PRPO: Paragraph-level Policy Optimization for Vision-Language Deepfake Detection

Tuan Nguyen[1], Naseem Khan[1], Khang Tran[2], Hai Phan[2], Issa Khalil[1]
[1]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[2]New Jersey Institute of Technology, NJ, USA

Keywords: *deepfake detection, vision language models, paragraph-level reasoning.*

Motivation

- Deepfakes are now near-indistinguishable from real images.
- Existing detectors give only a binary verdict - no “why”.
- Multimodal Large Language Models (MLLMs) can explain, but they hallucinate, reason shallowly, and contradict the image.

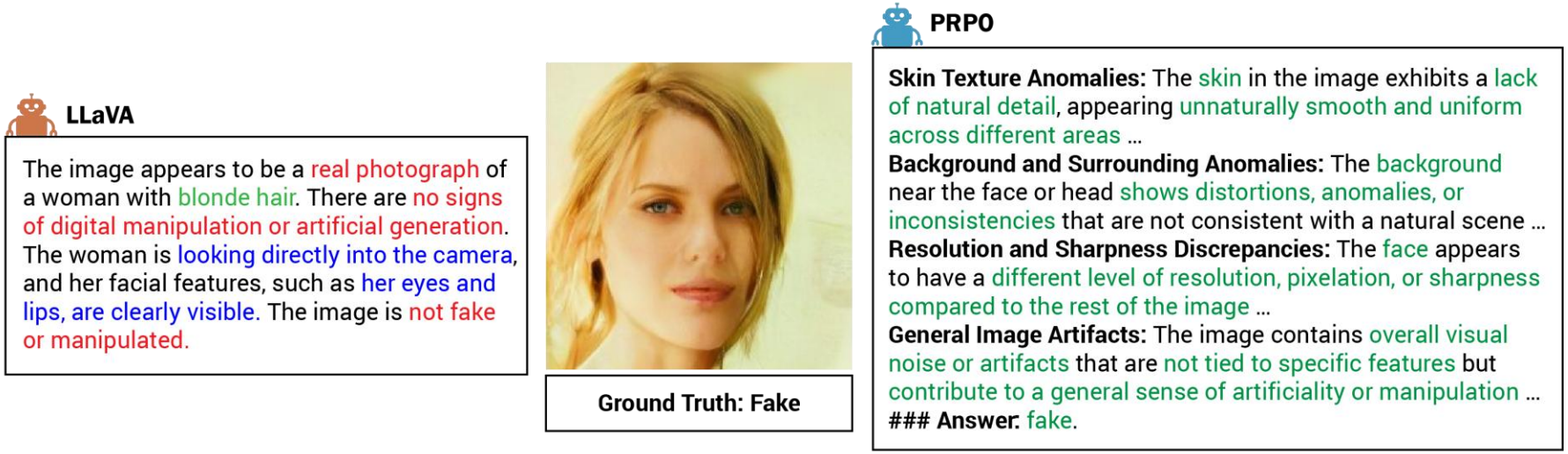


Figure 1: Reasoning quality comparison between LLaVA and the proposed PRPO.

Can a vision-language model detect deepfakes and produce faithful, evidence-grounded explanations?

Contributions

1. Dataset	• **DF-R5**: 115K image-reasoning pairs across 5 generative domains, with rich annotated reasoning.
2. Model	• **DX-LLaVA:** A detection + explainability framework: CLIP ConvNeXT vision encoder paired with a Vicuna LLM.
3. Algorithm	• **PRPO:** Paragraph-level Relative Policy Optimization which aligns each reasoning paragraph with visual evidence.
4. Results	• **New SOTA: +14.65%** F1 over prior SOTA, and a **4.55 / 5** explanation-quality score.

Dataset – **DF-R5**

- Built on the DF40 benchmark across five diverse generative domains, chosen to maximize difficulty and cross-domain generalization.

