# 類似度グラフの事前再計算による拡散式画像検索の効率化

加藤　辰弥†　　駒水　孝裕†　　井手　一郎†

† 名古屋大学　〒 464{8601 愛知県名古屋市千種区不老町

E-mail: ykatot@cs.is.i.nagoya-u.ac.jp, yytaka-coma@acm.org, yyyide@i.nagoya-u.ac.jp

あらまし　画像検索タスクにおいて，事前計算による高速な手法とリランキングに基づく高精度な手法がそれぞれ提案されているが，速度と精度を両立した手法の開発が課題になっている．我々はこれらの手法を統合することで，この課題に対処した画像検索のフレームワーク R-DiP （Re-ranking based Di usion Pre-computation）を提案する．具体的には，事前計算手法において類似度を計算する際に，リランキングで用いられる高精度な類似度再計算を導入することで，高速で高精度な検索を実現する．ベンチマークを用いた実験により，提案手法が最先端の手法と同等な検索精度を維持しつつ，高速な検索を実現できることを示す．提案手法は特に大規模なデータセットにおいて，mAP スコアを 2.0% 程度向上し，クエリごとの検索速度を平均で 75% 程度の削減をし，最先端の高精度手法を上回る検索精度を示した．提案する枠組みは，任意のリランキング手法を取り込むことが可能であり，今後出現するであろう高精度なリランキング手法に対しても効果を発揮し，画像検索技術の進展に貢献することが期待される．

キーワード　画像検索，Content-based Image Retrieval，Di usion，Re-ranking，効率化

## 1 はじめに

ディジタルカメラの低廉化やスマートフォンなどの携帯端末に搭載される高性能カメラの普及，画像編集ソフトウェアの大衆化により，個人がディジタル画像を作成することが容易になった．また，Web 上のソーシャルメディアなどのコンテンツ共有プラットフォームの普及に伴い，作成された画像を手軽に公開することができるようになり，我々がアクセスできる画像データの量が増大している．この状況に対して，画像データを管理・活用するために，画像検索は重要な技術のひとつである．画像は実世界を記録するメディアとして言葉だけでは表現することができない事象を記録することに向いており，画像検索は与えられた検索要求に基づいて，過去の事象の記録や関連する事象を探すための主要な手段の一つである．画像検索の中でも，画像を検索要求とする「Content-Based Image Retrieval（CBIR）」があり，キーワード検索やメタデータ検索では捉えきれない画像自身がもつ情報を活用するという特徴がある．

CBIR は 1990 年代 [17] から長い間研究されてきているものの，その性能向上は依然として研究課題である．その主な課題は，画像からの特徴抽出である．近年，ニューラルネットワーク技術の進歩に伴い，従来の人手による特徴量の設計から，大量のデータから自動的に適確な画像特徴を抽出する枠組みが主流となってきた．例えば，畳み込みニューラルネットワーク（Convolutional Neural Network；CNN）[1], [9], [31] や Vision Transformer（ViT）[6], [7] など，数多くの手法が提案されている．これらは検索の目的に合わせて構築されてきたわけではないため，必ずしも検索に適した特徴とは限らない．これに対して，より良い検索を実現するために，検索に適した結果の並び替えを行なうリランキングを利用する手法 [2], [22], [26], [35], [37] やランダムウォークに基づく拡散 [4], [11], [14], [46], [47] などが提案されている．この中でも，Correlation Verification Networks（CVNet）[22] のような，局所特徴をリランキングに用いる手法が高い検索精度を示している．一方で，このようなニューラルネットワークモデルを用いた類似度の推定は低コストとは言えず，検索のオーバヘッドが増し，検索速度が低下する．それに対応するために，大域特徴のみをリランキングに用いる手法 [34] も提案されているが，依然として十分に効率的とは言えない．

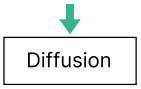



図 1: R-DiP の枠組み．

これに対して，高速な検索を実現するために，データベース画像間の類似性をランダムウォークにより事前計算する Offline Diffusion [44] と呼ばれる手法が提案されている．その基礎となっている拡散（Diffusion）モデルは，擬似適合性フィードバック [47] の考え方を用いることで，同一物体を異なる視点から写した画像や，遮蔽物により物体の一部が隠れている画像を検

索することができる手法である．具体的には，検索要求とデータベース画像から $k$ 近傍グラフを構築し，その上でランダムウォークを行なうことで，データの大域的な構造を考慮した効果的な検索を実現している．Offline Diffusion は，$k$ 近傍検索によるグラフ構築および画像間の類似度計算をデータベース側で事前に（つまりオフラインで）実行する．これにより，検索時に $k$ 近傍検索の結果との単純な線型結合によって検索を行なうことが可能になり，高速な検索を実現している．しかし，その検索精度は事前処理時のグラフ構築に大きく依存し，リランキングを利用する高精度な手法 [22], [34] と比較して劣っている．

