

# 221201 勉強会 cold-start問題

B4 上野

# 目次

◆ 深層能動学習とcold-start問題

◆ cold-start問題の対策

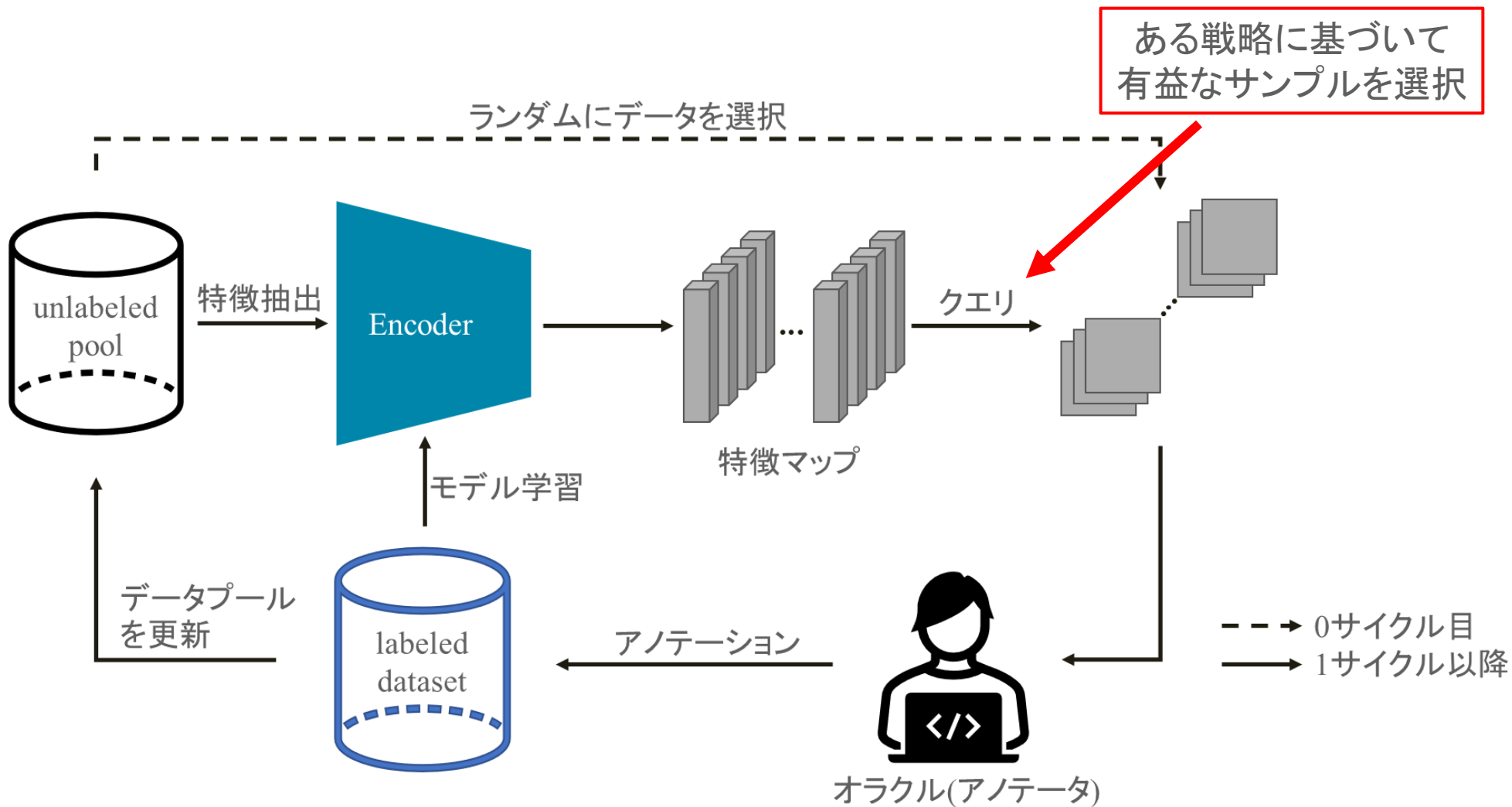
◆ まとめ

---

## 深層能動学習とcold-start問題

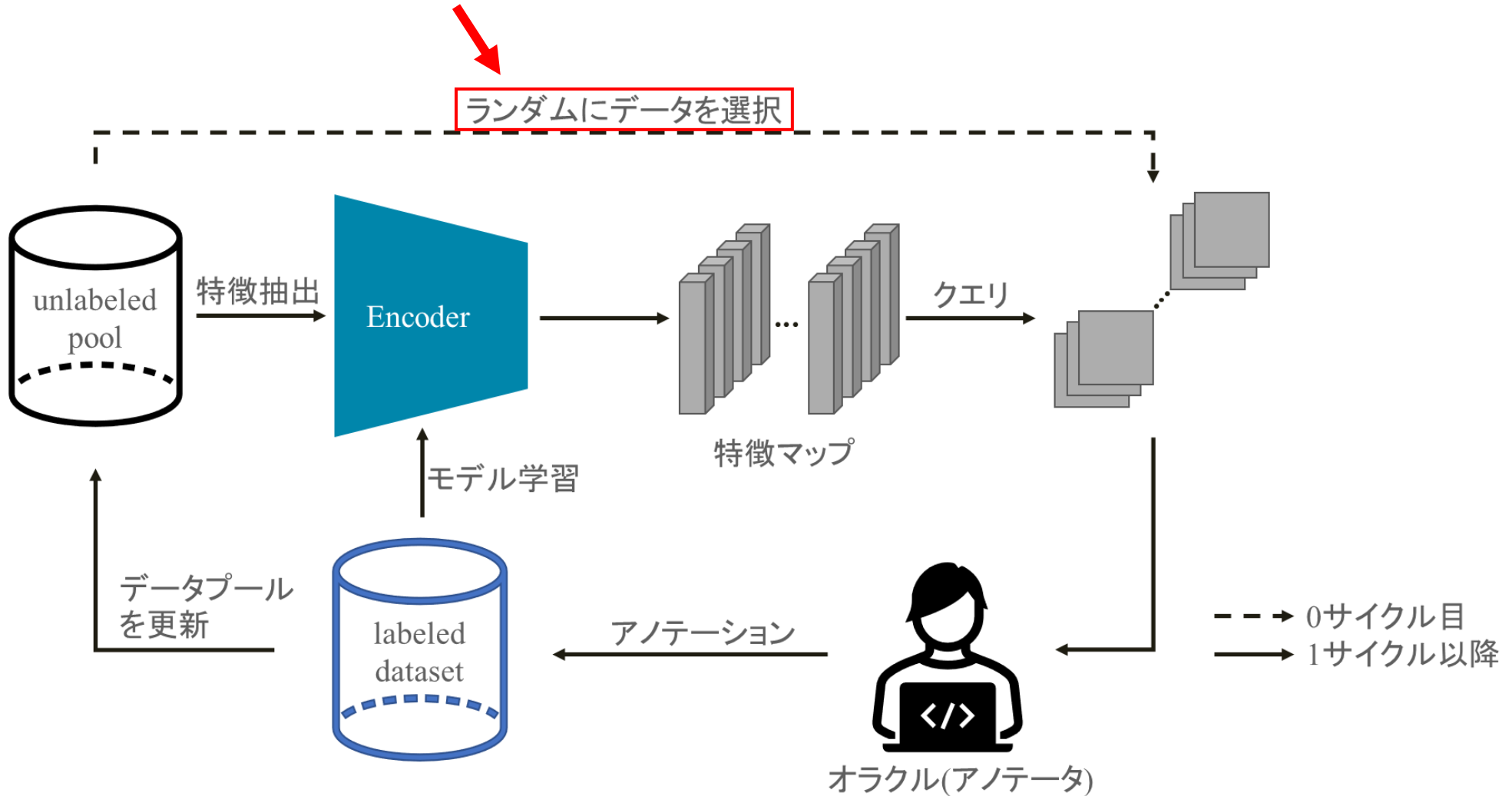
# 深層能動学習(DAL: Deep Active Learning)