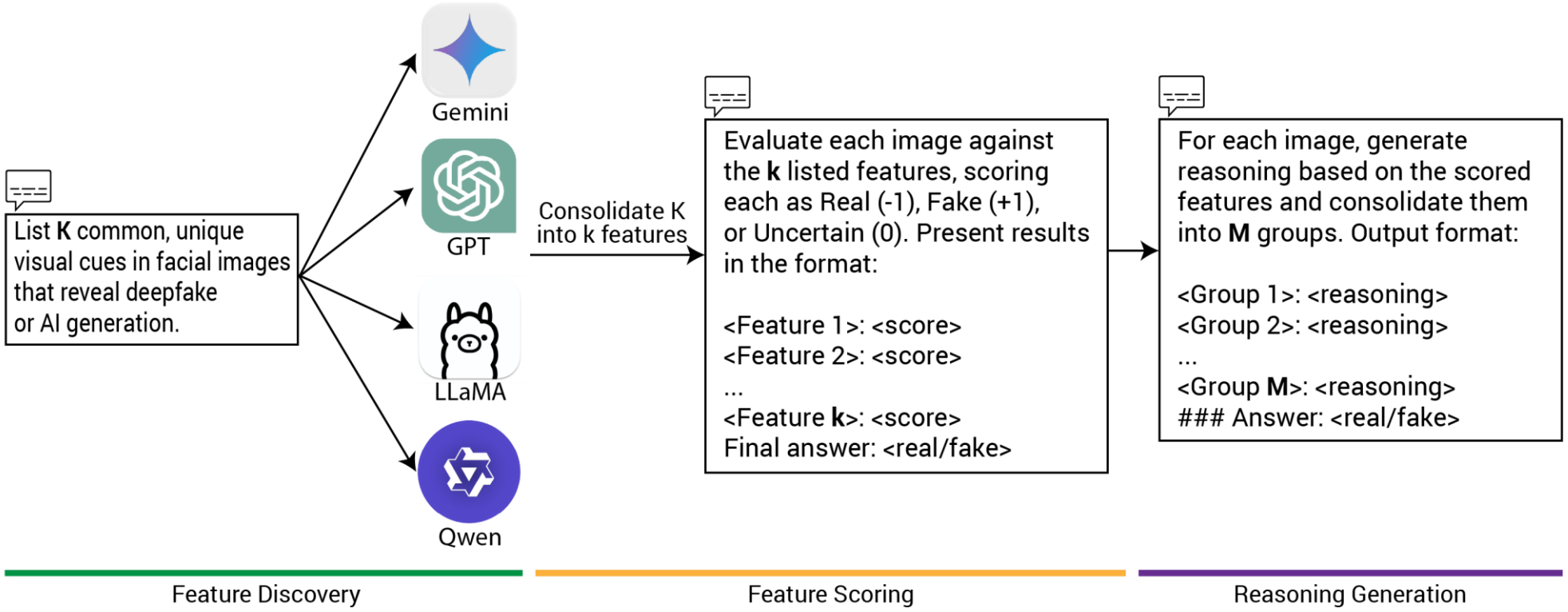


Figure 2: Three-step pipeline for generating high-quality reasoning annotations in DF-R5.

Dataset – **DF-R5**

- Built on the DF40 benchmark across five diverse generative domains, chosen to maximize difficulty and cross-domain generalization.

Skin Detail and Texture Issues: Observing the image, the skin appears abnormally smooth or plastic-like, lacking the typical variations in texture and detailed pore structure seen in natural skin.

Hair and Eyebrow Blending Issues: The eyebrows and hair exhibit unnatural blending with the surrounding skin and background.

Structural Anomalies: The facial structure shows signs of potentially excessive or unnatural symmetry or asymmetry which deviates from typical human variation.

Boundary and Transition Issues: The edges of the face, where it meets the neck, hair, or background, are poorly defined, overly sharp, or show visible signs of manipulation such as blending artifacts or ghosting.

Answer: fake

Figure 3: Sample reasoning annotation from the DF-R5 dataset.

Architecture – **DX-LLaVA**

- CLIP ViT captures global semantics - great for VQA, but it collapses on unseen domains, predicting almost everything as “real”.
- Deepfake detection needs sensitivity to local, high-frequency artifacts .

