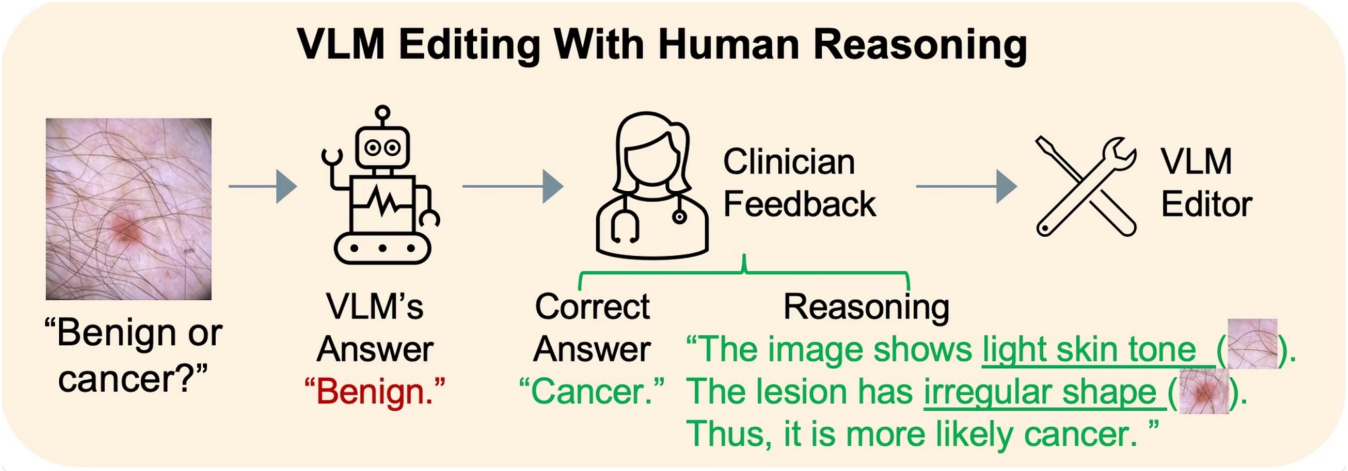
JIAXING QIU[1], KAIHUA HAO[2], ROXANA DANESHJOU[3], AHMED ALAA[2], THOMAS HARTVIGSEN[1]

[1] University of Virginia, Charlottesville, VA, USA

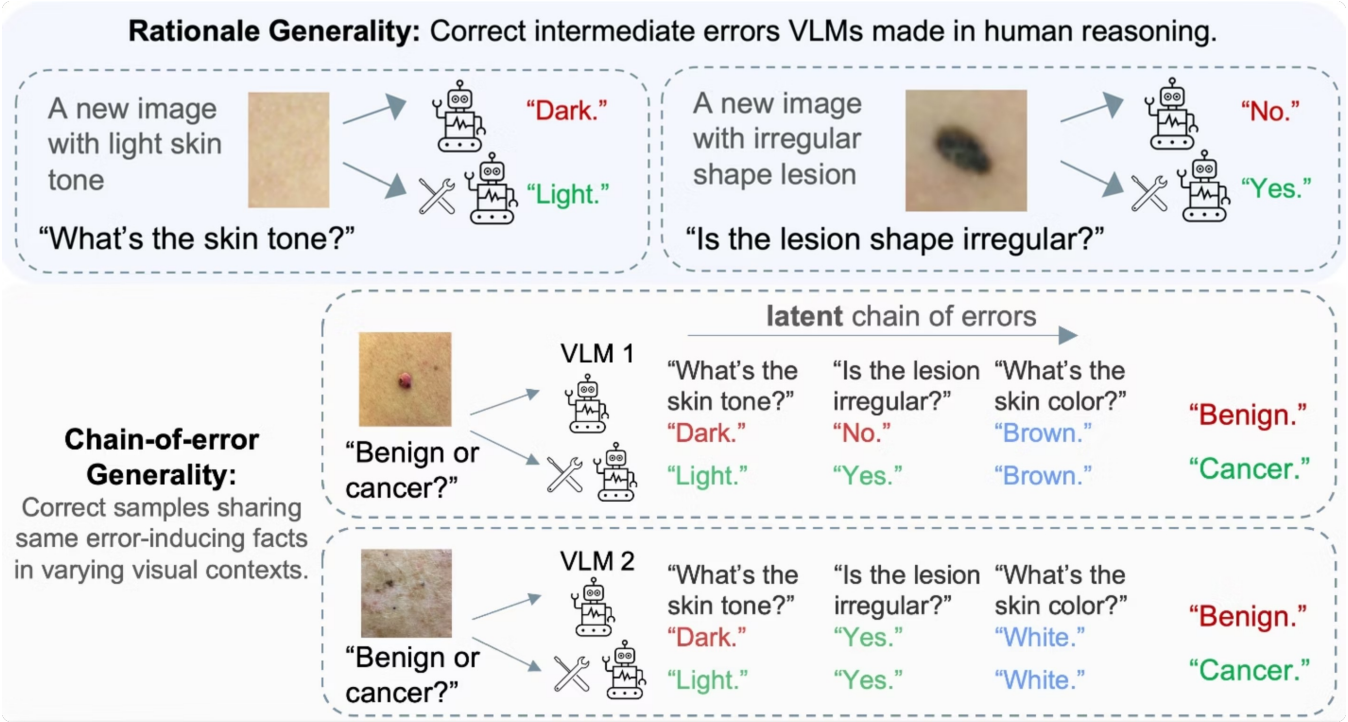
[2] University of California, Berkeley, Berkeley, CA, USA