◆DLの学習能力を保持したまま、ALによりアノテーションコストを削減



# cold-start問題とは

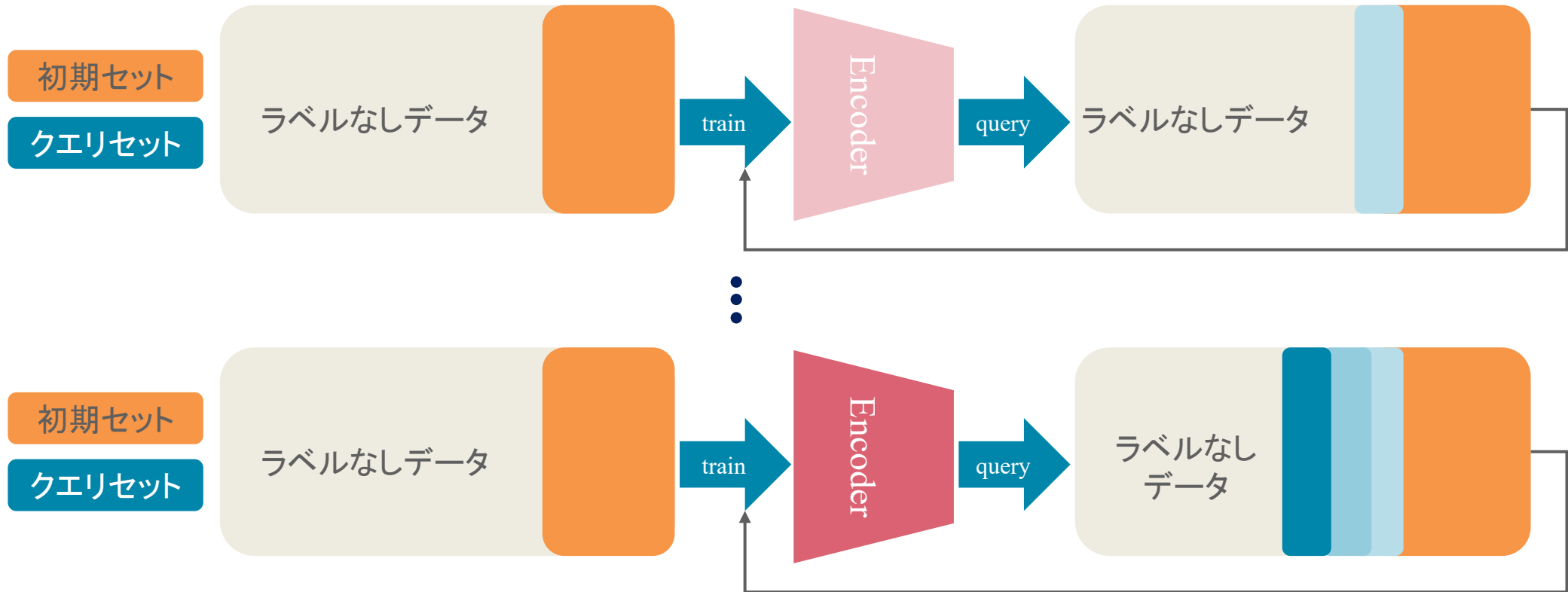
◆ 初期セットを何枚にすればよいか? という問題



## cold-start問題(1/5)

### ◆ 初期セットを何枚にすればよいか? という問題

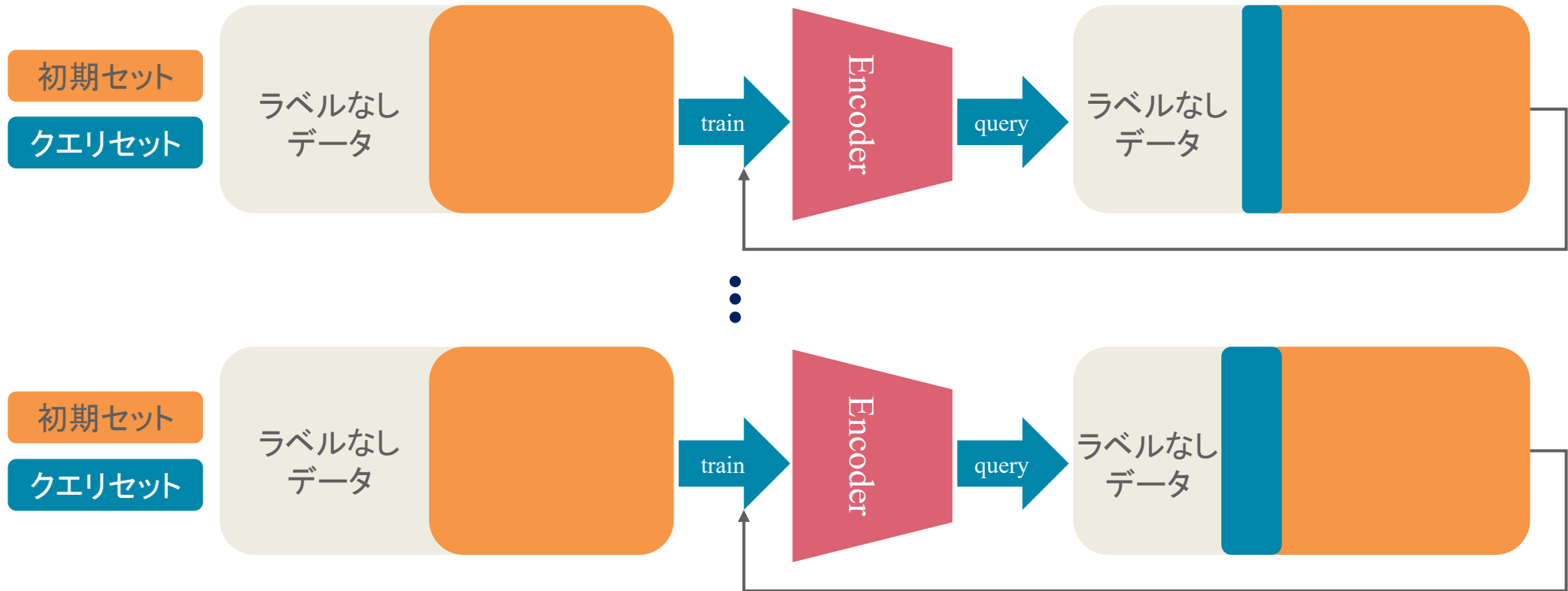
➤ 少ない場合: モデルが十分に学習できず, 以降のクエリ戦略が機能しない