本研究では，以上のような速度と精度の間のトレードオフを解決するための手法として，事前計算に基づく高速な手法である Offline Diffusion とリランキングを利用する高精度な手法を融合するための新しいフレームワーク R-DiP （Re-ranking based Diffusion Pre-computation）を提案する．R-DiP の枠組み（図 1）は，Offline Diffusion の近傍検索や関連する類似度検索に高性能なリランキング手法を取り込むことで精度の向上を図る．これにより，検索における関連度を計算する任意のリランキング手法を取り込むことができる柔軟性がある．本研究では，ベンチマークを用いた実験により，特に大規模データセットにおいて，提案手法がリランキングに基づく SuperGlobal [34] に匹敵する検索性能を維持しつつ，Offline Diffusion と同等に高速な検索を実現できることを示す．

本研究の主な貢献は以下の通りである．

- **フレームワーク R-DiP の提案**：Offline Diffusion を基礎とし，リランキング手法の利点を取り込むフレームワークを提案する．提案フレームワークは検索における関連度を計算する任意のリランキング手法を取り込める柔軟性があるため，今後より検索精度が高い高計算コストなリランキング手法が出現した場合にも適用可能である．
- **高速・高性能な検索の実現**：R-DiP は，リランキングに基づく最先端手法である SuperGlobal に匹敵する検索精度を維持しつつ，Offline Diffusion と同等に高速な検索を達成する．特に，大規模なデータセットにおいて，mAP スコアを Offline Diffusion の特性により，最先端の手法より 2.0% 程度向上したうえで，クエリごとの検索速度を平均で 75% 程度の削減をし，最先端の手法を上回る検索性能を示した．

## 2 関 連 研 究

画像検索における重要な課題は，画像特徴の設計である．機械学習の進歩に伴い，コンピュータビジョンに関する研究として，CNN [19], [21], Regional Maximum Activation of Convolutions（R-MAC）[42]，ViT [5]，GAN [8] などのニューラルネットワークによる手法が登場してきた．これらの手法は画像検索にも採用され [1], [6], [7], [9], [10], [18], [23], [25], [31], [32], [36], [38]，従来の問題であったセマンティックギャップを克服し，画像検索の精度を大幅に向上させている．

画像特徴の設計に続き，画像検索の特性を考慮した検索の最適化に向けた手法が提案されている．例えば，$k$ 近傍検索に基づく検索 [3], [24]，上位に順位付けられた検索結果の類似度を再計算するリランキング，ランダムウォークに基づく拡散 [4], [11], [14], [46], [47] などがある．その中で，リランキングを利用した手法は検索精度の面で優れており，拡散の中でも事前計算に基づく手法 [44] や $k$ 近傍検索に基づく手法は速度の面で優れていることが明らかにされている．以下に，リランキングおよび拡散に基づく手法について詳細に述べる．

a） 画像検索におけるリランキング

リランキングは，2 段階の検索手法であり，始めに軽量な手法を用いて検索結果を大まかにフィルタリングした後，高精度な類似度計算モデルによって順序付けを行なう．リランキングを利用した CBIR 手法では，最初の段階で大域的特徴に基づいて検索し，物体の局所的な関係性を捉える幾何検証 (Geometric Verification) により類似度を再計算する手法が主流である．幾何検証に関する研究としては，特徴点検出に焦点をあてた研究 [20], [27], [33], [45] や特徴抽出に焦点をあてた研究 [13], [16], [26], [33], [35], [40], [41], [45]，特徴点検出と抽出の順番を入れ替えた研究 [39] など，様々な手法が提案されている．

さらに，幾何検証の代替手法として，ニューラルネットワークを用いて抽出した局所特徴マップを用いる手法が登場した．Transformer を利用した Re-Ranking Transformer（RRT）[37]，局所特徴と大域特徴を効果的に統合した Super-Feature を提案する研究 [43]，局所特徴を用いて複数の異なる大きさで画像間の類似度を計算する CVNet [22] などが提案されている．これらの手法は高精度な検索結果を示すが，速度に関するオーバヘッドが大きいことが問題である．より低オーバヘッドな手法として，リランキングに大域特徴のみを用いる SuperGlobal [34] が提案されているが，事前計算に基づく高速な手法 [44] などと比較すると速度の面で改善の余地が残っている．

b） 拡散（Diffusion）/ Offline Diffusion

拡散 [14] の核となる考え方は，クエリ画像とデータベース内の画像の類似性をデータベース内の類似画像グラフを通して拡散させることである．拡散の主要な利点は，$k$ 近傍検索のように単に Euclidean 空間上での画像特徴の距離に依存するのではなく，近接性に基づくネットワーク構造を考慮することによって，間接的に類似している画像を発見できる点にある．このような違いにより，例えば，同一物体を異なる視点から写した画像や，遮蔽物により物体の一部が隠れている画像の検索が可能になる．しかし，特に大規模なデータセットに適用する際には，ランダムウォークの計算コストが高いため，検索に時間を要するという欠点がある．

