

All-Weather VLM: Enhancing Vision Language Models' Robustness Under Adverse Imaging Conditions

Tianfu Wang^{1*} Mingyang Xie^{1*} Haoming Cai¹ Tianyi Xiong¹ Xiyao Wang¹ Dongdong Fu² Guan-Ming Su² Paola Cascante-Bonilla^{1,3†} Christopher Metzler^{1†}

* Equal contribution † Equal advising ¹ University of Maryland ² Dolby Laboratories Inc. ³ Stony Brook University



Project webpage



1 Motivation

Can VLMs answer visual questions reliably under adverse imaging conditions?

Restoring the image first can still produce wrong answers. We train the VLM to reason through degradation.

Q: How many giraffes are in the image?

Correct answer: 1 giraffe

Base VLM
Rainy image

Restore, then VLM
RESTORED IMAGE

Ours
Same rainy image

0 giraffes
Incorrect

0 giraffes
Still wrong

1 giraffe
Correct

Restoration removes the giraffe: the VLM still fails. Ours answers correctly from the original rainy image.

Q: What is the highway speed limit?

Correct answer: 120 km/h

Base VLM
Original foggy image

Restore, then VLM
RESTORED IMAGE

Ours
Original foggy image

100 km/h: wrong

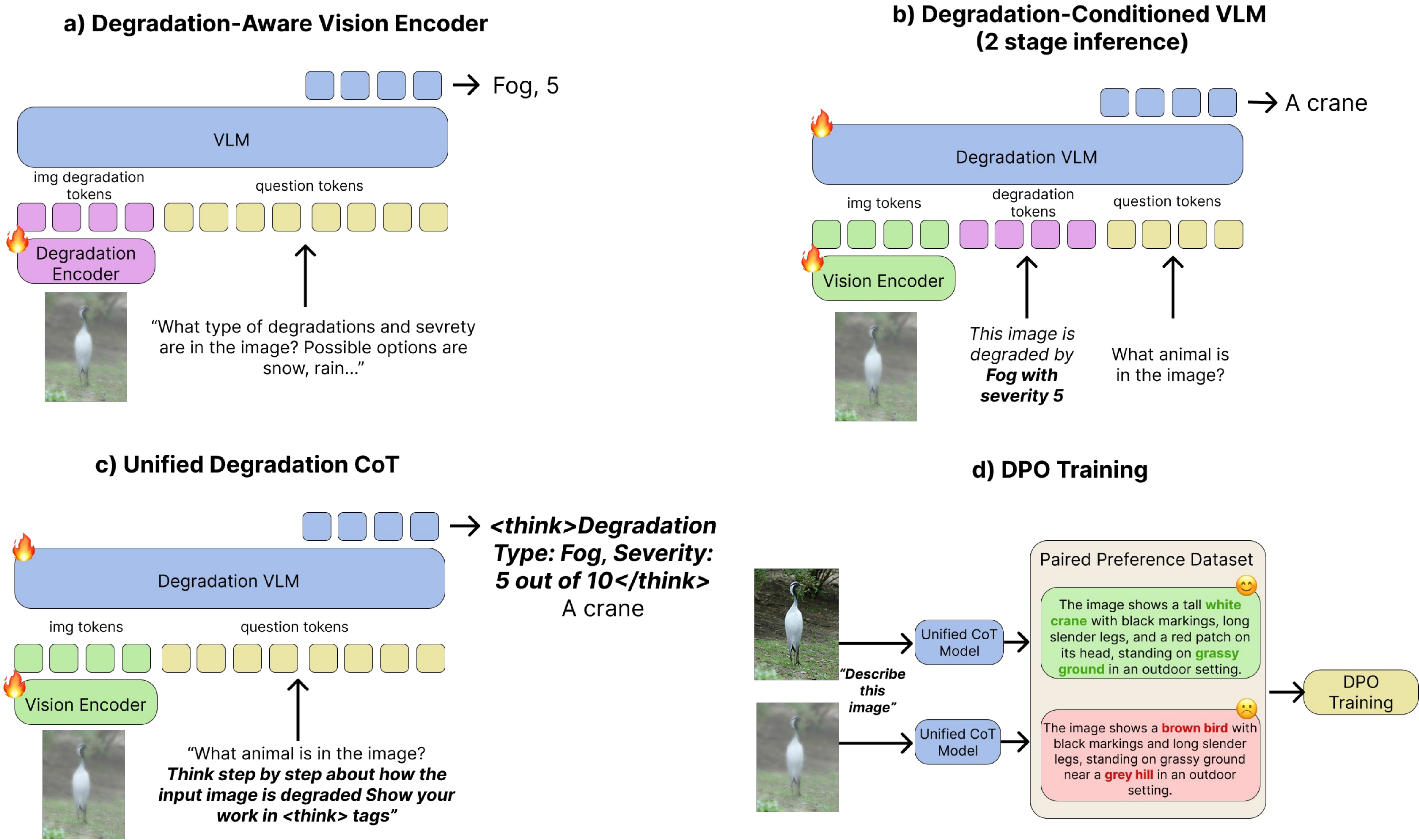
100 km/h: still wrong

120 km/h: correct

2 Degradation-aware training

Qwen2.5-VL-7B learns to identify degradation and use it to condition visual reasoning.

Paired training data: 8 corruption types, severity levels 1–10, plus clean images.



Inference After DPO training (d), we use Unified CoT (c) to reason about degradation type and severity, then answer a question about the image.

Takeaway

Degradation-aware reasoning improves robustness while preserving clean-image capability.

3 Results

Evaluation	Base	Ours
Real driving QA (%)	63.5	73.0
R-Bench overall score	52.2	56.1
RealWorldQA degraded (%)	57.0	61.0
RealWorldQA clean (%)	69.0	70.7

Base: Qwen2.5-VL-7B. Degraded scores average 8 corruption types.

Real-world visual question answering

Q: How many pedestrians are clearly visible?

Base: 1
Ours: 4

Q: How many buses are visible?

Base: 0
Ours: 1

Q: What do the neon lights say?

Base: M it's frensh
Ours: Oh it's fresh

Q: What best describes this road?

Base: One-way road
Ours: 3-way intersection