## cold-start問題(2/5)

### ◆ 初期セットを何枚にすればよいか? という問題

➤ 多すぎる場合: 初期セットの分布が強く残り, クエリセットの価値が下がる



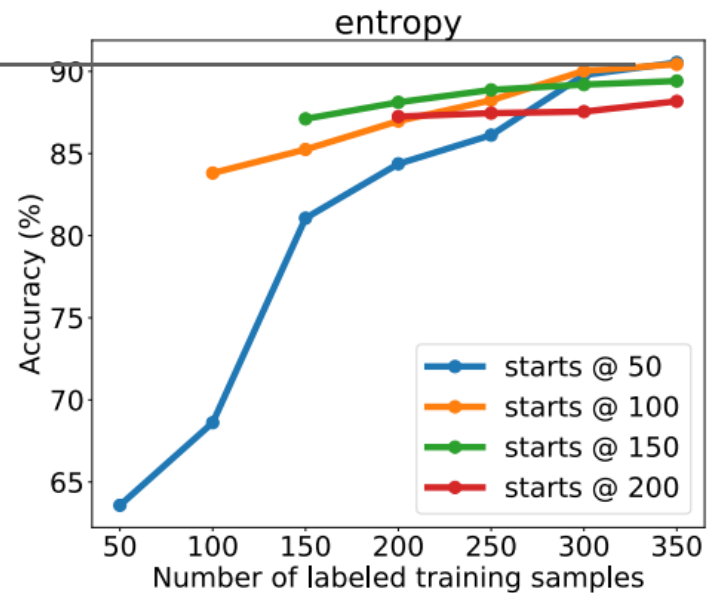
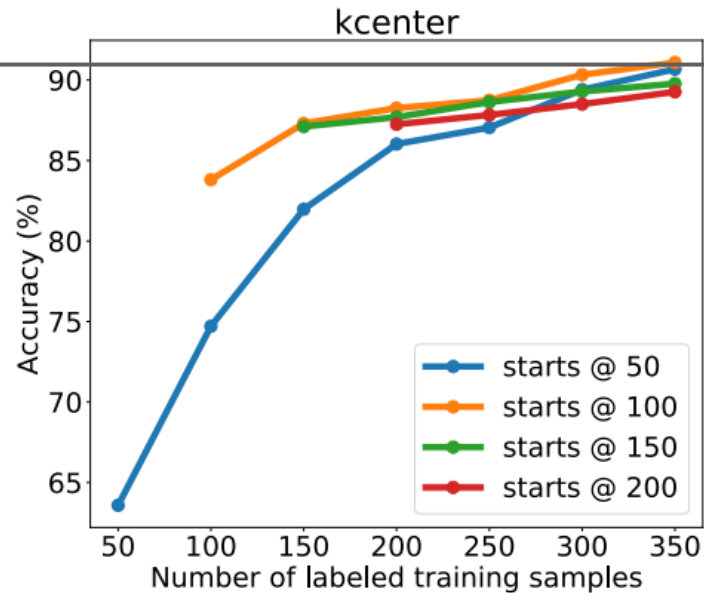
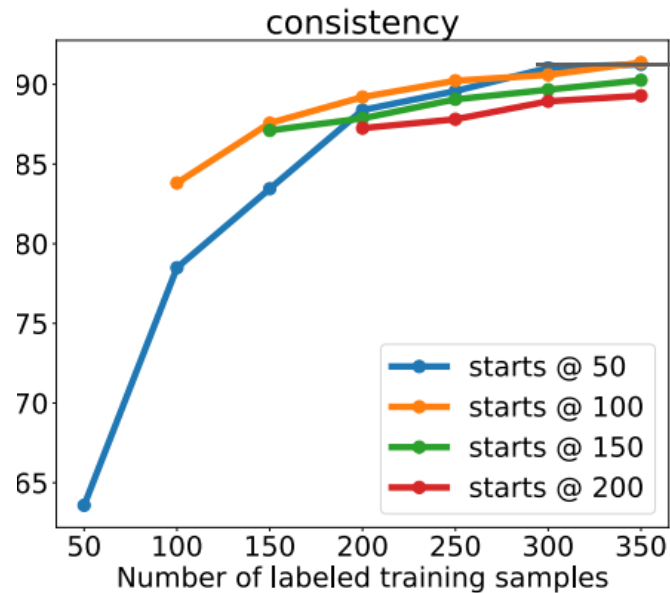
## cold-start問題(3/5)

◆ 初期セットの枚数次第で、後のサイクルでは同じ枚数でも精度が変わる

