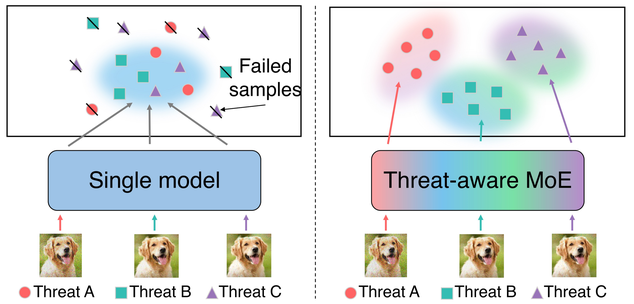
RoME: Robust Mixture of Low-Rank Experts against Multiple Adversarial Perturbations

Woo Jae Kim, Kyle Min, Suhyeon Ha, Joonsung Jeon, Sung-Eui Yoon

ECCV 2026 | MALMÖ | SEPT 8–12

Codes

Multi-Perturbation Robustness

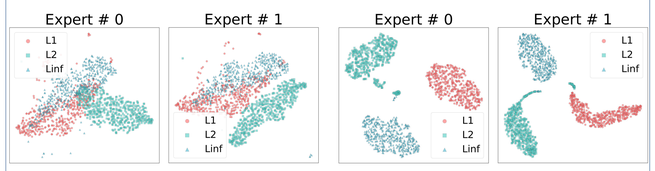


- AT on one threat leaves the model **vulnerable to different threats**
- AT on multiple threats (MAT) causes **cross-threat trade-off**
- Our idea: give each threat **its own model pathway** via a **mixture of experts (MoE)**

Why Naïve MoE Fails

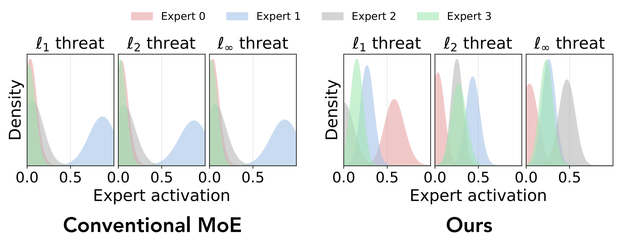
① Redundant Experts

- Experts redundantly capture **threat-common features**



② Threat-Agnostic Routing

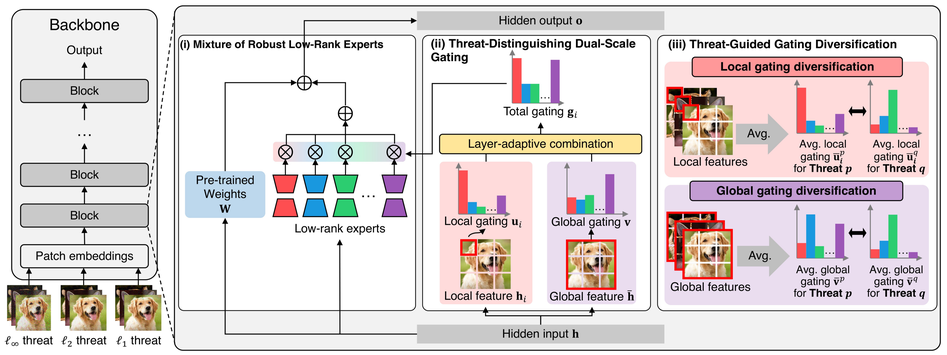
- Gating routes each threat to **nearly identical experts**



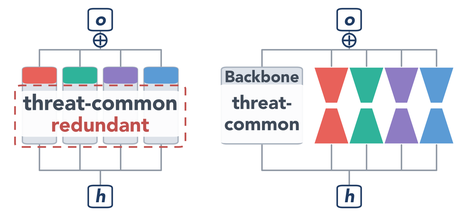
→ **Lacks specialization** of each expert on different threats

