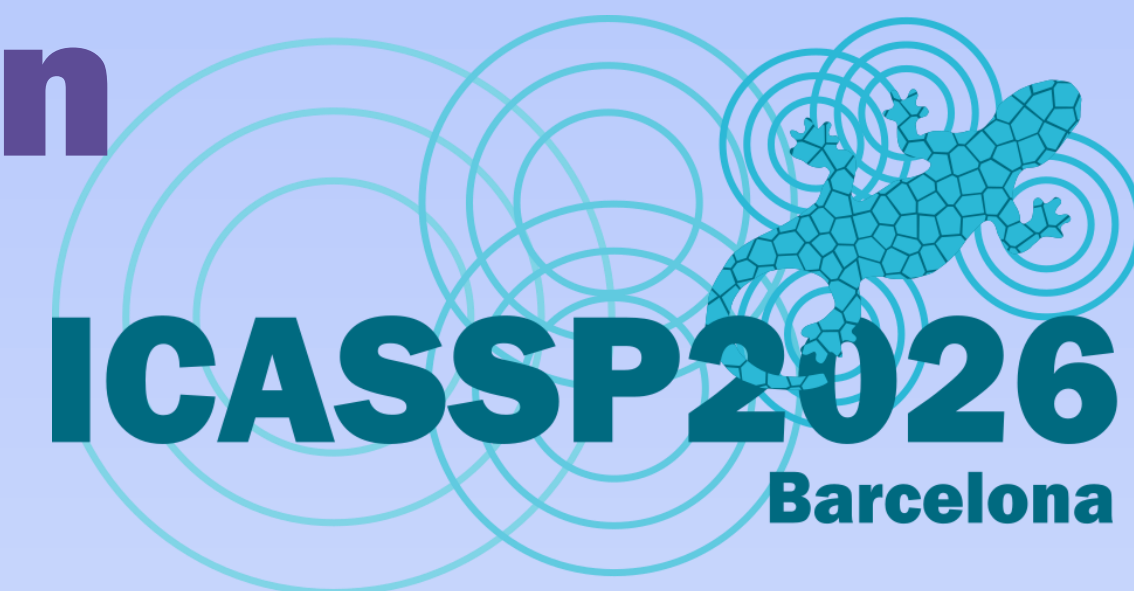


Performance-guided Reinforced Active Learning for Object Detection



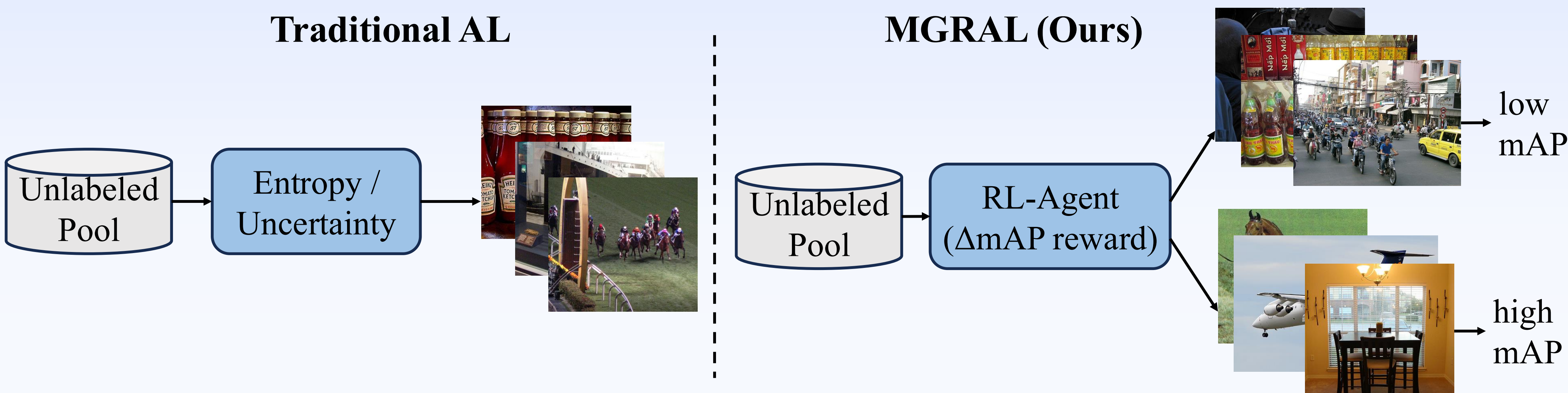
Zhixuan Liang^{1,2} Xingyu Zeng² Rui Zhao² Ping Luo¹