➤ 同じクエリ戦略でも、精度が大きく変化

◆ Q「300枚時点で一番優れた手法はどれ??」

➤ 初期セット=50枚ならconsistency





## cold-start問題(4/5)

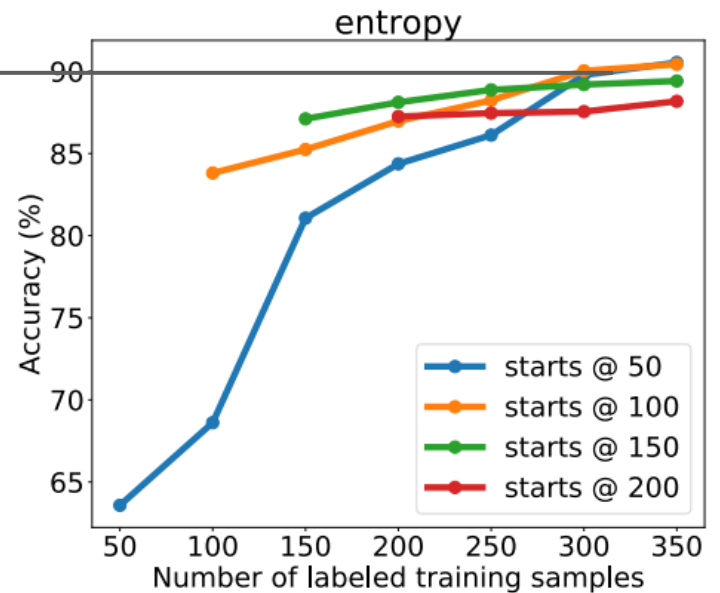
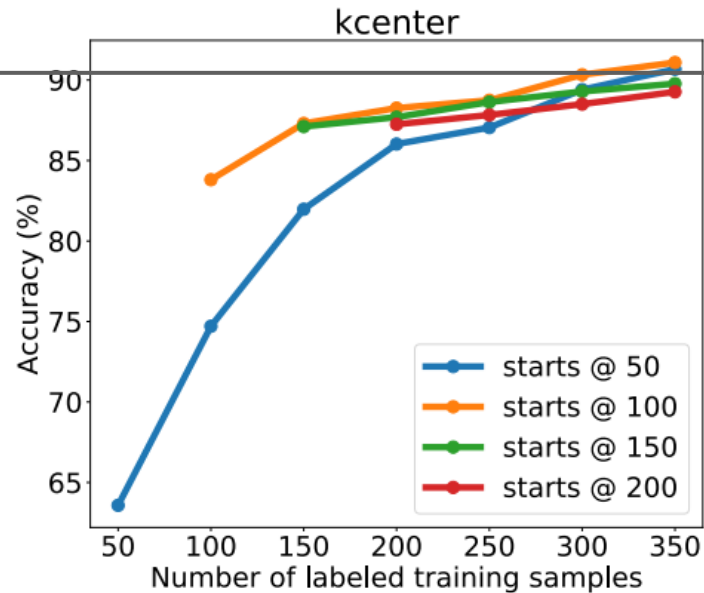
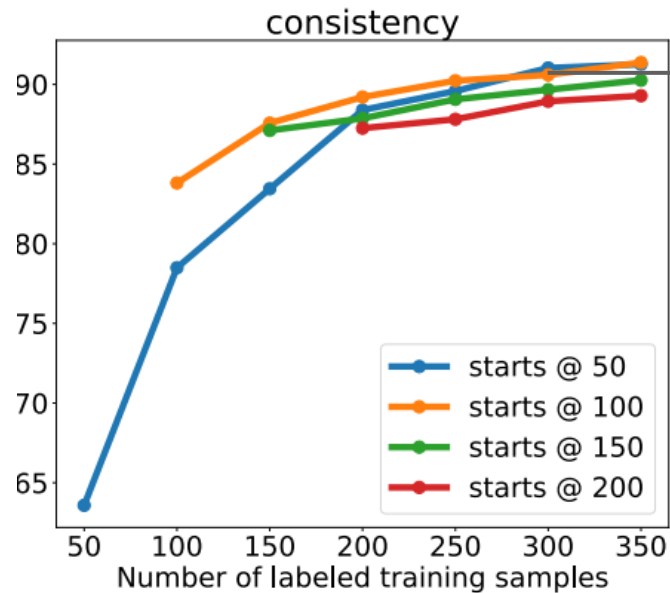
◆ 初期セットの枚数次第で、後のサイクルでは同じ枚数でも精度が変わる