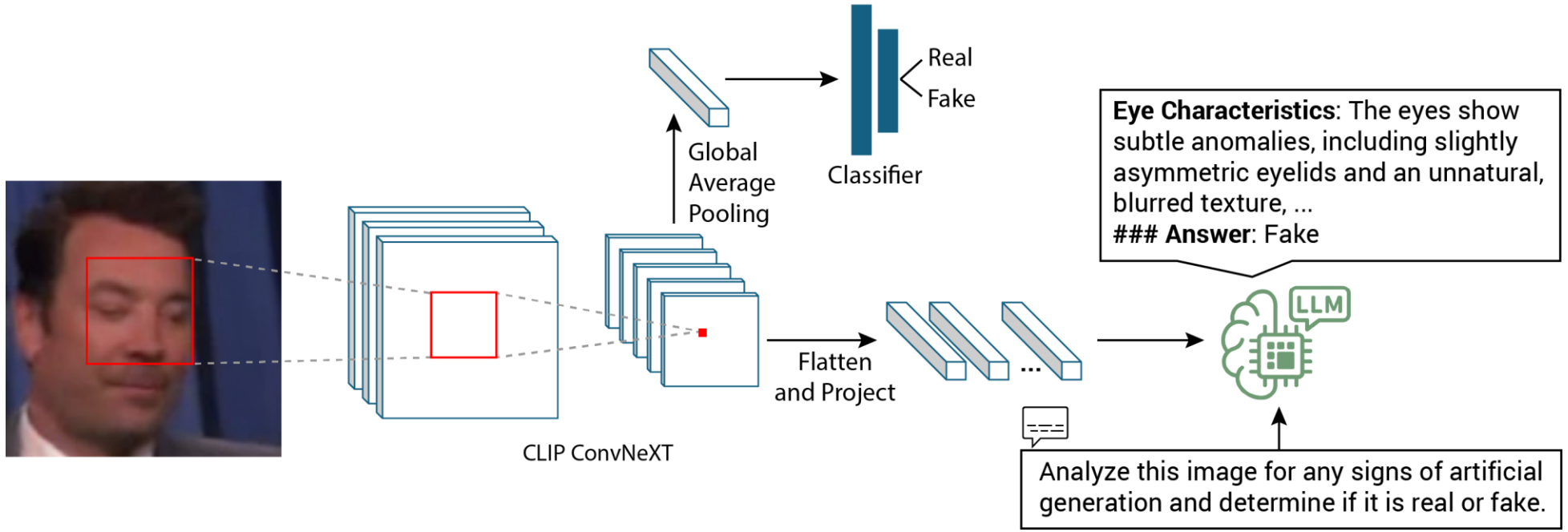
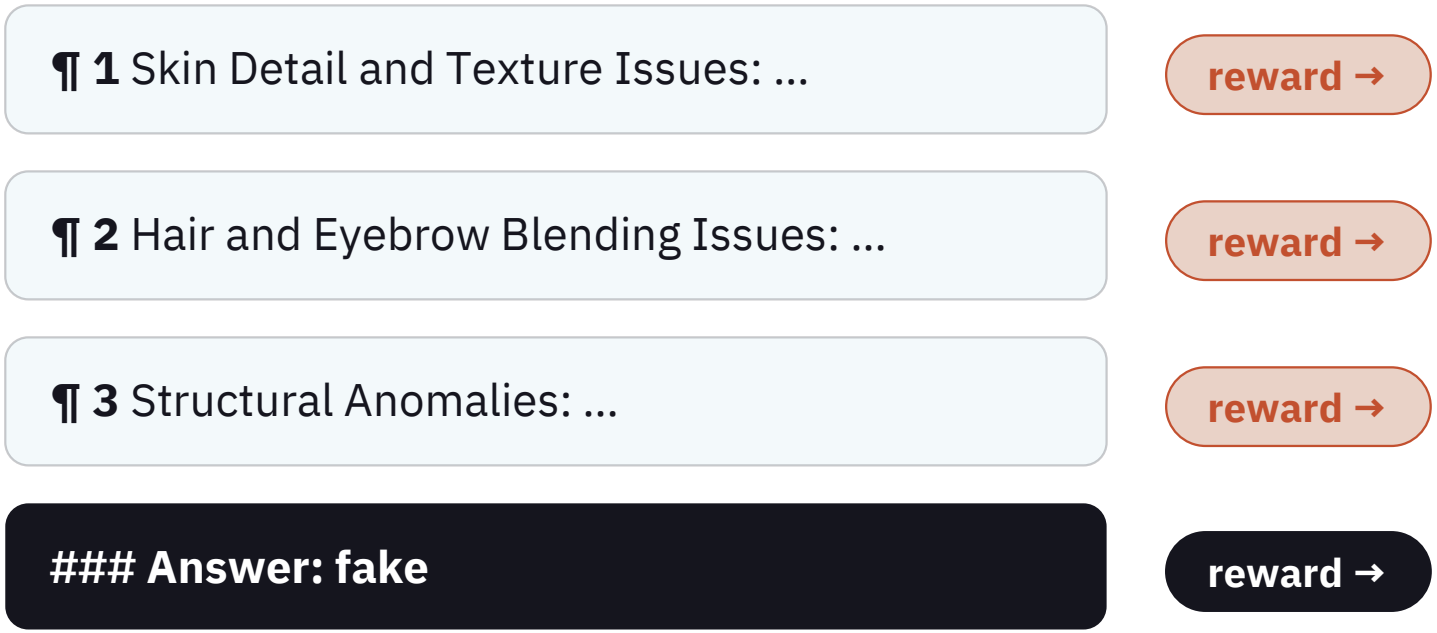


Figure 4: Proposed DX-LLaVA architecture.