¹The University of Hong Kong ²SenseTime Research



Motivation

- Existing AL methods select data by **distribution** or **uncertainty** — not directly by detection performance (mAP)
- Batch selection is **discrete & non-differentiable** → gradient-based optimization fails
- We propose directly using ΔmAP as reward to guide sample selection via RL



Our Approach: MGRAL

(a) Training Phase

- At each AL cycle, the RL-based sampling agent queries a candidate batch from the unlabeled pool and feeds it into a semi-supervised detector to approximate ΔmAP .
- This ΔmAP serves as the **policy gradient reward**, driving the agent to select batches that genuinely improve detection.
- A **Look Up Table pre-trains multiple detector variants in parallel**, replacing thousands of sequential re-trainings with Wasserstein distance lookup (**1000× speedup**).

(b) Inference Phase

- The trained agent scores every unlabeled batch in a single forward pass.

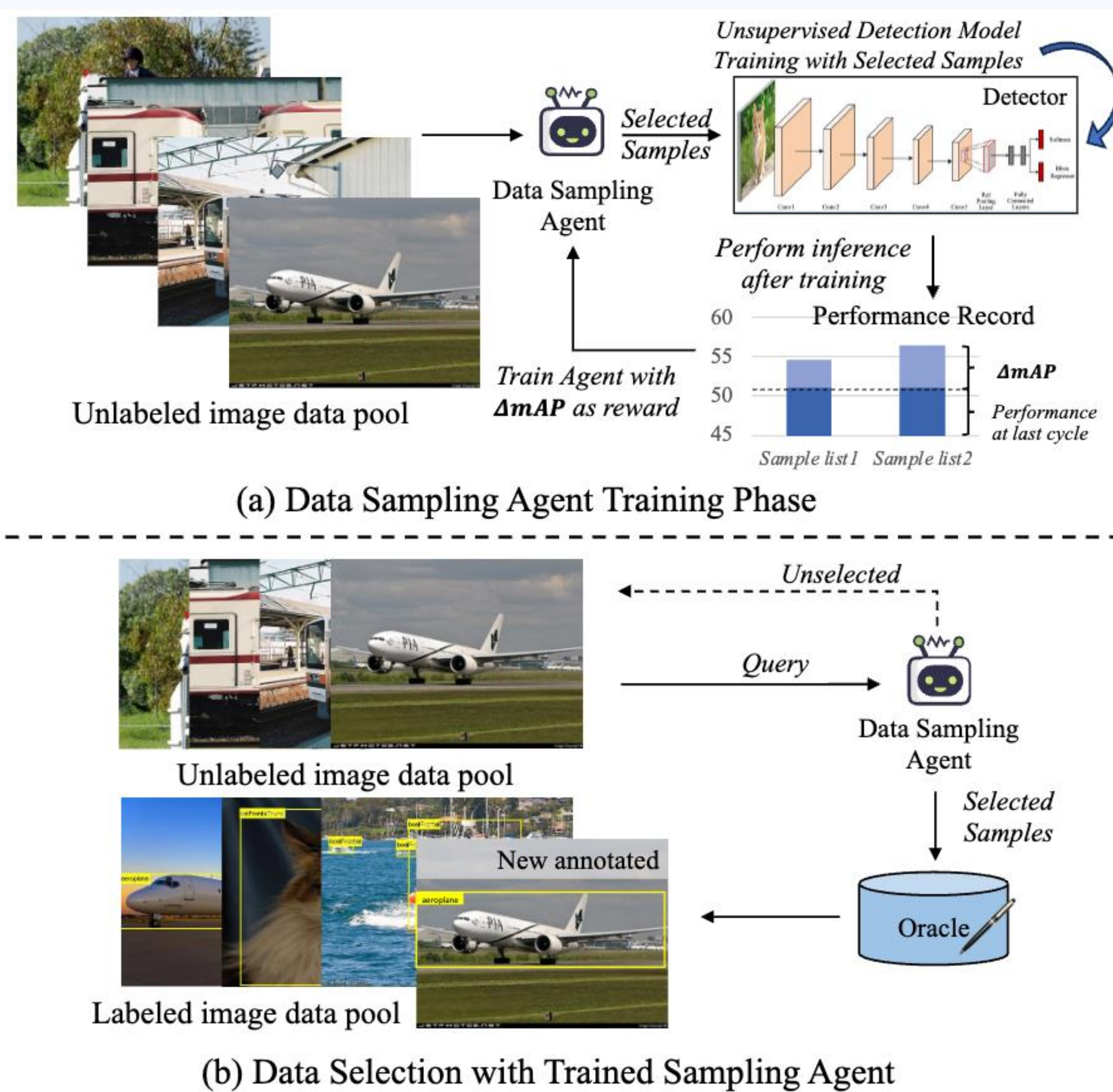


Fig. 1: Overview of Performance-guided (mAP-guided) Reinforced Active Learning (MGRAL) for object detection.

