

クラス名に依存しない 視覚的記述子探索によるCLIPのOOP適応

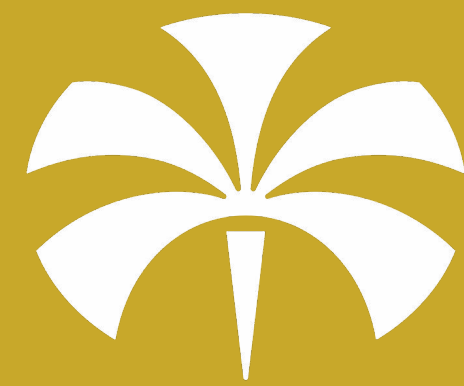
IS2-160

川越 壮

内海 ゆづ子

岩村 雅一

大阪公立大学 大学院情報学研究科

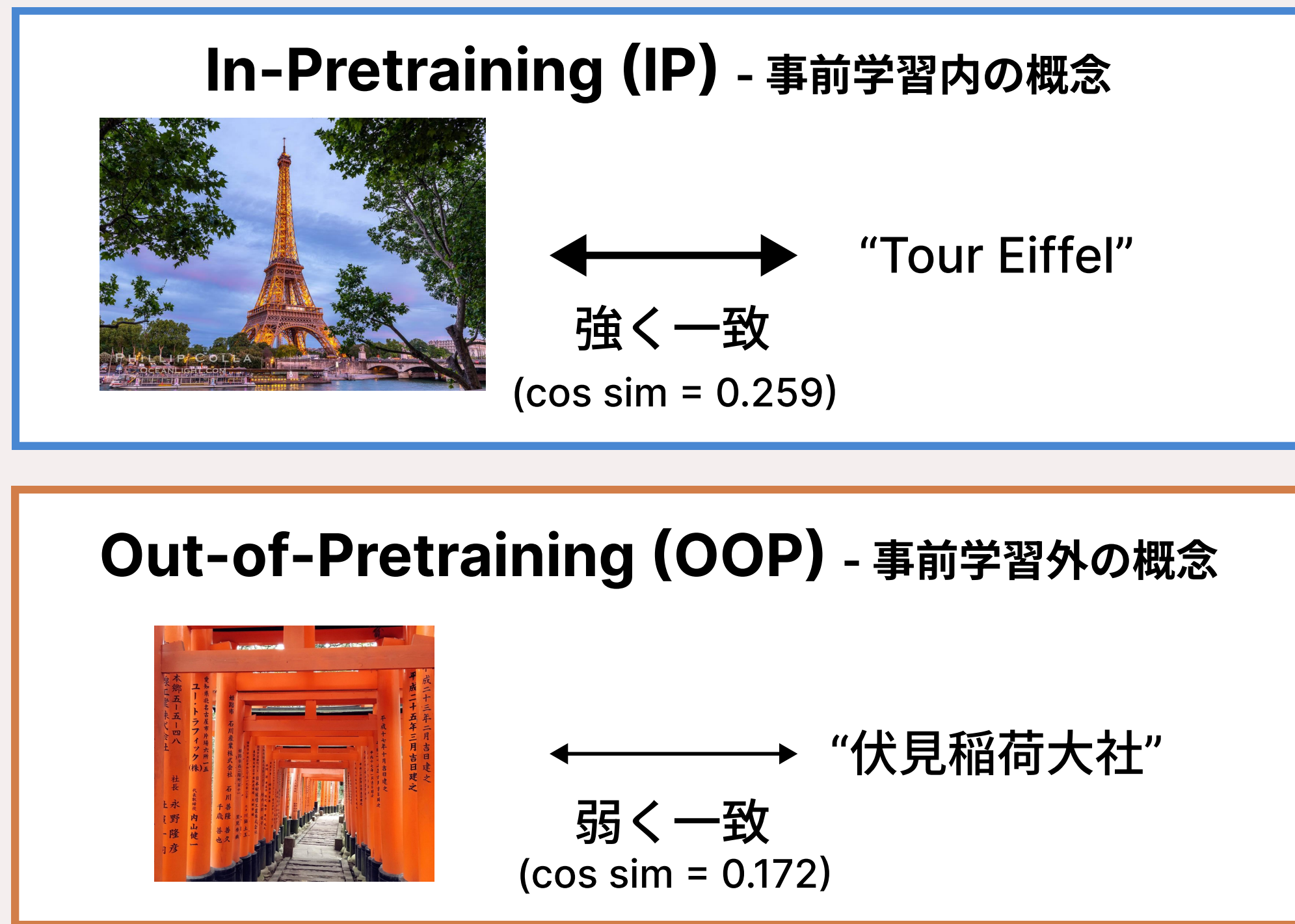


Osaka
Metropolitan
University

1. 背景と問題設定

CLIPのOOP問題

- IPではクラス名と画像が対応
- OOPではクラス名と画像が対応せず、Zero-shot精度が低下
