



Title: Iterative Finetuning for Image Geolocalization Based on Proximity Graph Segmentation
(近接グラフセグメンテーションに基づく画像地理位置推定のための反復的微調整)

Authors: Takayasu Fushimi and Rinto Koike
(伏見 卓恭(東京工科大学・CS 学部・講師), 小池 倫斗 (研究室の学生))

Journal: Applied Network Science, Vol. 10, No. 40

掲載年月: 2025 年 9 月

研究概要: 本研究は、画像の地理位置推定の精度向上を目的に、近接グラフを用いた地域分割と Vision Transformer (ViT) の反復的ファインチューニングを統合した新手法を提案する。画像の位置情報をノードとしたグラフを構築し、画像特徴の類似度をエッジ重みに設定する。ルーヴァン法でコミュニティ抽出し、地域ラベルを得て ViT を反復学習する。これにより、地理的クラスタリングと特徴表現が相互に最適化され、高精度な地理分類を実現した。

研究背景: 従来の地理位置推定は PlaNet や CPlaNet に代表されるグリッド分割を用いた手法が主流であるが、固定的な空間区分により境界誤差や地理的な連続性の無視といった課題があった。近年、画像テキスト整合を試みる方法もあるが、分類タスクへの適用性は限定的である。本研究は、地理的近接性を考慮した動的グラフ構築と、画像特徴と空間クラスタリングの反復改善を融合し、この課題に対処している。

研究成果: 提案手法により、従来の固定格子区分や事前学習画像特徴のみを用いた方法に比べ、位置推定の分類精度が向上した。近接グラフの種類（ドロネー三角分割 (DT)、Gabriel グラフ (GG)、相対近傍グラフ (RNG)）を比較しつつ、ViT の反復ファインチューニングによって画像特徴の地域表現が段階的に最適化されることで、より自然で意味的に整合した地理クラスタリングを実現した。表 1 は 4 つの都府県データに対する分類精度を表し、いずれも提案手法による 3 回反復した際の結果が最高精度を示している。

社会への影響: 本研究の高精度な画像地理位置推定技術は、位置情報付きソーシャルメディア解析、自治体の地域分析、観光関連サービス、災害時の迅速な現地把握など多様な分野に寄与する。多様な地域に適応可能な動的な地域分割モデルの開発は、ビッグデータ解析やスマートシティ構築の基盤技術として期待される。今後の大規模データ対応も見据えた発展が期待される。

表 1：分類精度

| | | Gunma | | Tochigi | | Ibaraki | | Saltama | |
|---------|------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | | Acc | F1 | Acc | F1 | Acc | F1 | Acc | F1 |
| DT | w/o | 0.289 | 0.321 | 0.331 | 0.342 | 0.307 | 0.267 | 0.179 | 0.212 |
| | FT1 | 0.677 | 0.607 | 0.592 | 0.638 | 0.649 | 0.612 | 0.567 | 0.513 |
| | FT2 | 0.793 | 0.817 | 0.766 | 0.802 | 0.791 | 0.796 | 0.696 | 0.690 |
| | FT3 | 0.851 | 0.872 | 0.817 | 0.838 | 0.860 | 0.839 | 0.753 | 0.746 |
| GG | w/o | 0.340 | 0.382 | 0.378 | 0.362 | 0.311 | 0.358 | 0.209 | 0.216 |
| | FT1 | 0.635 | 0.668 | 0.619 | 0.619 | 0.667 | 0.621 | 0.526 | 0.498 |
| | FT2 | 0.779 | 0.814 | 0.778 | 0.807 | 0.782 | 0.795 | 0.658 | 0.649 |
| | FT3 | 0.882 | 0.863 | 0.815 | 0.864 | 0.848 | 0.869 | 0.719 | 0.751 |
| RNG | w/o | 0.233 | 0.279 | 0.292 | 0.266 | 0.219 | 0.240 | 0.125 | 0.188 |
| | FT1 | 0.552 | 0.578 | 0.508 | 0.556 | 0.536 | 0.512 | 0.412 | 0.470 |
| | FT2 | 0.696 | 0.690 | 0.662 | 0.660 | 0.637 | 0.652 | 0.569 | 0.544 |
| | FT3 | 0.774 | 0.727 | 0.757 | 0.743 | 0.738 | 0.746 | 0.596 | 0.593 |
| | Mesh | 0.383 | 0.367 | 0.341 | 0.332 | 0.314 | 0.329 | 0.283 | 0.252 |
| CPlaNet | | 0.555 | 0.515 | 0.568 | 0.536 | 0.565 | 0.505 | 0.417 | 0.470 |

Bold indicates the highest value in each column

専門用語:

- **近接グラフ (Proximity Graph)** 地理的に近い地点同士をエッジでつないだグラフ構造
- **ルーヴァン法 (Louvain Method)** グラフのコミュニティ (地域) を効率的に抽出するクラスタリング手法

- **Vision Transformer (ViT)** 画像特徴を抽出するトランスフォーマーベースの深層学習モデル
- **反復的ファインチューニング (Iterative Fine-tuning)** 学習済みモデルを繰り返し最適化し、特徴表現と分類精度を同時に改善する手法