$$\text{Reward: } \Delta mAP_i^t = mAP_i^t - mAP_{i-1}^t$$

$$\text{Loss: } \mathcal{L} = -(mAP_i^t - mAP_{\text{ref},i-1}^t)$$

Our Approach: MGRAL

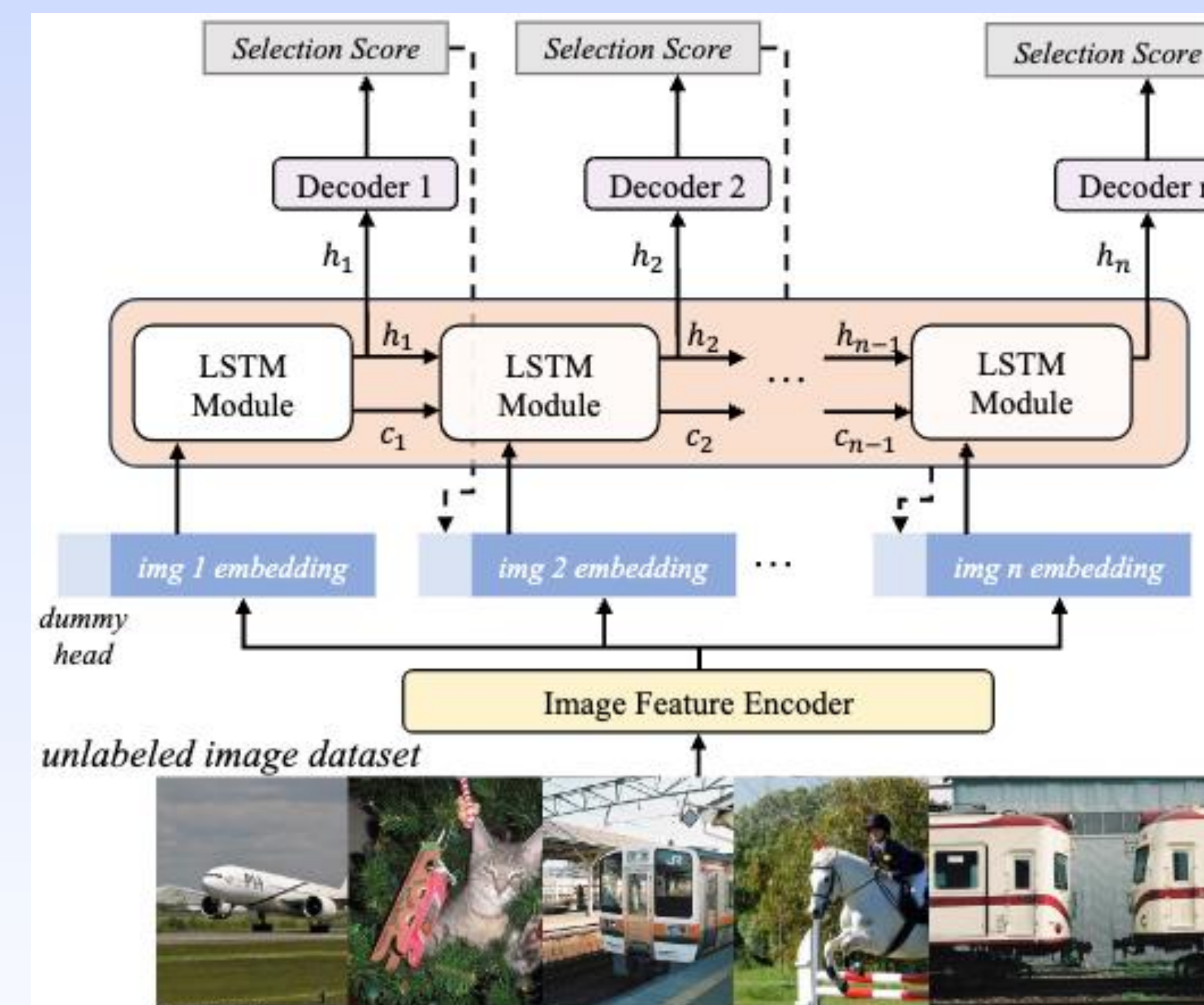


Fig. 2: Data sampling agent architecture.