- 少数のラベル付き画像からOOPクラス表現を適応する必要がある



着眼点

OOPクラスも
既知の視覚概念に分解できる

固有クラス名は未知でも、色・形・模様などの構成要素はCLIPが既に知っている



研究課題：少数ショット画像から、クラス名に依存しないOOPクラス表現を構築できるか？

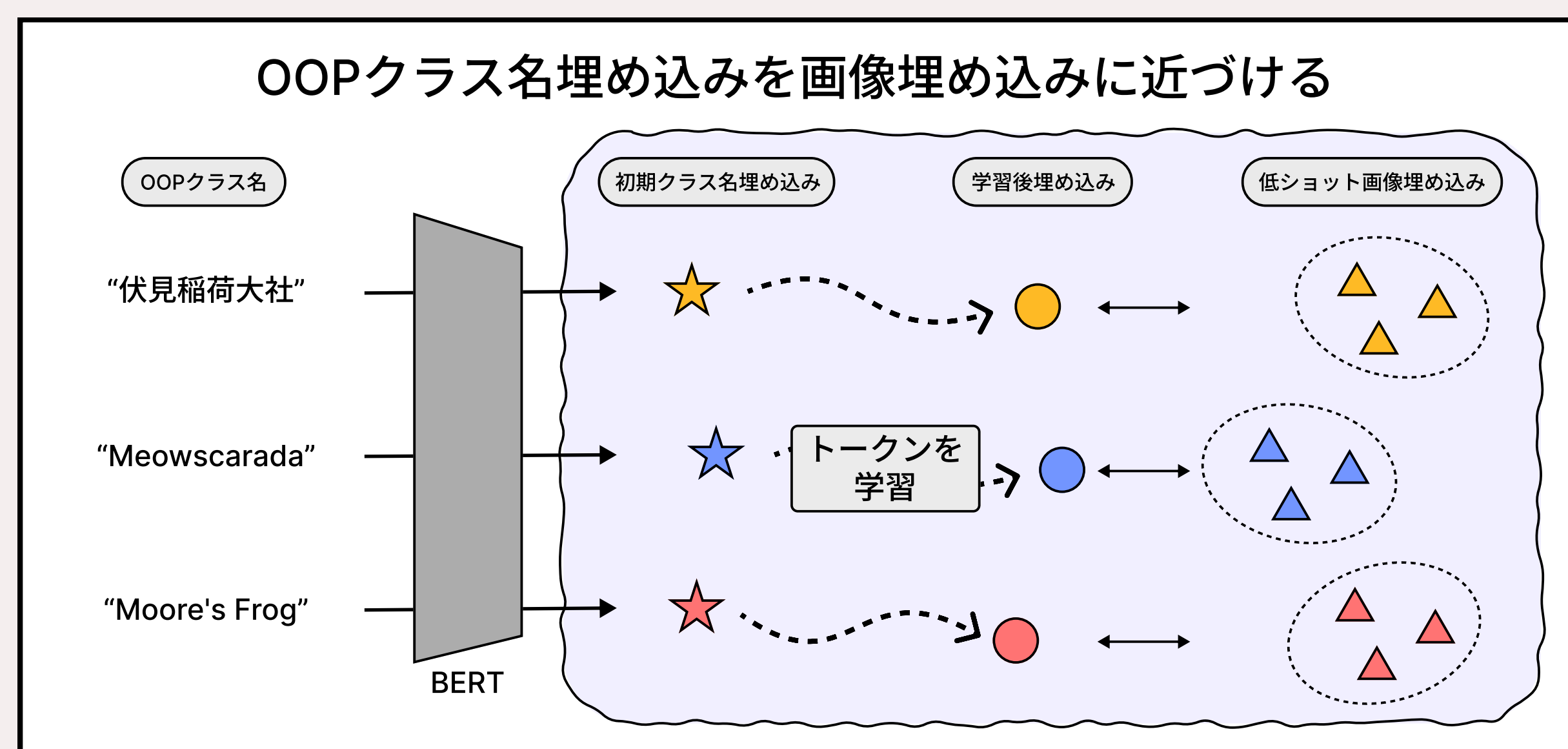
2. 既存手法と課題

手法：Few-shot Name Learning^[1] (FSNL)

