RoCA: Robust Cross-Domain End-to-End Autonomous Driving

Rajeev Yasarla Shizhong Han Hsin-Pai Cheng Apratim Bhattacharyya
Shweta Mahajan Litian Liu Yunxiao Shi Risheek Garrepalli
Hong Cai Fatih Porikli

Qualcomm AI Research*

ICML 2026

One-sentence takeaway

RoCA learns a **probabilistic codebook of driving behaviors** with a Gaussian process, enabling **better cross-domain generalization**, **uncertainty-aware adaptation**, and **no extra inference latency** when used as training regularization.

* Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Motivation: Cross-domain E2E driving remains brittle

- End-to-end planners map **multi-view images** directly to future trajectories.
- In practice, performance drops under **domain shift**: new cities, weather, lighting, camera characteristics, and long-tail driving events.
- Existing datasets are dominated by **common easy scenarios**; rare but safety-critical cases are underrepresented.
- LLM/MLLM-based approaches may help with world knowledge, but **do not guarantee cross-domain robustness** and are expensive to adapt.

Goal

Design an E2E planner that **generalizes across domains** and **adapts efficiently** to new target environments, with or without target labels.

Target scenarios

- Sim $\rightarrow$ real
- Boston $\leftrightarrow$ Singapore
- Low light / blur / snow / fog
- Long-tail maneuvers

Core challenge

How can we exploit similarities to known behaviors while quantifying uncertainty on ambiguous or unseen scenes?

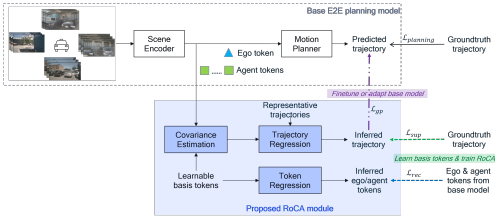
RoCA in one slide: basis tokens + Gaussian process

Key idea

- Learn a **codebook of basis tokens** that spans diverse ego and agent behaviors.
- Pair each basis token with a **representative trajectory**.
- For a new scene, compare its token embedding to the basis codebook and infer trajectories via a **Gaussian process (GP)**.

Why this helps

- **Probabilistic prediction**: not a single hard anchor lookup.
- **Uncertainty estimation**: variance highlights hard or out-of-domain cases.
- **Better adaptation**: uncertainty can prioritize which target samples to learn from.



Framework overview

Base E2E model

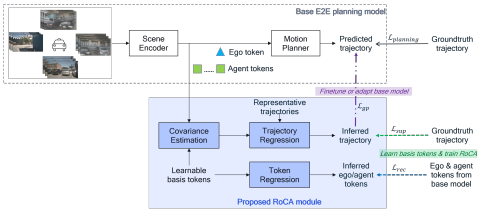
- Scene encoder extracts tokenized scene representation from multi-view images.
- Motion planner predicts:
  - ego future waypoints and class
  - surrounding-agent trajectories and classes

RoCA module

- Learns **basis token groups** and a matching **trajectory codebook**.
- Reconstructs ego/agent tokens and predicts trajectories with a GP.
- Uses predictive variance to **reweight training** and guide adaptation.

Deployment benefit

- When used as a **training regularizer**, the deployed planner is unchanged: **no extra inference cost**.



GP formulation: prediction + uncertainty

Given a token x from the current scene, RoCA first assigns it to a basis group c, then predicts the future trajectory by conditioning on the corresponding basis tokens and codebook trajectories.

GP-based trajectory prediction

$$\hat{p}_w = w_{\text{anchor},c} + \kappa(x, B_c)\,\kappa(B_c)^{-1}\,\bar{W}_c$$
$$\sigma_w^2 = \kappa(x) - \kappa(x, B_c)\,\kappa(B_c)^{-1}\,\kappa(x, B_c)^\top + \sigma_{\text{noise}}^2 I$$

Interpretation

- B_c: selected basis-token group
- $\bar{W}_c$: zero-mean trajectories in that group
- $\kappa(\cdot,\cdot)$: kernel similarity (RBF in the paper)

Why uncertainty matters

- Larger σ_w^2 indicates **ambiguous**, **rare**, or **out-of-domain** scenes.
- RoCA uses this to emphasize difficult samples during source training and target adaptation.

Training and target-domain adaptation

Source-domain training

1. Pretrain the base E2E planner.
2. Learn basis tokens and GP parameters using:
   - token reconstruction loss
   - GP-based trajectory supervision
   - classification + triplet regularization
3. Finetune the base planner with **RoCA as a teacher**.

Target-domain adaptation

- **Supervised**: use target labels when available.
- **Unsupervised**: use GP predictions and uncertainty without labels.
- **Active learning**: prioritize high-variance target samples for annotation.
- Also supports **online adaptation**.

Practical value

RoCA improves generalization and also reduces adaptation cost by identifying the most informative target data.

Main results across domains

Bench2Drive (closed-loop)

Using ORION as the base model, RoCA improves Driving Score from **77.74** to **80.38**, and mean ability from **54.72** to **61.11**.

Sim-to-real: Bench2Drive → nuScenes

With ORION, zero-shot full-val L2 improves from **0.98** to **0.68**. After adaptation, RoCA reaches **0.44** with GT labels and **0.52** without GT labels.

Cross-city: Boston → Singapore

With SparseDrive-S, direct finetuning with GT gives **0.55 m / 0.12%**, while RoCA adaptation with GT improves to **0.49 m / 0.09%**.

Robustness under image degradations

RoCA improves performance under:

- low light
- motion blur
- snow
- fog

It also remains strong under **unsupervised adaptation**.

Takeaway

Across closed-loop, open-loop, sim-to-real, cross-city, and degraded-image settings, RoCA consistently improves planning robustness.

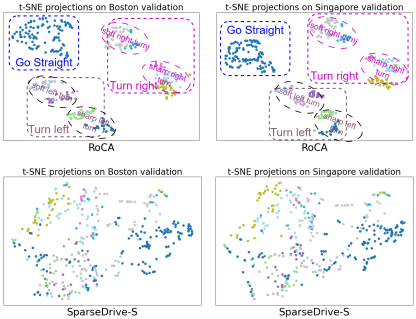
Why RoCA works: better token structure + informative uncertainty

Better token geometry

- RoCA yields more separable clusters for ego and agent trajectory modes.
- Driving commands such as **go straight**, **turn left**, and **turn right** become more structured in token space.
- This improves stability under perturbations and domain shift.

Better active learning

- GP predictive variance selects **more informative target samples**.
- RoCA-based sampling outperforms random selection and stronger uncertainty baselines like MC Dropout and Deep Ensembles.



Conclusion

- We propose **RoCA**, a GP-based framework for robust cross-domain end-to-end driving.
- RoCA learns a probabilistic codebook over ego and agent behaviors, enabling:
  - robust source-domain training,
  - uncertainty-aware target adaptation,
  - effective active learning,
  - stronger performance on long-tail and degraded-image scenarios.
- A key advantage is that RoCA is **plug-and-play**: when used for training regularization, it adds **no extra deployment latency**.

Thank you!

Questions and feedback are welcome.