➤ 同じクエリ戦略でも、精度が大きく変化

◆ Q「300枚時点で一番優れた手法はどれ??」

➤ 初期セット=50枚ならconsistency

➤ 初期セット=100枚ならentropy



## cold-start問題(5/5)

◆ 初期セットの枚数次第で、後のサイクルでは同じ枚数でも精度が変わる

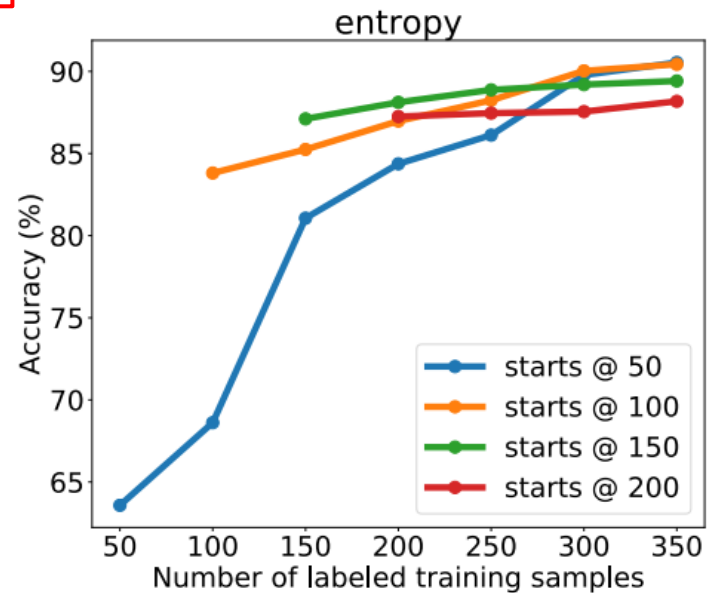
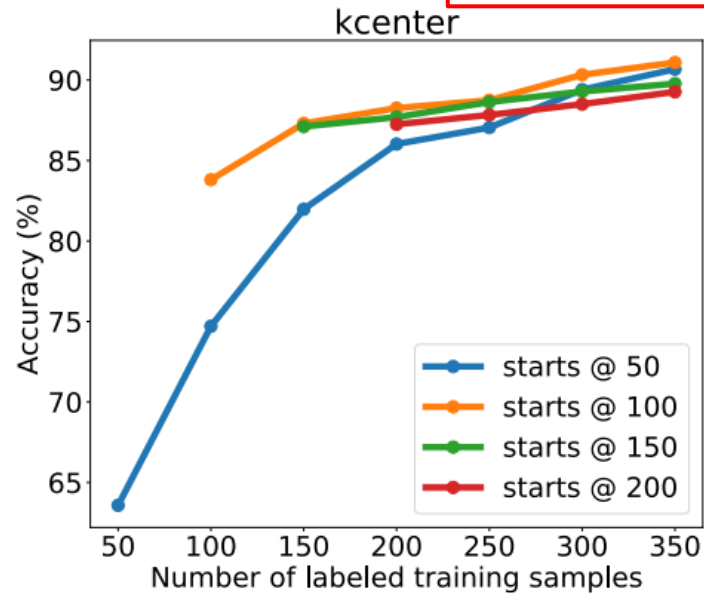
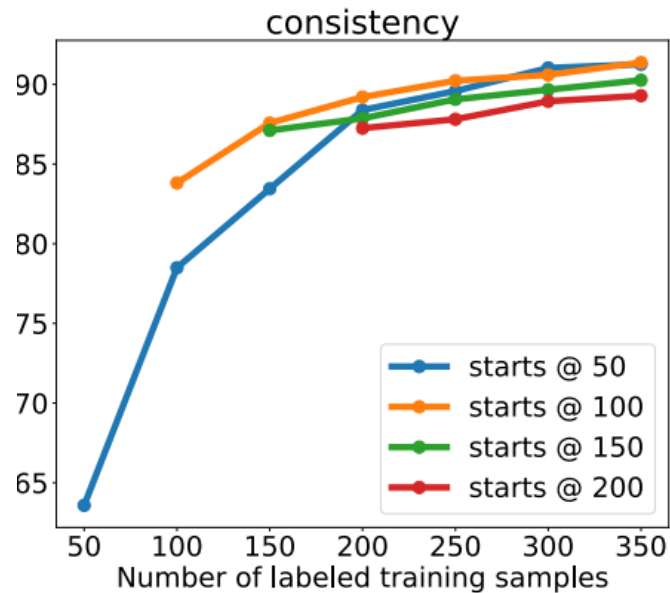
➤ 同じクエリ戦略でも、精度が大きく変化