Algorithm – **PRPO**

- Generated reasoning can still be hallucinated, irrelevant, or inconsistent with the final prediction.
- Treat reasoning as **paragraphs,** and give every paragraph its own reward, instead of rewarding only the final answer.

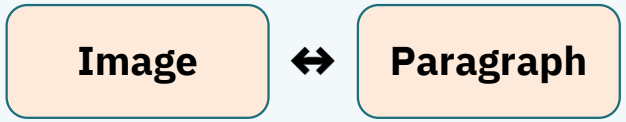


Algorithm – **Reward Design**

REWARD 1

Visual Consistency (VCR)

Does each paragraph actually match the image?



CLIP ConvNeXT embeddings + YAKE keyword extraction → cosine similarity, normalized to [0,1].

REWARD 2

Prediction Consistency (PCR)

Does the verdict agree with its own paragraphs?

Reward = 1 if the conclusion matches the paragraph majority vote, else 0.

- No ground-truth labels at reward time, encouraging generalization to unseen images.

Results

Detection

- A new state of the art on unseen domains

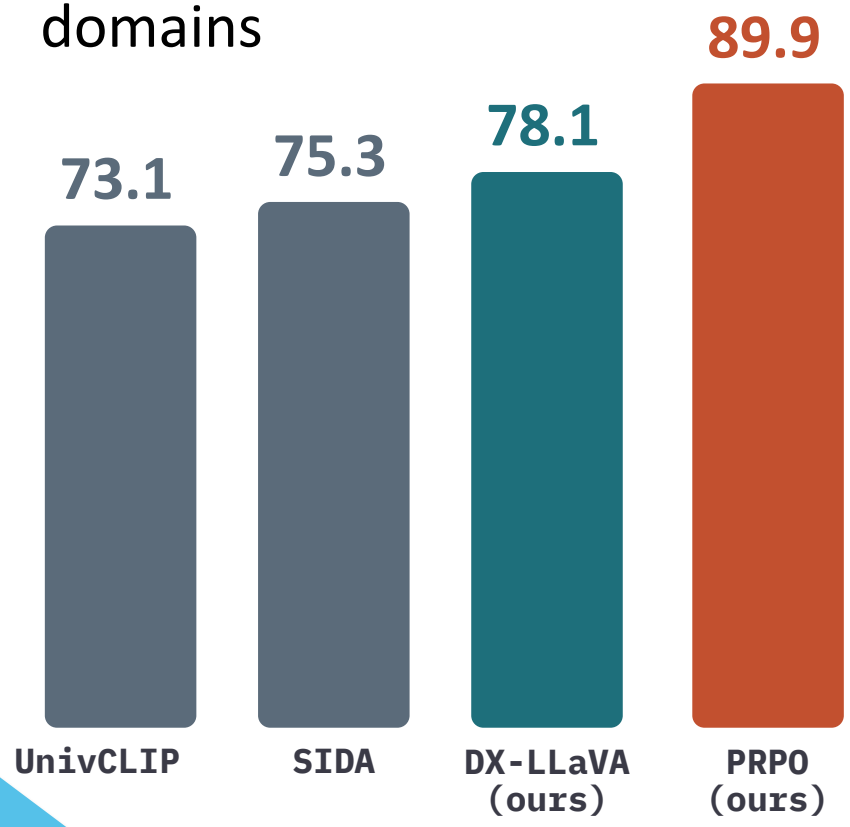


Figure 4: Average F1 Score Across Five Held-Out Domains (%)

Explanation Quality

- PRPO surpasses Gemini-2.5 on evidence grounding and classification consistency.

MODEL	OVERALL
DX-LLaVA (ours)	4.02
Gemini-2.5	4.20
PRPO (ours)	**4.55**

Table 1: Reasoning quality evaluation conducted by GPT-4o.

Key Takeaways and Future Work

- DF-R5: first large-scale deepfake dataset with reasoning annotations.
- DX-LLaVA: ConvNeXT with joint loss fixes cross-domain detection.
- PRPO: paragraph rewards align reasoning with visual evidence at test time.
- New SOTA: 89.9% F1 and 4.55/5 explanation quality.
- Extend PRPO's paragraph-level rewards to other multimodal reasoning - VQA, visual entailment, and beyond.

Paper

Code & Dataset

Thank You for Watching!