- クラス埋め込みを画像の埋め込みに対応付け
- 少数ショットを手がかりに、クラス名に相当するトークンを学習

課題

- 勾配学習を伴う最適化によりOOPクラスの埋め込みを更新するが、計算コストが高い
- 構築されたクラス表現はベクトルのため、解釈性が低い

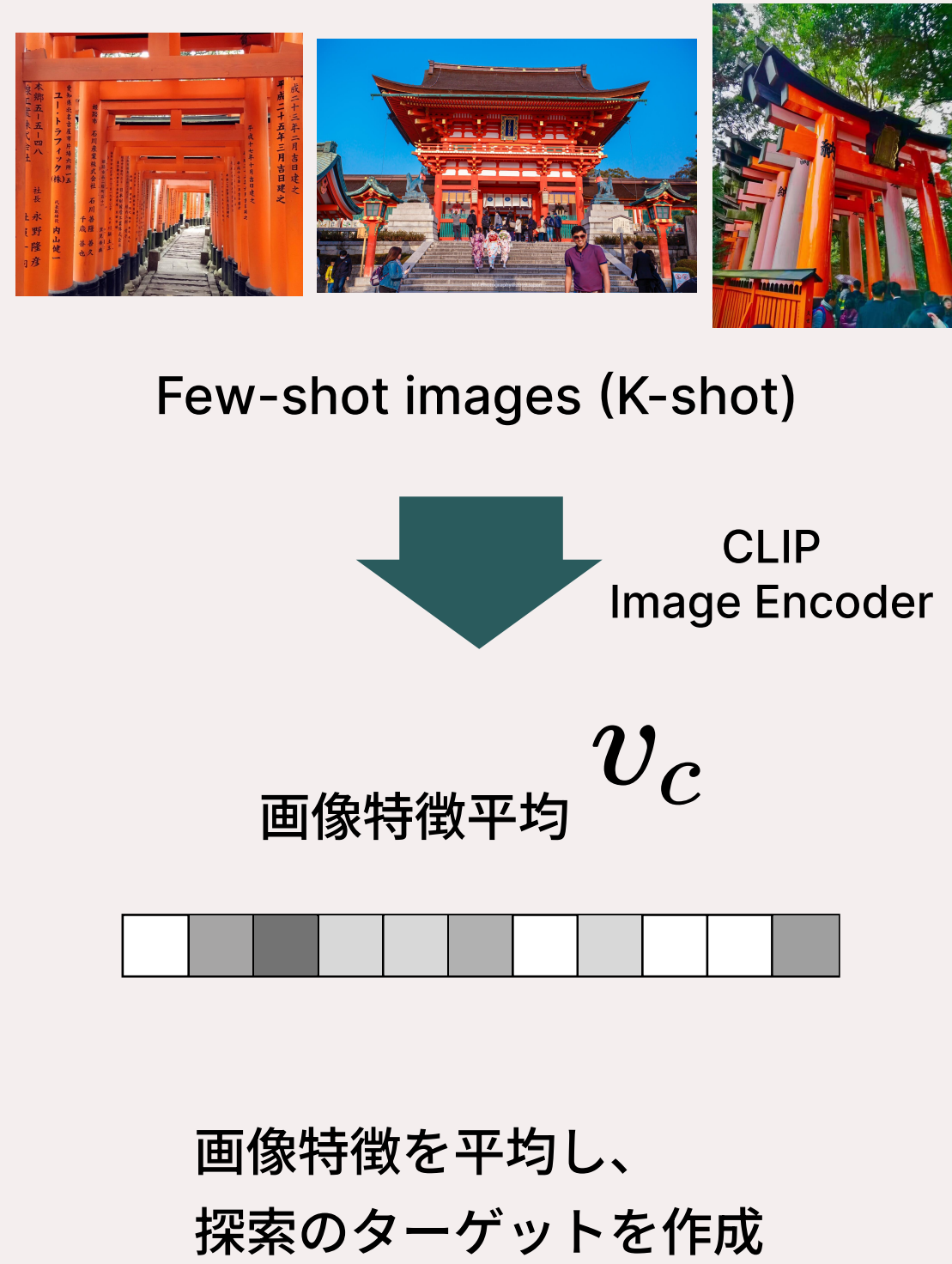


[1] Huang, X., Chen, Z., Fan, X., Wang, K., Zhou, Y., Guan, Q. and Lin, L.: Reproducible Vision-Language Models Meet Concepts Out of Pre-Training, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2025).