RoME: Mixture of Low-Rank Experts

Overall Framework



(i) Mixture of Low-Rank Experts

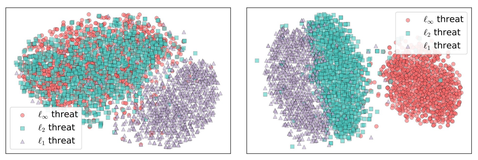


Conventional MoE — Ours (RoME)

- **Low-rank (LoRA) experts** on the shared backbone
- Backbone: **threat-common** features
- Low-rank experts: **threat-specific** features

→ ① **Threat-specialized experts**

(ii) Threat-Distinguishing Dual-Scale Gating

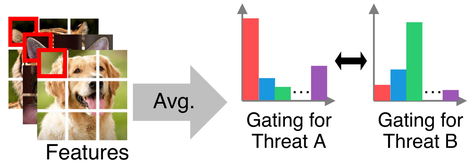


(a) Patch-level — (b) Image-level

- Each threat is distinguishable at **different levels of granularity**
- **Dual-scale gating** at local (patch) and global (image) level

→ ② **Threat-aware routing**

(iii) Threat-Guided Gating Diversification

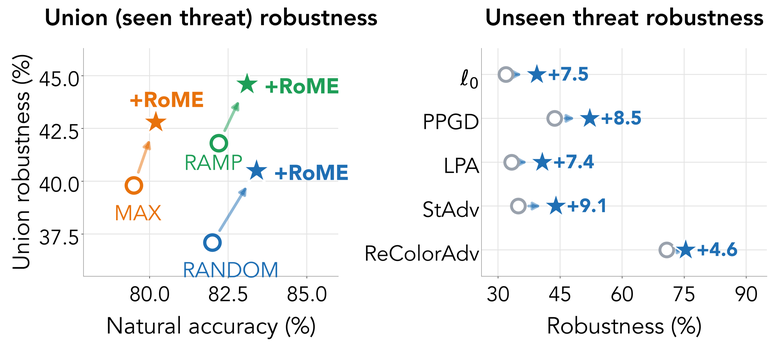


- Maximise pairwise distance between **per-threat gating** patterns
- Enforces **distinct expert combinations** for each threat

→ ② **Threat-aware routing**

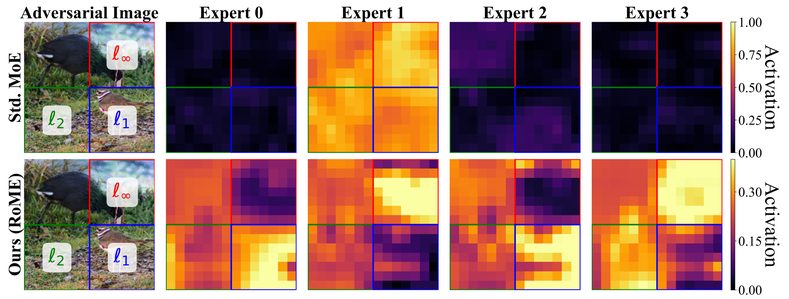
Experimental Evaluations

Robustness



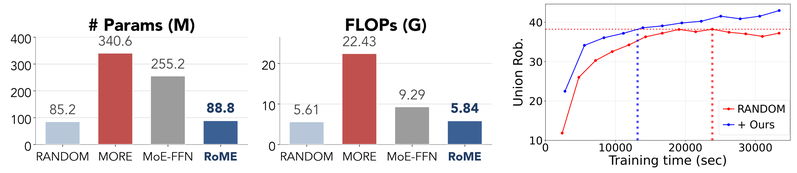
- Improves union robustness **and** natural accuracy on **different MAT methods**
- Also improves robustness against **threats unseen during training**

Learned Expert Routing



- Standard MoE activates the **same** experts for each threat
- RoME activates **different experts per threat**

Computational Cost



- SoTA union robustness at only **1.04× params** and **1.04× FLOPs**