Making Your First Choice: To Address Cold Start Problem in Vision Active Learning

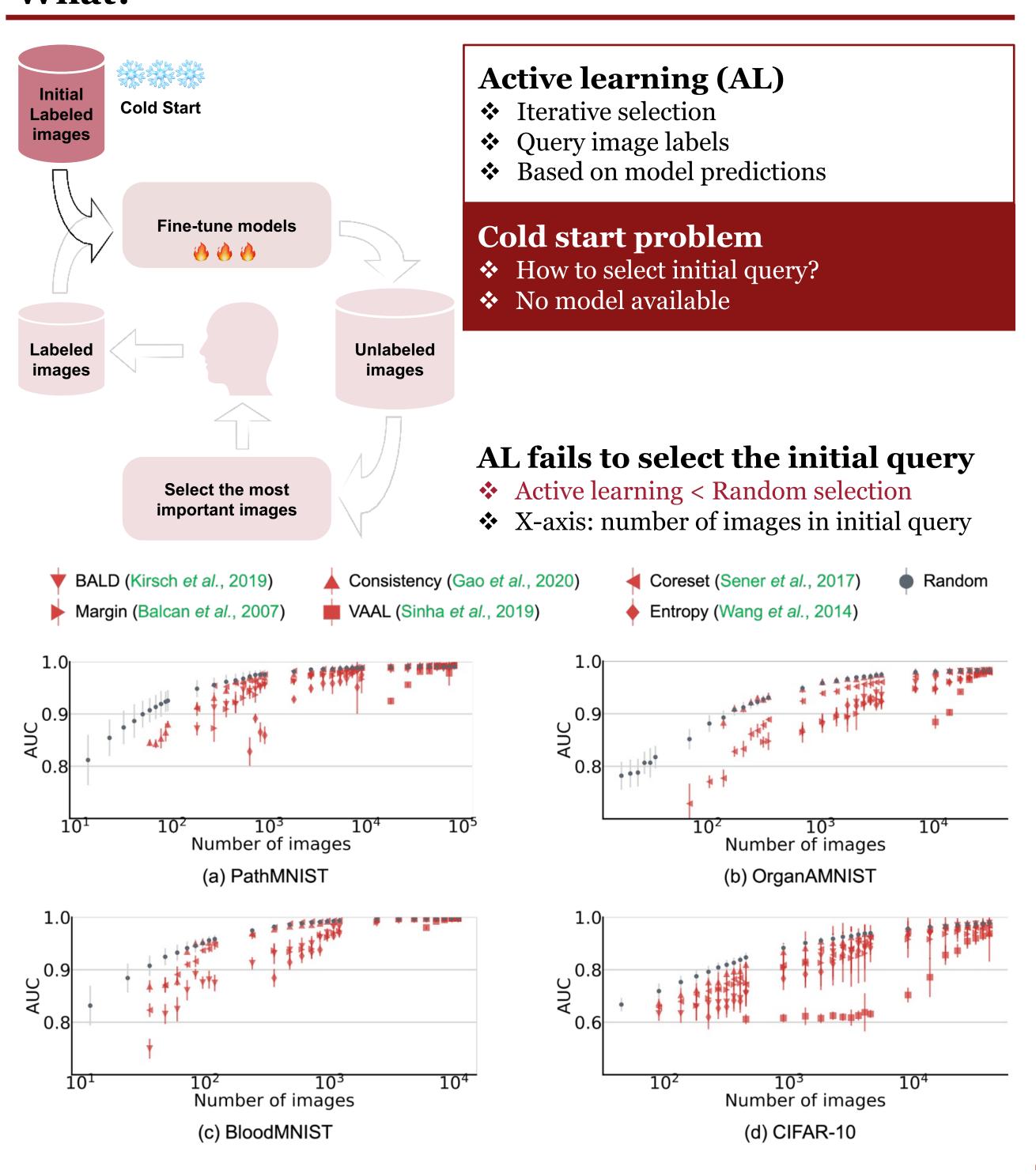


Liangyu Chen¹, Yutong Bai², Siyu Huang³, Yongyi Lu², Bihan Wen¹, Alan L. Yuille², Zongwei Zhou² ¹Nanyang Technological University, ²Johns Hopkins University, ³Harvard University

