

画像検索精度評価のためのオープンソースツール

画像検索 (Image Retrieval) の評価には、平均適合率 (mAP) や Precision@K, Recall@K, NDCG などの指標が用いられます。近年、これらの評価指標を簡便に計算でき、主要なベンチマークデータセットにも対応したオープンソースの評価ツールがいくつか公開されています。以下、代表的なツールについてリンクと特徴を示し、簡単に比較します。

PyRetri – 統合型画像検索評価ライブラリ (PyTorchベース)

PyRetriはPyTorch上で実装された画像検索の統合ライブラリで、特徴抽出からインデックス構築、評価まで一貫して提供します ① ②。初の深層学習ベース画像検索用統合フレームワークとして開発され、柔軟なモジュール構成で拡張性に優れています ③。評価モジュールではRecallとmAPを主要指標としており、全クエリの平均適合率(mAP)や検索結果の再現率(Recall)を計算できます ②。Oxford5kやParis6kなどの標準データセットにも対応したデータ準備・評価スクリプトが含まれており、インストール手順や使用方法はREADMEおよびドキュメントに詳しく記載されています ④。また、PyRetriはモジュール差し替えにより様々な特徴量抽出手法に対応可能です。実際、R-MACやGeMといったグローバル記述子抽出、PCAやSVDによる次元圧縮など多数の手法が実装済みであり ⑤、CNN特徴はもちろんViTや拡散モデルによる埋め込み特徴にも応用できます (モデルを読み込んで特徴ベクトルを取得・評価する設定が可能)。GitHub上で公開されており、Apache 2.0ライセンスで自由に利用できます。

- **対応指標:** mAP, Recall (他の指標も拡張可能) ②
- **対応データセット:** Oxford5k, Paris6kなど (データセット準備スクリプトあり)
- **実装/使用:** PyTorchベース。pip install で導入後、設定ファイルまたはスクリプトで評価実行。ドキュメントとモデル動作例あり。
- **汎用性:** 高い (モジュール構成により様々なモデル・特徴量に対応) ③ ⑤

RevisitOP – Oxford/Paris再評価ベンチマーク用コード

RevisitOP (Revisiting Oxford and Paris) は、Oxford Buildings 5kやParis 6kデータセットの評価手法を刷新したベンチマーク ⑥ で、公式に提供されている評価プログラムです ⑦ ⑧。Filip RadenovićらのCVPR 2018論文に基づき、従来の注釈誤りを修正した新GT (真値) とEasy/Medium/Hardの難易度プロトコルを導入しています ⑨。GitHubで公開されているコードには、Matlab版とPython版が含まれ、**画像のダウンロード・前処理から評価までのスクリプト**が用意されています ⑩ ⑪。評価スクリプトを実行することで、クエリごとの平均適合率 (mAP) や上位1,5,10件でのPrecision (適合率) などが難易度別に出力されます ⑫。例えばROxford5kデータセットに対し、Easy/Medium/Hard各設定でmAPやP@Kが報告される形式です ⑫。Oxford5kやParis6k本体の画像もスクリプト経由で自動ダウンロードでき、インストール手順もREADMEに記載されています。特定のデータセット専用ではありますが、その分セットアップが簡単で再現性の高い評価が可能です。

- **対応指標:** mAP, Precision@K (各難易度設定で算出) ⑫
- **対応データセット:** Oxford 5k, Paris 6k (および拡張版のROxford5k, RParis6k)
- **実装/使用:** Python 3系 or MATLAB。スクリプト (example_evaluate 等) を実行するだけで評価可能 ⑧。
- **汎用性:** データセット依存 (他データセットに流用する場合はGTフォーマットの準備が必要)。主にランドマーク画像検索に特化。

Deep Image Retrieval (DIR) – グローバル記述子評価用ツール

Deep Image Retrieval (DIR)は、NAVER Labs Europeの研究により公開されたエンドツーエンド画像検索モデル実装および評価コードです¹³。ResNet+GeMなどのグローバル記述子による画像検索モデルを提案したIJCV 2017¹⁴やICCV 2019論文の再現コードであり、**評価スクリプトがPython (PyTorch)で提供**されています¹³。あらかじめ学習済みモデル (ResNet+R-MAC/GeM等) が付属し、それらを用いたOxford5KやParis6K、および再注釈版ROxford5K・RParis6Kでの検索性能 (mAP値) がREADMEで報告されています¹⁵。例えばResNet101+GeMモデルではOxford5Kで約85.7%、ROxford (Medium)で64.5%のmAPを達成しています¹⁶。このツールを使えば、**コマンドラインからモデルを指定して各種データセットで評価**でき、`--dataset` オプションでOxfordやParis等を選択すると自動で対応する評価が行われます¹⁷。Precision@Kやランキング出力も含め、論文と同じ指標算出が可能です。GitHubリポジトリにはセットアップ方法や必要データのリンクが記載され、PyTorch 1.0+で動作します。

