

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

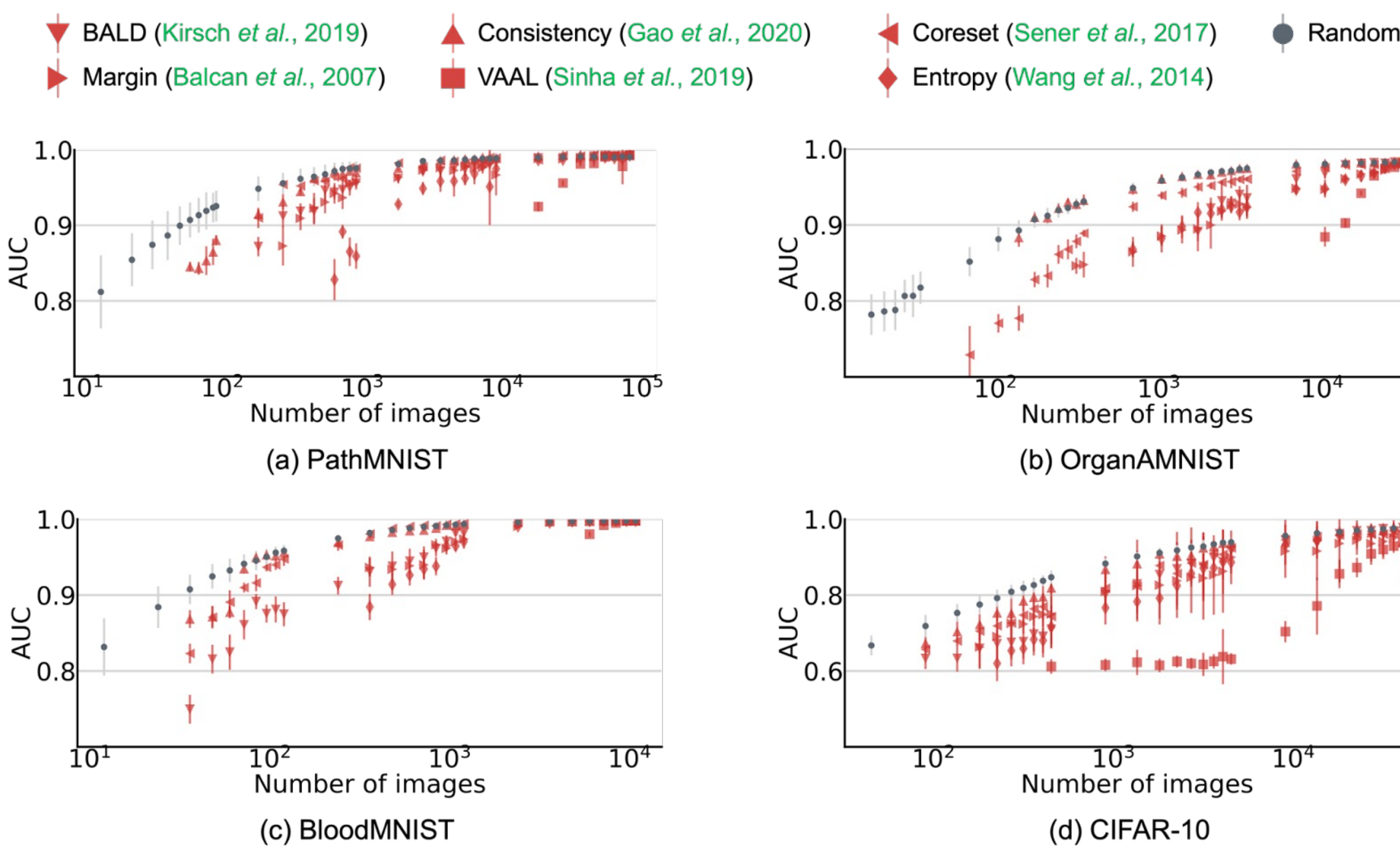