Key takeaways

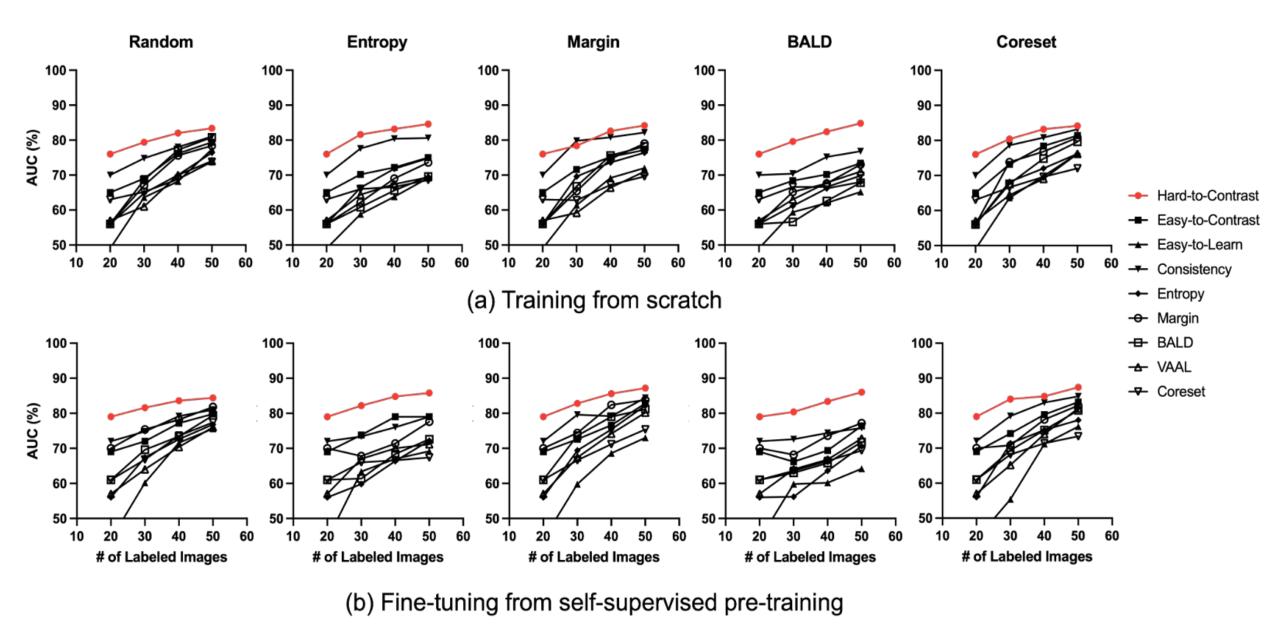
- ❖ We identify the cold start problem in vision active learning.
- ❖ A good initial query foresees improved active learning results.
- * Cold start problem is caused by 1 biased query and 2 outlier query.
- ❖ We propose "hard-to-contrast" strategy to select a balanced and typical initial query, as a simple baseline to address cold start problem.

What?



Why it matters: better initial query → better AL performance

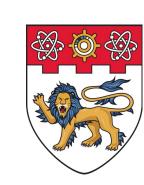
- ❖ AUC at initial query is strongly correlated with AUC at the last query
- Plot: active learning strategy
- Line: initial query
- ❖ The red line is our proposed "Hard-to-Contrast" strategy



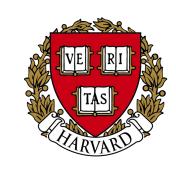
Paper



Code



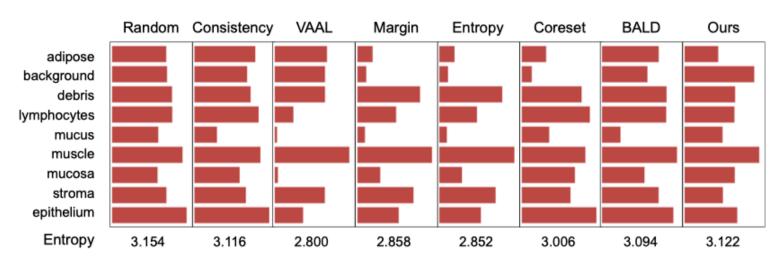




Why?

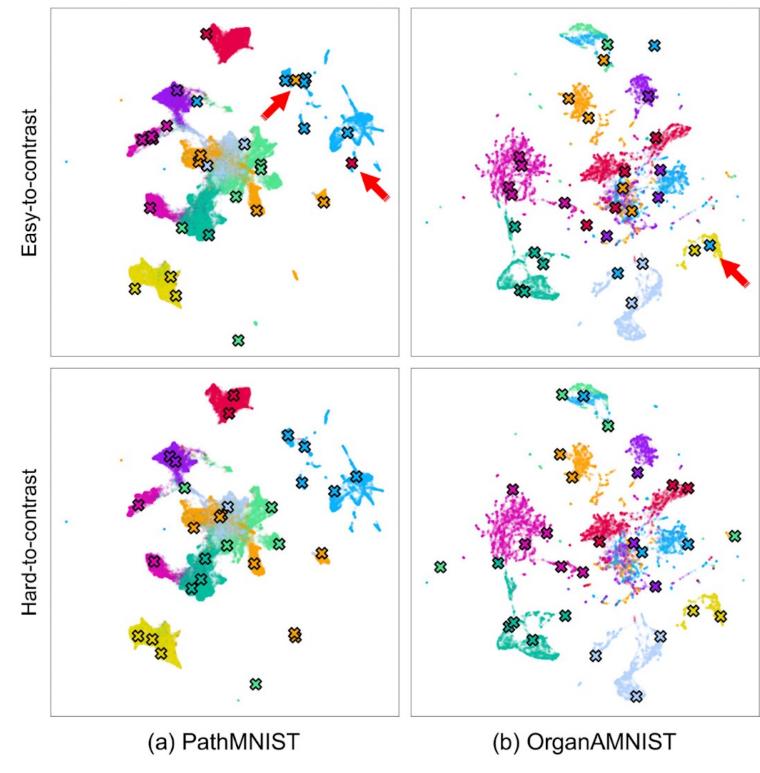
1) Biased query

- ❖ Initial queries are class-imbalanced
- Hypothesis: label diversity is significant



2 Outlier query

- ❖ Local outliers: out-of-distribution samples in a cluster
- Uncertainty-based AL selects local outliers
- * Hypothesis: outliers are harmful in cold start



How?

Hard-to-Contrast Initial Query Contrastive pre-training Data Map by pseudo-labels basophil lymphocyte platelet eosinophil monocyte erythroblast neutrophil 1.0 0.8 confidence 0.6 0.4 Easy-to-learn Easy-to-contrast 0.4 variability variability Hard-to-learn Data Map by ground truth Data Map by pseudo-labels Cluster image features

Query the most typical images among each cluster

- ❖ State-of-the-art with all AL strategies
- No manual annotation needed