- **対応指標:** mAP (難易度別にも計算)、ほか平均精度に関連する指標 (論文に準拠)
- **対応データセット:** Oxford5K, Paris6K, ROxford5K, RParis6K (Landmarksデータで学習済みモデル付属)¹⁵
- **実装/使用:** PyTorch 1.x。GitHubからコード取得後、モデルチェックポイントとデータパスを指定しスクリプト実行¹⁷。
- **汎用性:** 主に提供モデル (CNNベース) の評価に最適化。独自モデルを利用する場合はコード拡張が必要だが、グローバル記述子であればフレームワークを流用可能。

TREC_eval系ツール – 汎用的な検索評価ライブラリ

画像検索を含む情報検索タスク全般の評価には、**TREC評価ツール (trec_eval)** やそのPythonインターフェースである**pytrec_eval**も利用できます。trec_evalは情報検索大会(TREC)で標準的に使われてきた評価ソフトウェアで、検索結果とクエリごとの関連リスト (qrels) から多様な評価指標を計算します¹⁸。

pytrec_evalはそのPythonラッパーであり、`pip install pytrec_eval`で導入して辞書形式の入力を与えるだけで、mAP, Precision@K, Recall, NDCGなど主要指標を高速に計算できます¹⁹²⁰。例えば、各クエリに対する関連画像リストを用意すれば、pytrec_evalで**平均適合率 (MAP)** や**NDCG**を一括算出でき、評価指標ごとの詳細も取得可能です²⁰。このツール自体は特定のデータセットに依存しないため、OxfordやCOCOなど任意のデータセットに適用できます (評価用の関連アノテーションさえ用意すれば汎用的に利用可能)。ドキュメントや使用例も充実しており、カスタム実装を避けて信頼性の高い評価が行える利点があります¹⁹。

- **対応指標:** mAP, Precision@K, Recall@K, NDCG, MRR, R-Precisionなど多数 (TREC標準に準拠)²⁰
- **対応データセット:** 汎用 (クエリと関連画像の対応情報を用意すればどのデータセットでも可)。
- **実装/使用:** C言語製のtrec_evalをPythonから呼び出し。シンプルなAPIで統計量を取得可能²¹。
- **汎用性:** 極めて高い (テキスト・画像を問わずランキング評価全般に利用可能)。大規模データでも高速。

なお、この他にも**TorchMetrics** (PyTorch Lightningの評価指標集) には情報検索向けのmAPやNDCG実装が含まれており、カスタムモデルの評価指標算出に利用できます。また、Python製の評価フレームワーク**ranx**などもPrecisionやRecall, NDCGといったランキング指標を簡便に計算・比較できるツールとして登場しています。こうした汎用ライブラリを組み合わせることで、CNN特徴抽出モデルからViT・拡散モデルまで様々な画像検索手法の精度を統一的に評価できるでしょう。

各ツールのまとめ:

評価ツールは用途に応じて選択できます。例えば、特定ベンチマーク (Oxford5k等) の再現実験にはRevisitOPやDIRが便利です。一方、独自の特徴量や新しいモデルを評価したい場合は、PyRetriのような拡張

性の高いフレームワークやpytrec_eval等の汎用ライブラリが適しています。それぞれ公式のGitHubリポジトリでインストール方法や使用例が公開されており、最新の情報に基づいて評価環境を構築できます ⁴ ²²。

References:

- PyRetri GitHub リポジトリ ²³ ² (READMEおよび論文より)
- Revisiting Oxford & Paris (RevisitOP) GitHub ²⁴ ¹²
- Deep Image Retrieval (DIR) GitHub ¹³ ¹⁵
- pytrec_eval GitHub ¹⁹ ²⁰ and TREC Evalツール説明 ²⁵

¹ ² ³ PyRetri: A PyTorch-based Library for Unsupervised Image Retrieval by Deep Convolutional Neural Networks

<https://arxiv.org/pdf/2005.02154>

⁴ ⁵ ²³ GitHub - PyRetri/PyRetri: Open source deep learning based unsupervised image retrieval toolbox built on PyTorch

<https://github.com/PyRetri/PyRetri>

⁶ ⁷ ⁸ ⁹ ¹⁰ ¹¹ ¹² ²⁴ GitHub - filipradenovic/revisitop: Revisiting Oxford and Paris: Large-Scale Image Retrieval Benchmarking

<https://github.com/filipradenovic/revisitop>

¹³ ¹⁴ ¹⁵ ¹⁶ ¹⁷ GitHub - naver/deep-image-retrieval: End-to-end learning of deep visual representations for image retrieval

<https://github.com/naver/deep-image-retrieval>

¹⁸ ²⁵ usnistgov/trec_eval: Evaluation software used in the Text Retrieval ...

https://github.com/usnistgov/trec_eval

¹⁹ ²⁰ ²¹ ²² GitHub - cvangysel/pytrec_eval: pytrec_eval is an Information Retrieval evaluation tool for Python, based on the popular trec_eval.

https://github.com/cvangysel/pytrec_eval