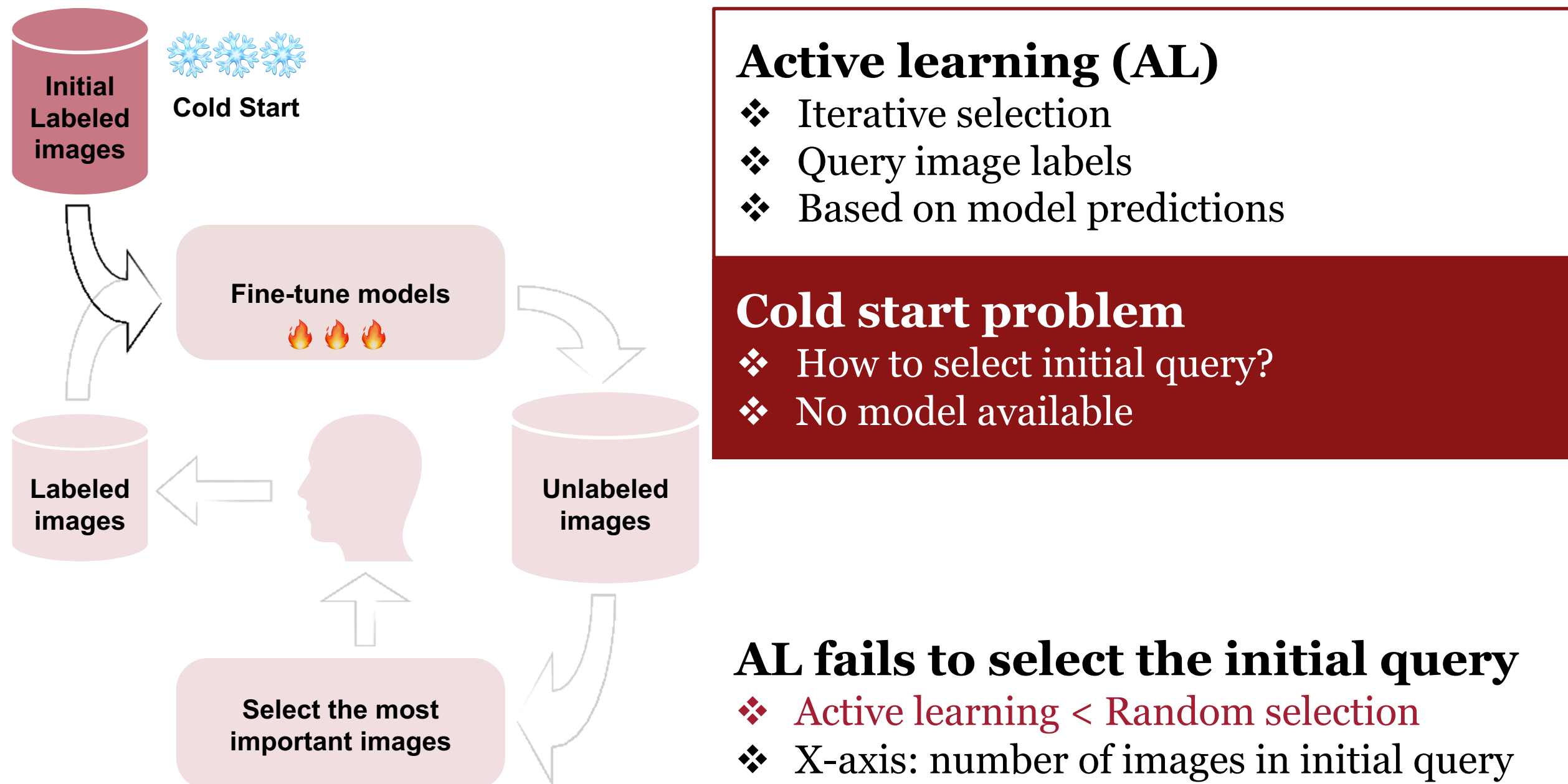
Liangyu Chen¹, Yutong Bai², Siyu Huang³, Yongyi Lu², Bihan Wen¹, Alan L. Yuille², Zongwei Zhou²