[3] Stanford University, Stanford, CA, USA

The Problem: VLMs Fail on Reasoning-Heavy Tasks

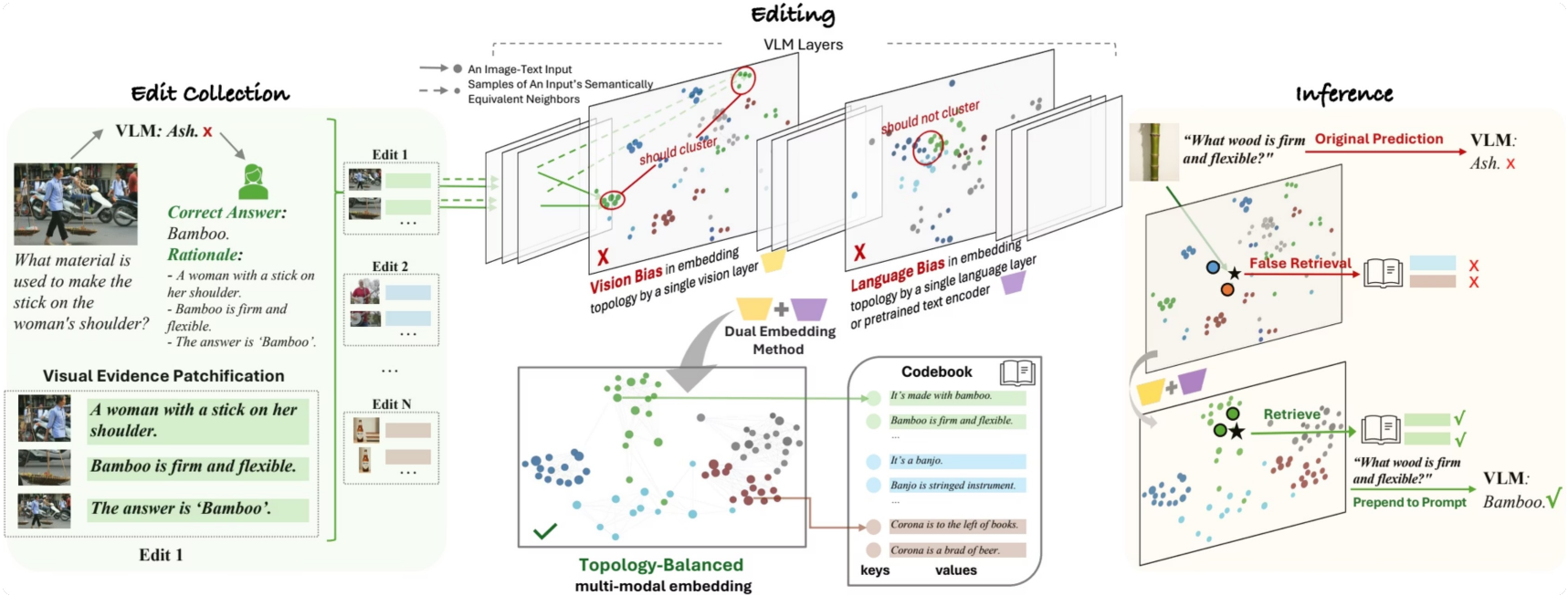


Human Reasoning Enables New Forms of Generalization



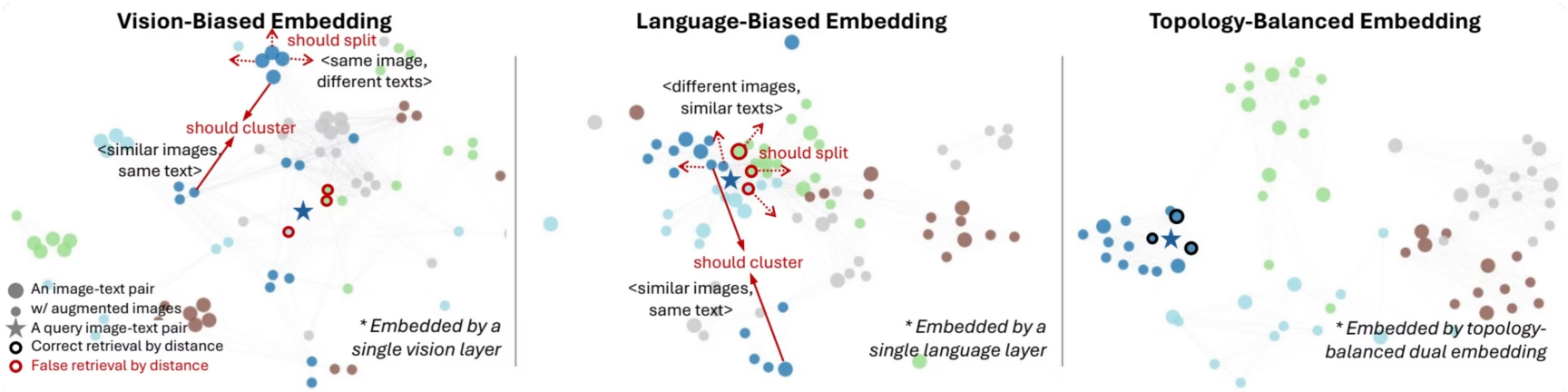
ReasonEdit: The First Reasoning-Enhanced VLM Editor

ReasonEdit avoids weight updates. Each edit becomes <**key** = image-text embedding, **value** = reasoning statement> entries stored in a **codebook**, retrieved at inference time as context in prompts.



The “Key” of ReasonEdit: Topology-Aware Multimodal Embedding

- **Single-layer VLM embeddings are biased** — vision layers cluster by image, language layers by text.
- ReasonEdit proposes thinking of relationships between image-text embeddings as a **graph**, using **Newman's modularity** as a principled criterion to **evaluate and select** multimodal embeddings for balanced joint-modality topology.



Vision Modularity

Measures alignment with image-only similarity.

Language Modularity

Measures alignment with text-only similarity.

Bimodal Modularity

Measures joint image-text clustering — used to select the optimal vision layer.

State-of-the-Art Results

Editing Performance

- ReasonEdit outperforms baseline editors across **VLMs** and **datasets**.
- It achieves high locality through its retrieval-based design.
- It achieves high image-/text-generality, due to its **topology-balanced multimodal embedding** method.
- Human reasoning enables new **reasoning-based generality**.

Stable Sequential Editing Performance

- ReasonEdit maintains consistently good performance over 5,000+ sequential edits, while weight-updating methods suffer catastrophic degradation.

Robust to Noisy Reasoning

- ReasonEdit is robust to injected noise into the reasoning.