◆ Q「300枚時点で一番優れた手法はどれ??」

➤ 初期セット=50枚ならconsistency

➤ 初期セット=100枚ならentropy

手法間の  
比較が困難



---

cold-start問題の対策

# cold-start問題の対策

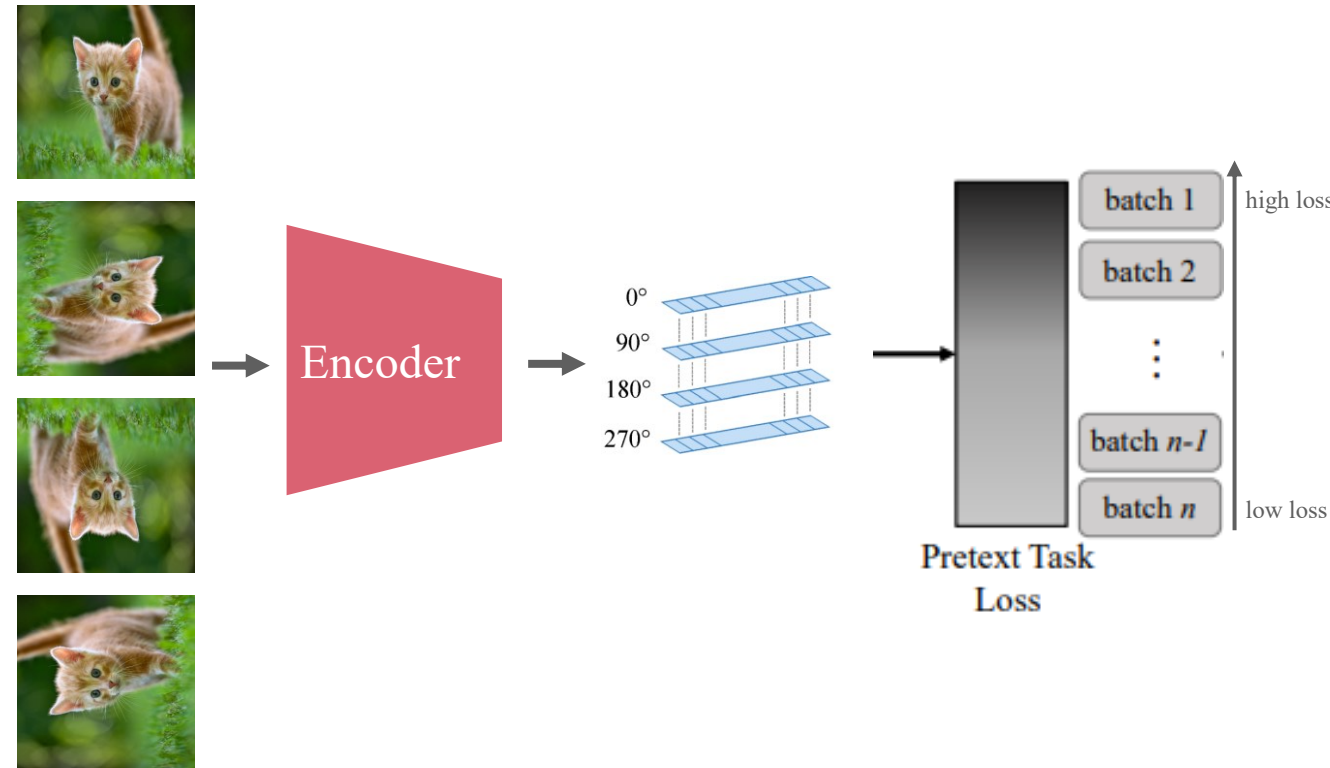
## ◆ 主な問題点は2つ

1. 初期セットをランダムサンプリング
  - 初期セットを選択可能なクエリ戦略
2. 最適な初期セット枚数が分からない
  - 最適な初期セット枚数を推定
    - 具体的な方法がない

# PT4AL: Using Self-Supervised Pretext Tasks for Active Learning (ECCV '22)

## ◆手順

1. モデルを回転予測タスクで事前学習
2. 各データの回転予測タスクのLossでデータを降順ソート
3. ソートされたデータをサイクル数分のバッチに分割
4. 初回クエリは最初のバッチから等間隔にサンプルを選択
  - Lossでソートされているためサンプルの難易度を考慮