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.

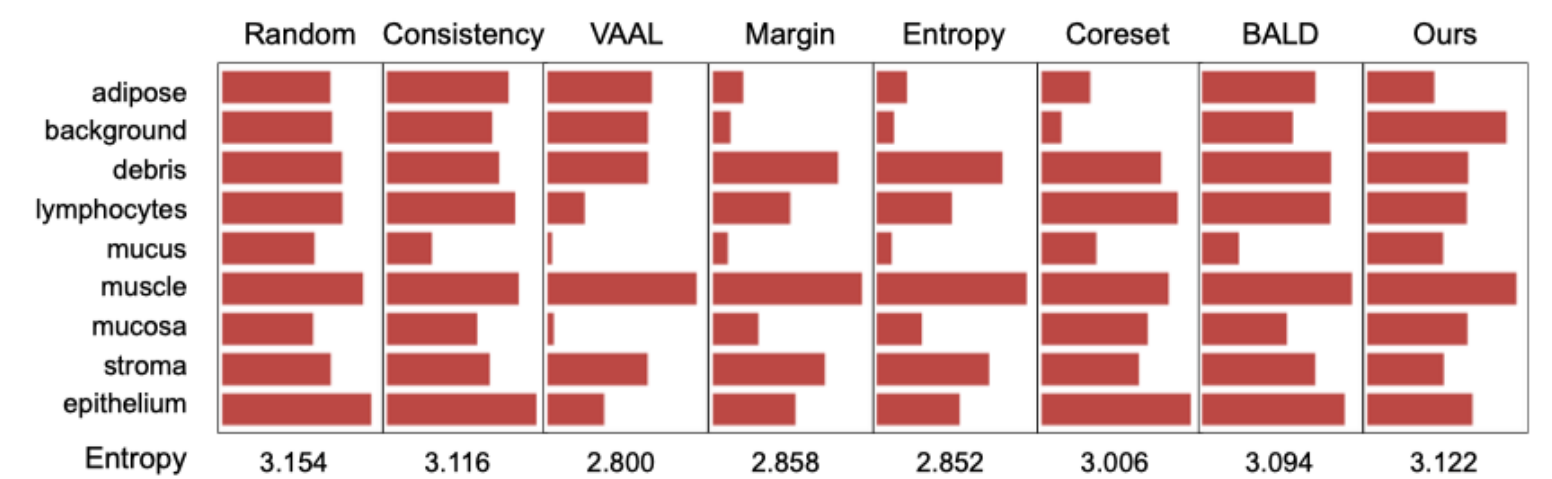
What?



Why?