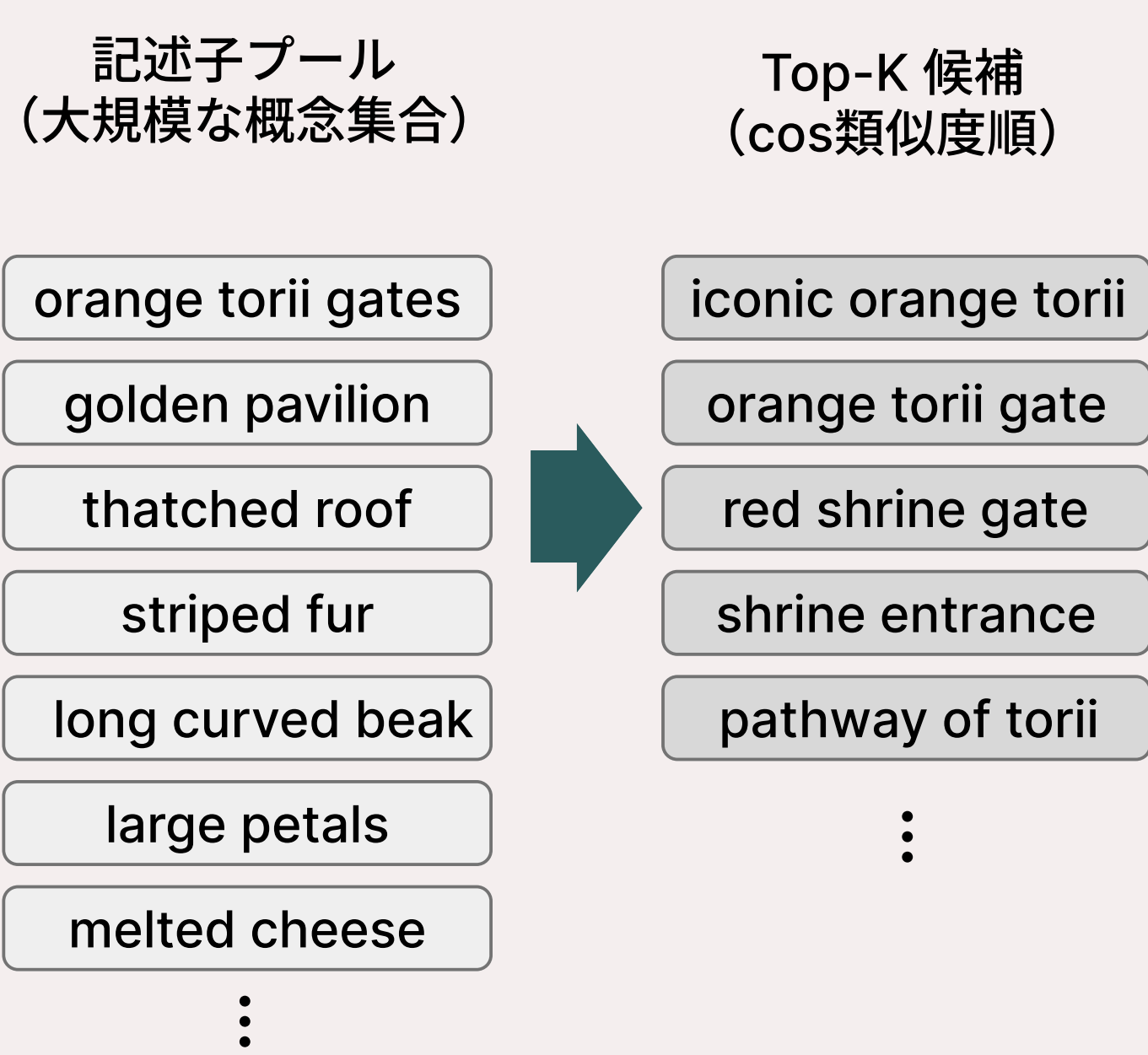
本研究の狙い：勾配学習なしで、高速かつ解釈可能なOOPクラス表現を構築する

3. 提案手法：視覚的記述子探索