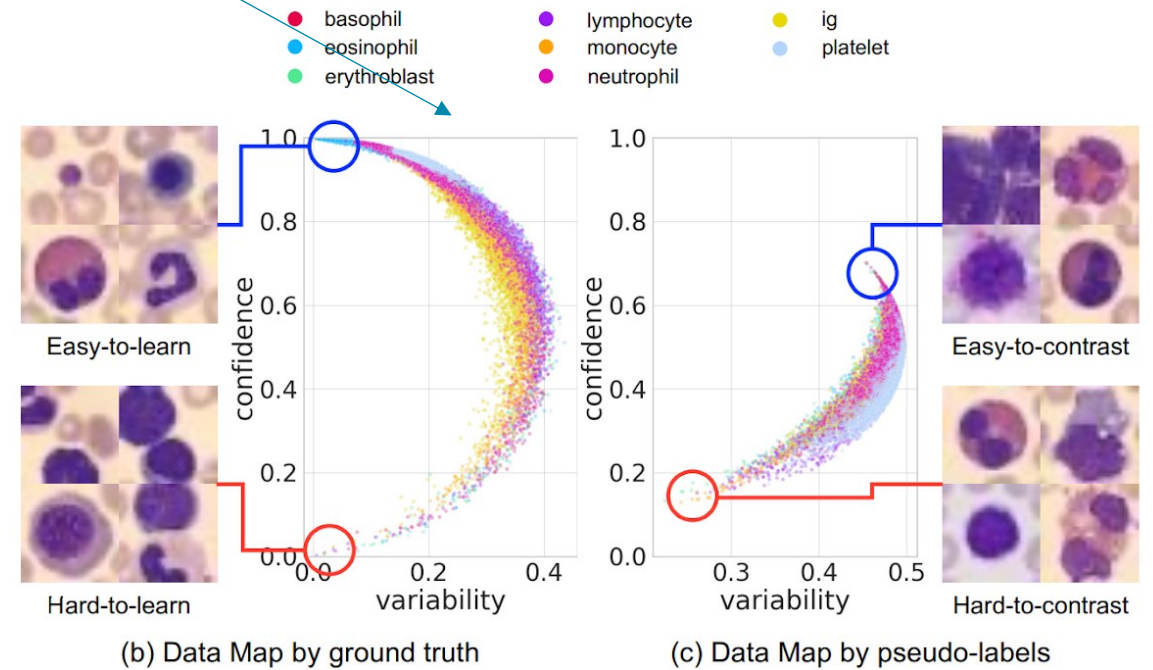
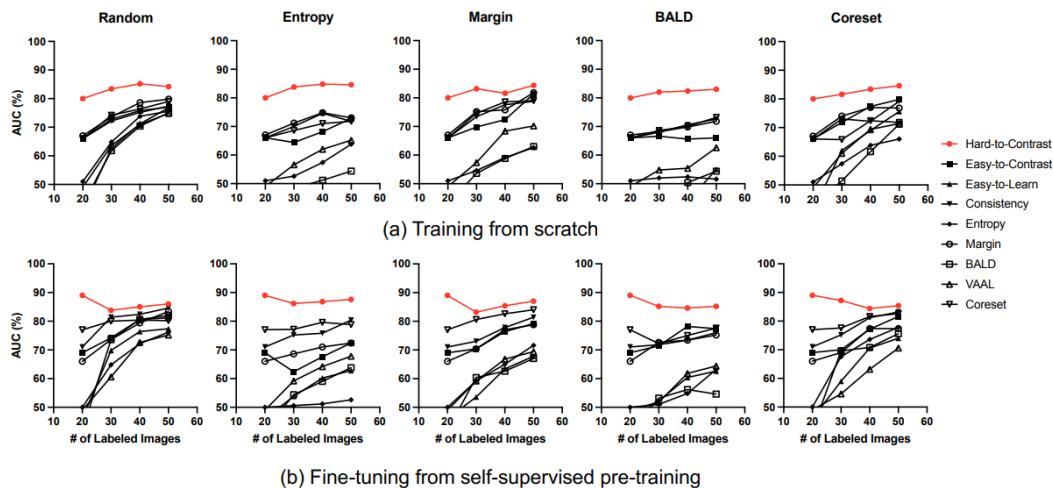


# Making Your First Choice: To Address Cold Start Problem in Vision Active Learning (arXiv '22)

## ◆手順

1. MoCo v2を用いて事前学習, k-meansでクラスタ生成
2. ground truthを疑似ラベルに置き換え, Dataset Mapを生成
3. マップ内から対象学習が難しいデータをクエリ
  - ▶ 対象学習が難しい = 他との区別がつかない = 典型的なデータ