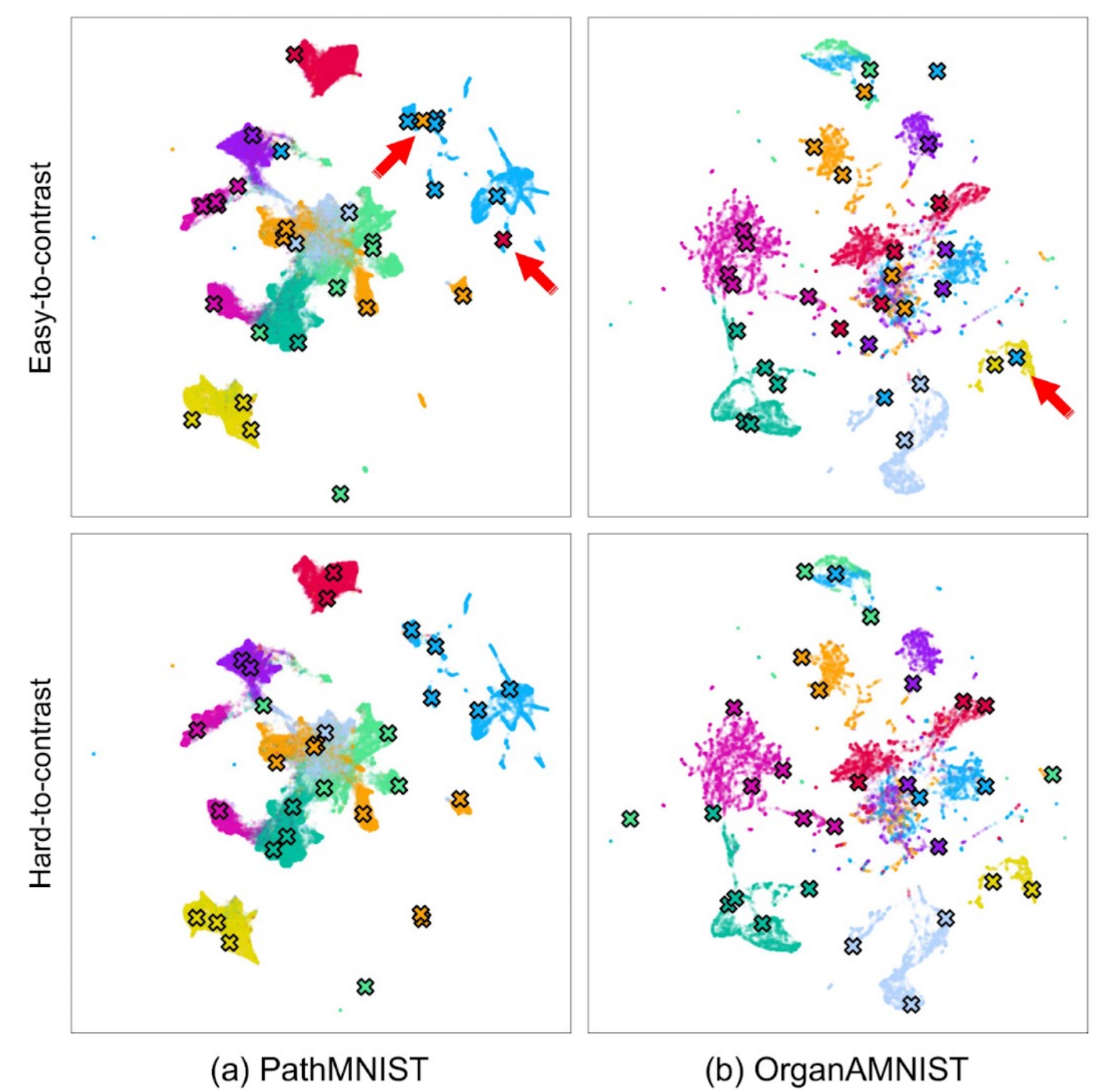
① Biased query

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



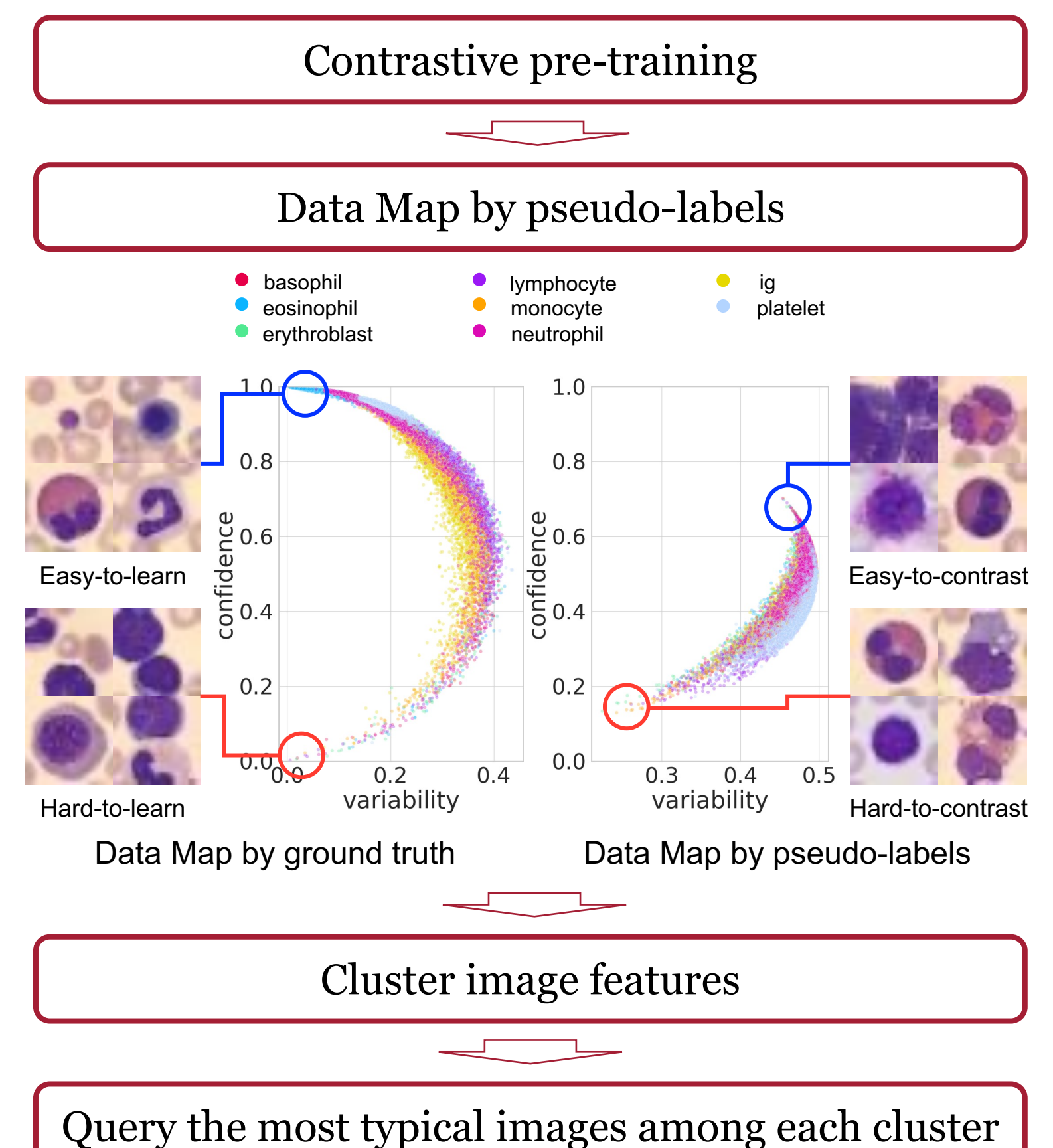