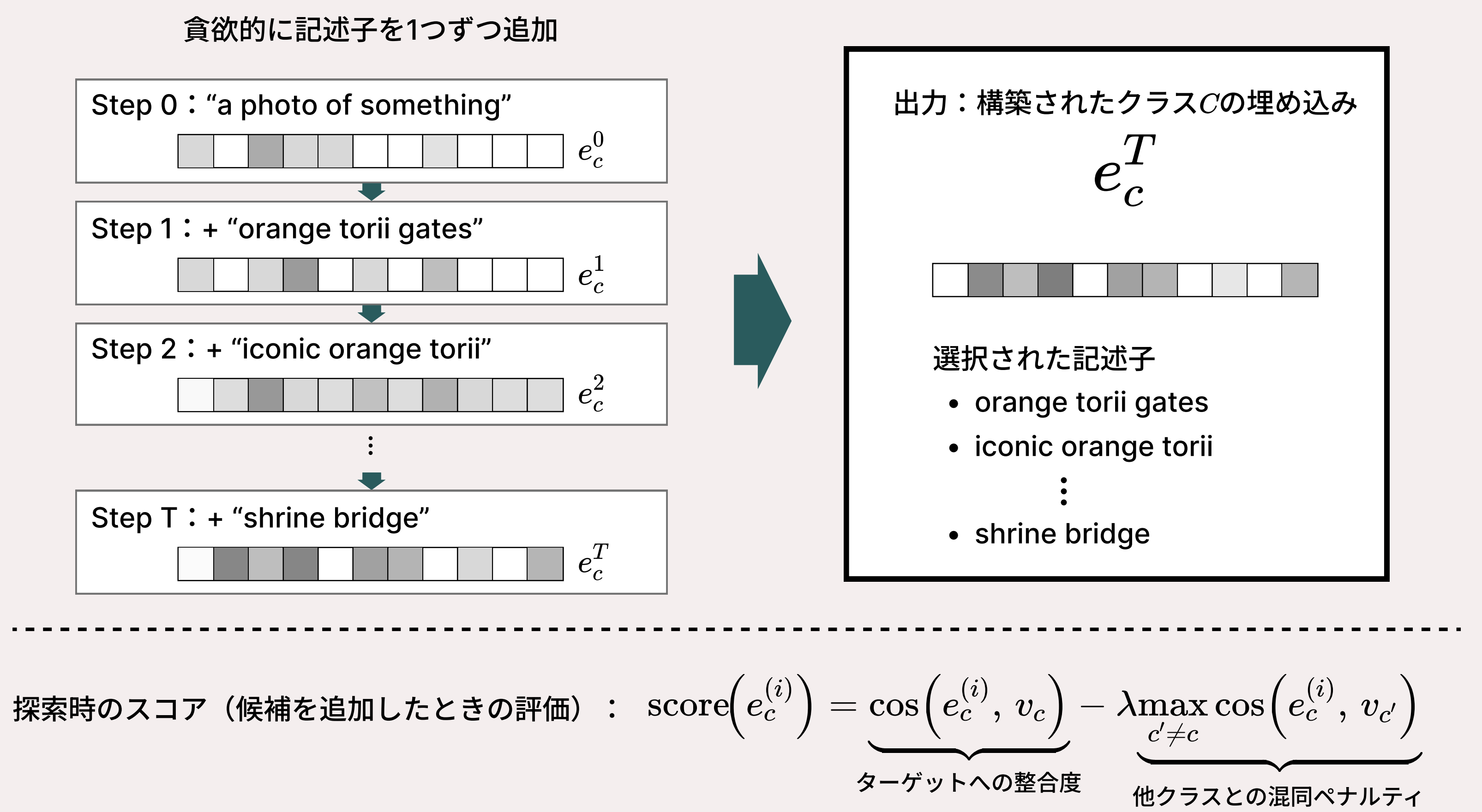
① 画像の特徴平均を算出



② cos類似度で上位候補を選択



③ 貪欲探索でクラス埋め込みを構築



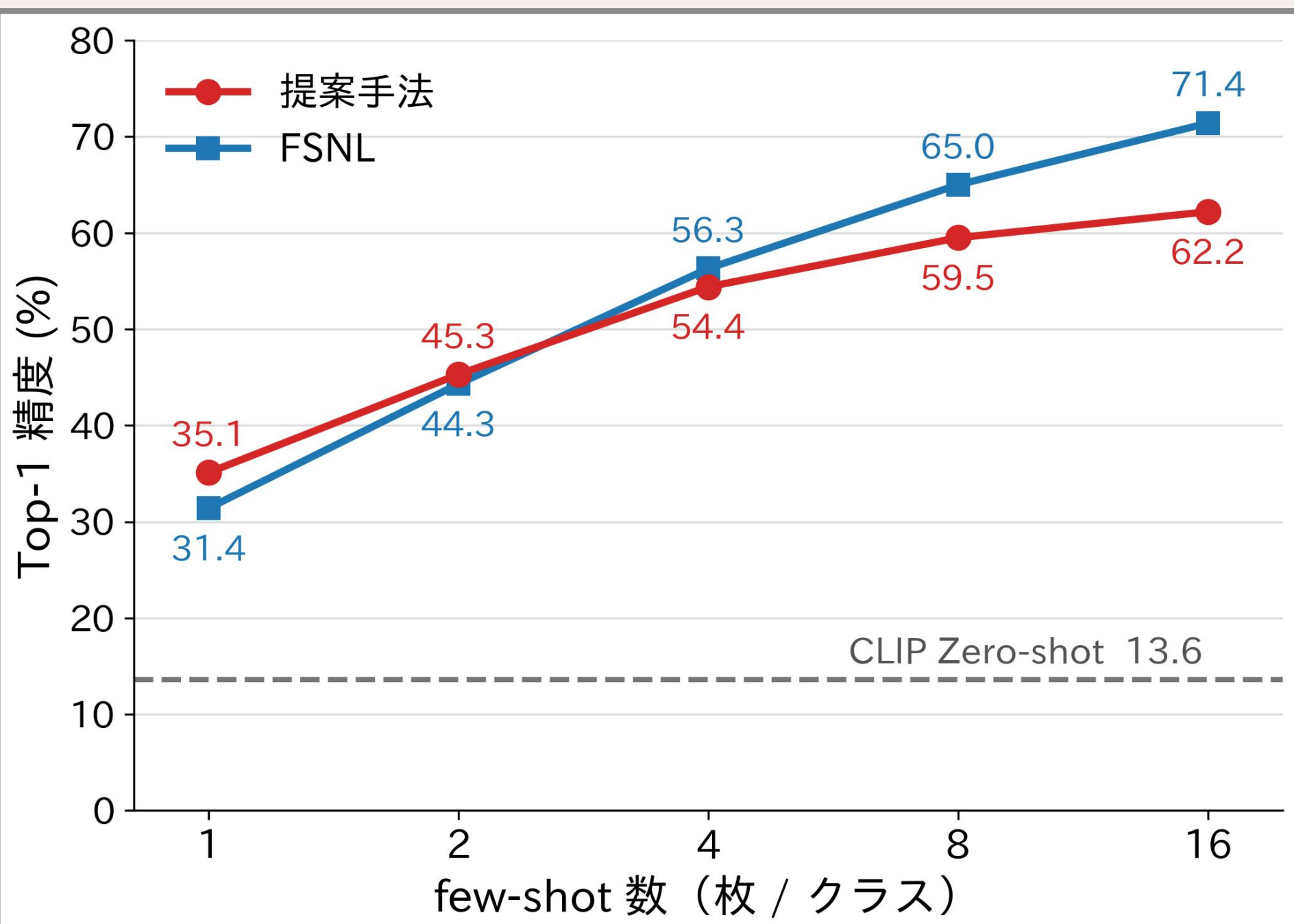
4. 評価 実験設定

- バックボーン：OpenCLIP B/16
- データセット：LAION-Beyond^[1] (9ドメイン、OOP674クラス)
- 評価指標：各ドメインにてTop-1精度を算出後、マイクロ平均を報告