## 結果



## まとめ

### ◆ cold-start問題の対策はまだ少ない

- 既存論文はすべて自己教師あり学習, 半教師あり学習を使用
  - スクラッチからの対策はない

### ◆ 枚数の推定は困難

- 教師なしデータを使う前提が高難易度

### ◆ 現状はどの論文も無視

---

補助資料



## SSL: Self-Supervised Learning(自己教師あり学習)

### ◆機械的に生成したラベルを用いて学習する教師なし学習の一種

- 人によるアノテーション不要のため, ALの問題設定にマッチ
  - ALでは事前学習として利用
- 教師あり学習と同等の精度が得られる

### ◆Pretext Tasks

- 回転予測, ジグソーパズルのように1枚の画像に前処理を与え, 前処理を基に疑似ラベルを作成し学習

### ◆対象学習

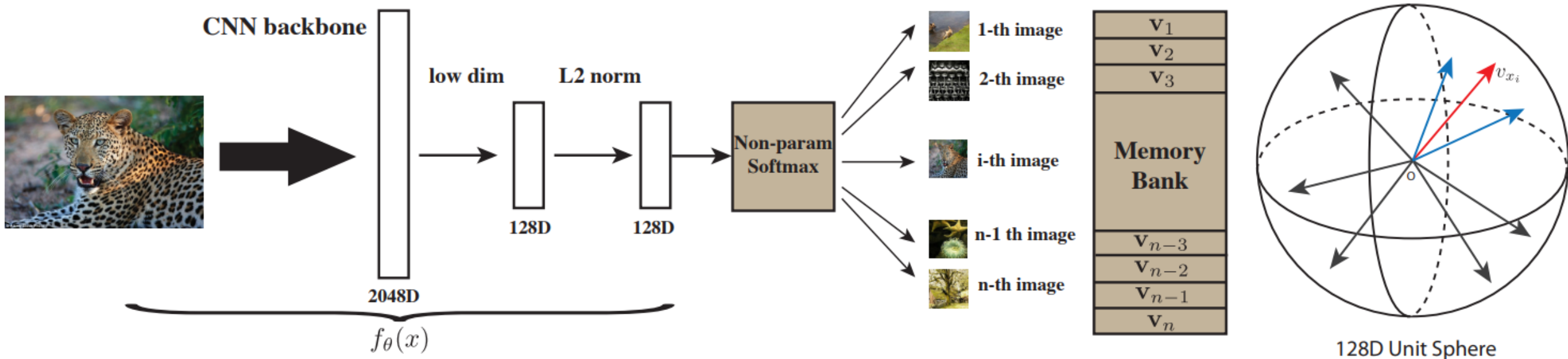
- 同じ画像から前処理されたデータのペアを正例, 別の画像からのペアを負例とし正例の類似度を近づけるように学習

# Instance Discrimination

## ◆ 同じ画像か否かを分類する学習

➤ タスク | 正例/負例の分類問題

$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{v}_i^T \mathbf{v} / \tau)}{\sum_{j=1}^n \exp(\mathbf{v}_j^T \mathbf{v} / \tau)}$$



## ◆ 対象学習のベース

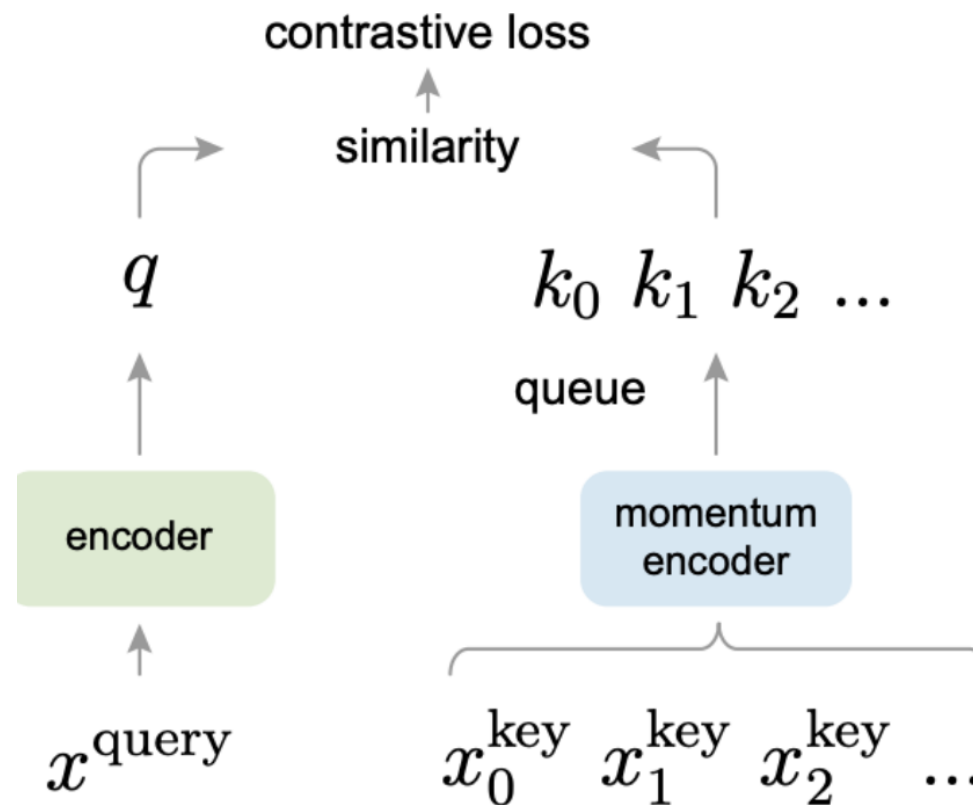
# MoCo(Momentum Contrastive Learning)

## ◆ Instance DiscriminationをKey-Queryタスクと解釈

- Key数が大きい状態を保つため queue形式を使用

## ◆ Momentum Encoderで学習を安定化

- Encoderのパラメータを0.1%混ぜてキーの一貫性を保ちつつモデルを最適化



# MoCo v2

## ◆ MoCoにSimCLRの構造を導入

## ◆ 対象学習を行うことでv1 と比べ精度向上

- ▶ SimCLRと比べ小さいバッチサイズ  
少ないエポック数で収束

case	MLP	unsup. pre-train			batch	ImageNet acc.
		aug+	cos	epochs		
MoCo v1 [6]				200	256	60.6
SimCLR [2]	✓	✓	✓	200	256	61.9
SimCLR [2]	✓	✓	✓	200	8192	66.6
<b>MoCo v2</b>	✓	✓	✓	200	256	<b>67.5</b>

results of **longer** unsupervised training follow:

SimCLR [2]	✓	✓	✓	1000	4096	69.3
<b>MoCo v2</b>	✓	✓	✓	800	256	<b>71.1</b>