Links & Contact



Personal Website



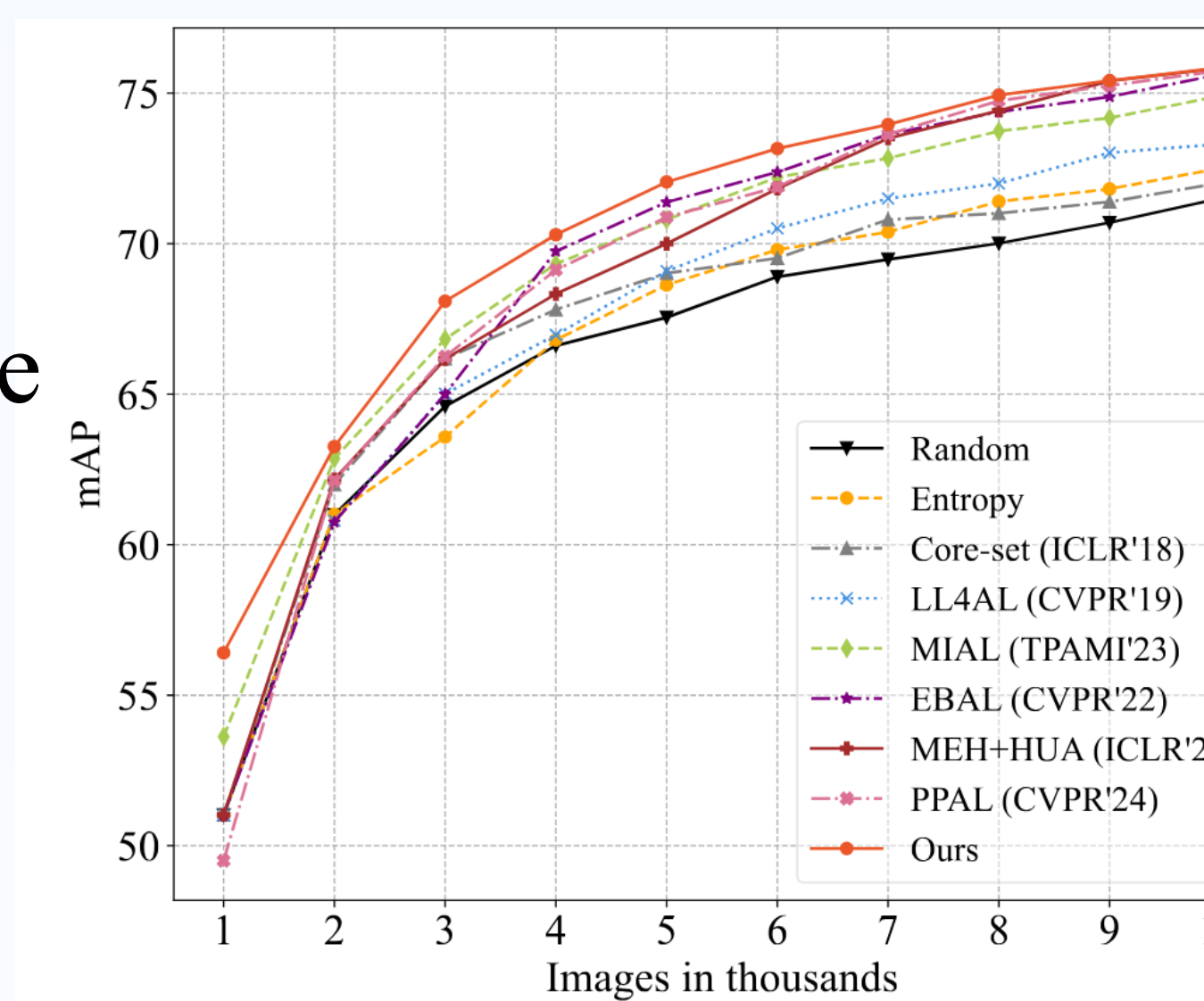
WeChat

LSTM-based agent processes each image embedding sequentially, outputting a **selection score** per sample; top-B chosen as the query batch.