② Outlier query

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



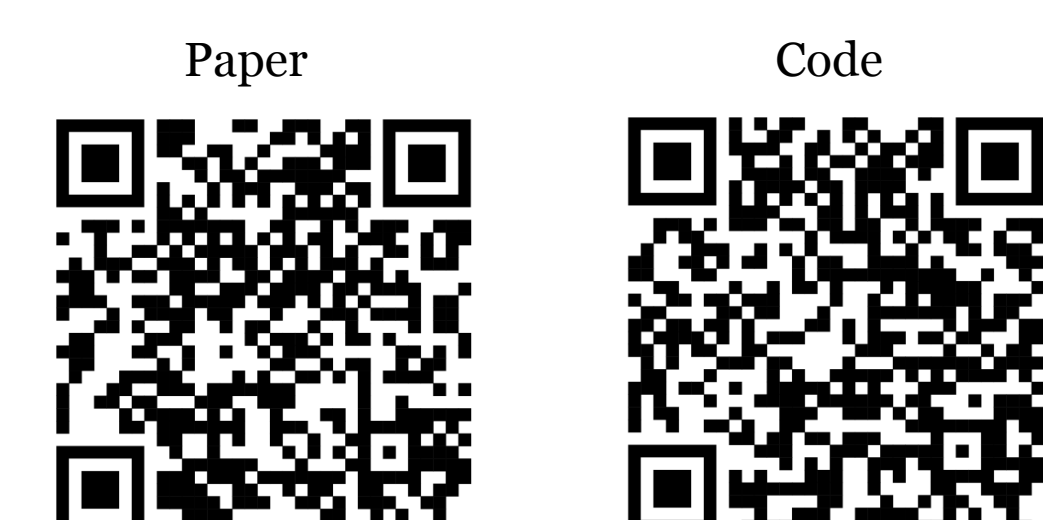
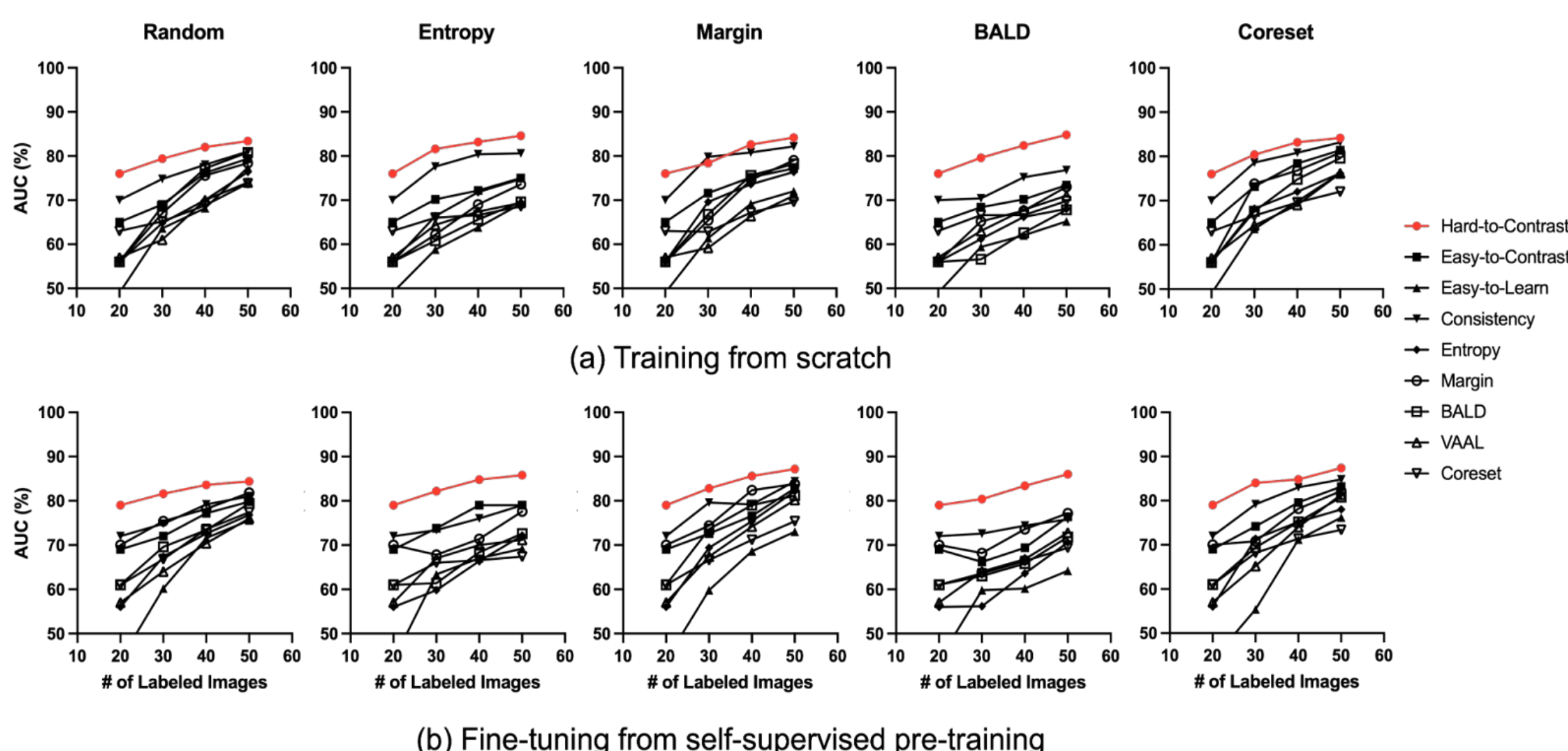
How?

Hard-to-Contrast Initial Query



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



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