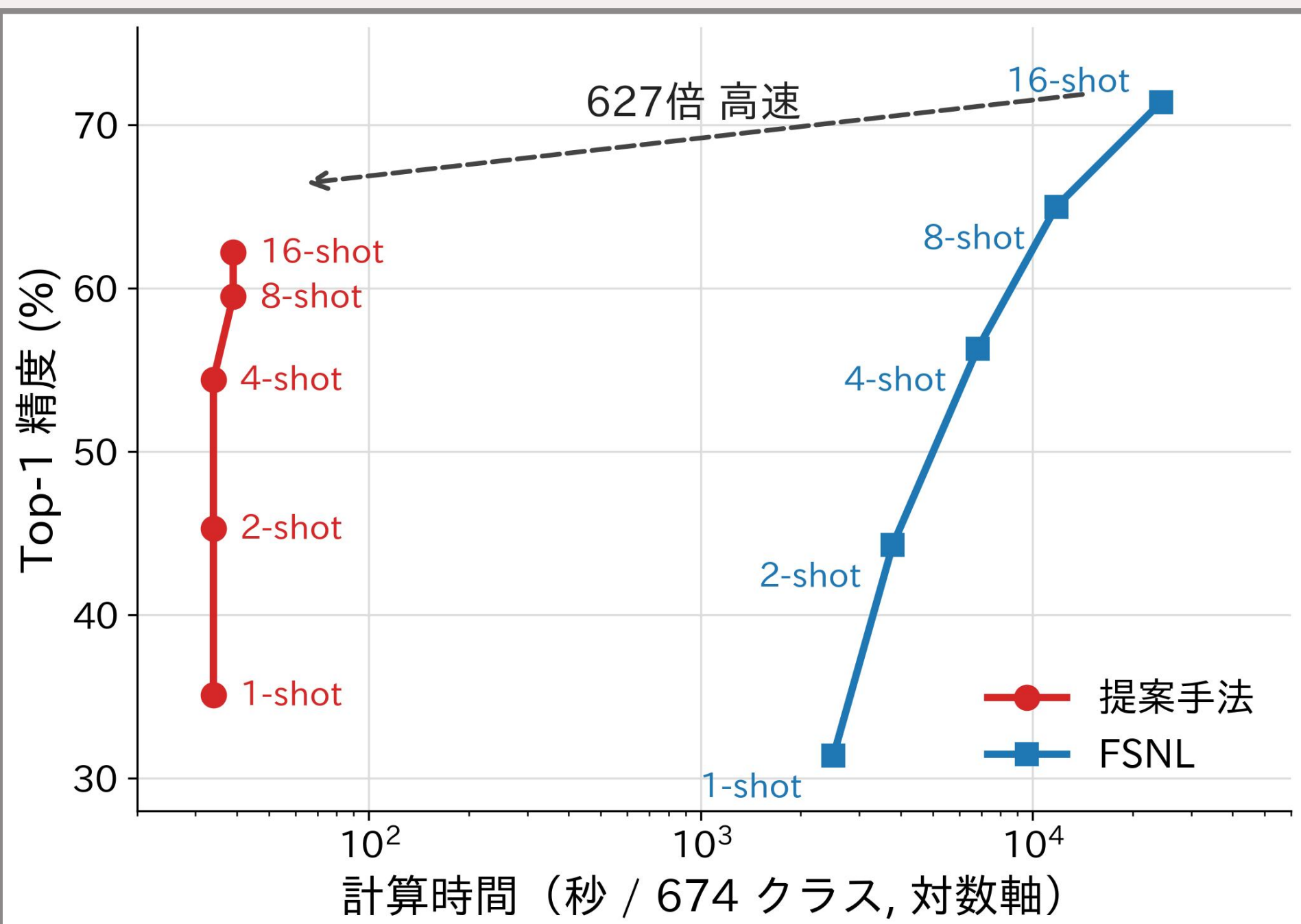
ドメイン一覧

- Animal
- Architecture
- Attire
- ForkArt
- Food
- Insects Spiders
- Landmark
- Plants
- Pokemon

A. ショット毎のTop-1 精度



B. 高速性の比較（精度 vs 計算時間）



C. 解釈性の例：伏見稲荷大社