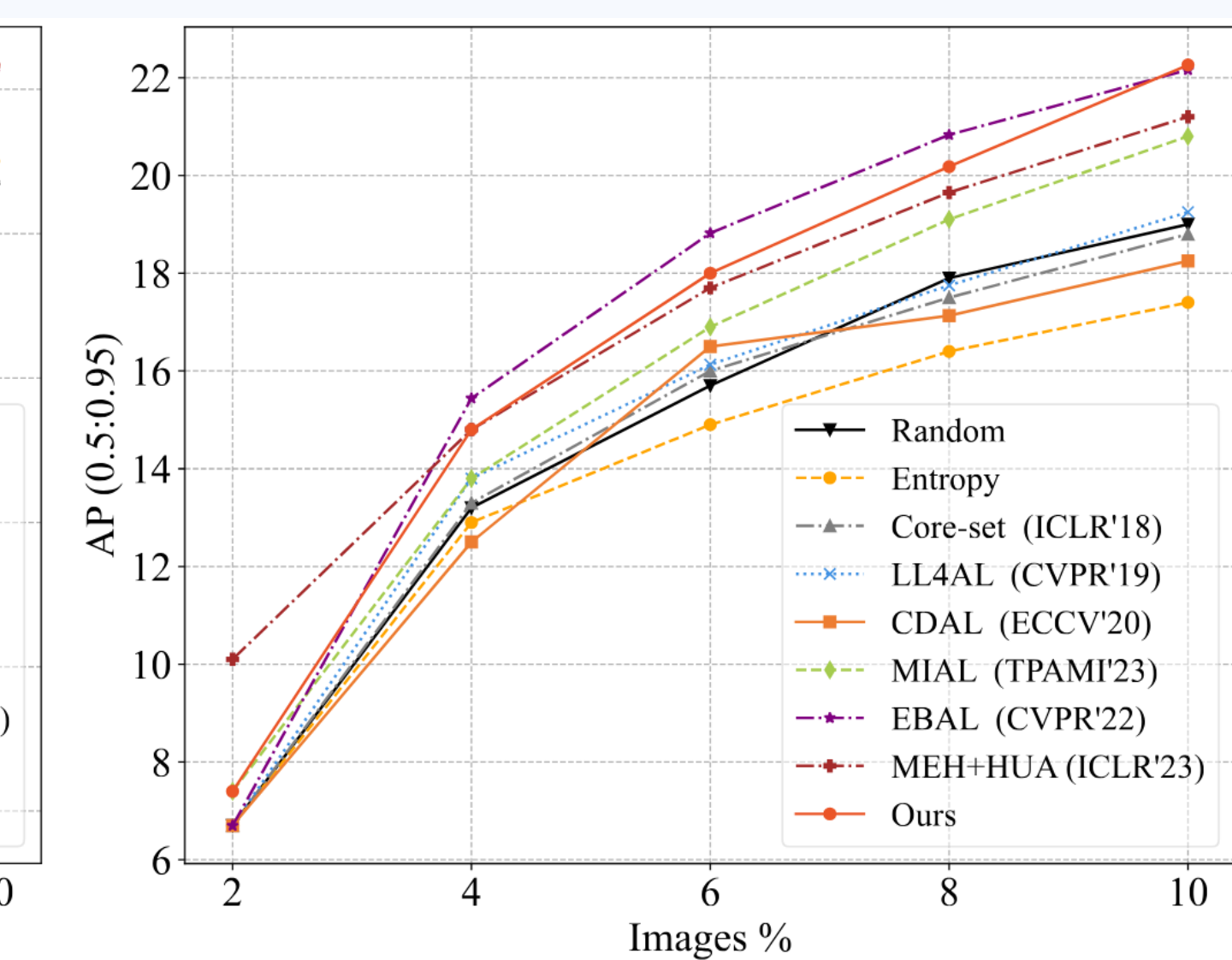
Experiments

➤ Main Results

- MGRAL achieves the highest AL curve on both VOC and COCO, outperforming all baselines including PPAL & MEH+HUA.



(a) SSD on Pascal VOC



(b) RetinaNet on MS COCO

➤ MGRAL promotes category diversity while maintaining clean supervision signal.



➤ Fastest inference; LUT overhead amortized across cycles. (See Tab. 1 & 2 in paper)

- Training Time of One Cycle: (9 h) + 7 h 8 min; Inference Time of One Cycle: 0.5 min