この問題に対処するために，Offline Diffusion [44] が提案されている．Offline Diffusion では，拡散過程がデータベース側で事前に実行されることで，検索時の計算負荷を大幅に軽減できる．具体的には，拡散を行ない事前計算した到達確率と $C = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n : \mathbf{c}_i \in \mathbb{R}^n\}$（ここで，$n$ はデータベース画像枚数）と $k$ 近傍検索で求める $y$ の線型結合として，以下のようにして画像間の類似度を高速に取得する．

（1） $k_q$ 近傍検索：$k_q$ は初期の $k$ 近傍検索における検索枚数である．クエリ画像 $q$ に対して，類似画像のインデックス

表 1: 提案手法と比較手法の精度評価（mAP）.

| Method | Medium | | | | Hard | | | |
|---|---|---|---|---|---|---|---|---|
| | $\mathcal{R}$Oxf | +$\mathcal{R}$1M | $\mathcal{R}$Par | +$\mathcal{R}$1M | $\mathcal{R}$Oxf | +$\mathcal{R}$1M | $\mathcal{R}$Par | +$\mathcal{R}$1M |
| Offline Diffusion [44] ($k_q$=100) | 89.12 | 85.78 | 95.22 | 92.09 | 75.95 | 71.11 | 88.99 | 84.63 |
| Offline Diffusion [44] ($k_q$=400) | 89.88 | 85.91 | 95.37 | 92.75 | 77.06 | **72.10** | 89.29 | 85.20 |
| SuperGlobal [34] ($k_q$=100) | 88.26 | 81.76 | 92.20 | 83.24 | 76.67 | 67.74 | 83.79 | 67.66 |
| SuperGlobal [34] ($k_q$=400) | **90.79** | 84.08 | 93.36 | 85.53 | **80.05** | 70.93 | 86.77 | 72.52 |
| R-DiP($k_q$=100) | 89.71 | **86.55** | 95.39 | 92.94 | 76.94 | 71.97 | 90.68 | **85.61** |
| R-DiP($k_q$=400) | 90.40 | **86.55** | **95.79** | **93.15** | 77.67 | 71.80 | **91.30** | 85.46 |

表 2: 最適なパラメータ.

| Method | Medium | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{R}$Oxf | | | | +$\mathcal{R}$1M | | | | $\mathcal{R}$Par | | | | +$\mathcal{R}$1M | | | |
| | $k_m$ | $k_d$ | $k_t$ | $\gamma$ | $k_m$ | $k_d$ | $k_t$ | $\gamma$ | $k_m$ | $k_d$ | $k_t$ | $\gamma$ | $k_m$ | $k_d$ | $k_t$ | $\gamma$ |
| Offline Diffusion [44] ($k_q$=100) | — | 200 | 400 | 15 | — | 60 | 400 | 15 | — | 200 | 800 | 5 | — | 200 | 800 | 5 |
| Offline Diffusion [44] ($k_q$=400) | — | 200 | 200 | 10 | — | 40 | 100 | 20 | — | 200 | 800 | 10 | — | 100 | 800 | 10 |
| R-DiP($k_q$=100) | 100 | 60 | 800 | 10 | 400 | 100 | 800 | 10 | 400 | 200 | 800 | 5 | 400 | 200 | 800 | 5 |
| R-DiP($k_q$=400) | 100 | 60 | 800 | 10 | 400 | 60 | 800 | 20 | 400 | 200 | 800 | 10 | 400 | 200 | 800 | 10 |

| Method | Hard | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mathcal{R}$Oxf | | | | +$\mathcal{R}$1M | | | | $\mathcal{R}$Par | | | | +$\mathcal{R}$1M | | | |
| | $k_m$ | $k_d$ | $k_t$ | $\gamma$ | $k_m$ | $k_d$ | $k_t$ | $\gamma$ | $k_m$ | $k_d$ | $k_t$ | $\gamma$ | $k_m$ | $k_d$ | $k_t$ | $\gamma$ |
| Offline Diffusion [44] ($k_q$=100) | — | 200 | 400 | 15 | — | 100 | 400 | 10 | — | 200 | 800 | 5 | — | 200 | 800 | 5 |
| Offline Diffusion [44] ($k_q$=400) | — | 200 | 200 | 10 | — | 40 | 100 | 20 | — | 80 | 400 | 15 | — | 100 | 800 | 10 |
| R-DiP($k_q$=100) | 400 | 400 | 400 | 10 | 400 | 100 | 800 | 10 | 400 | 200 | 800 | 10 | 400 | 80 | 800 | 5 |
| R-DiP($k_q$=400) | 400 | 400 | 200 | 10 | 400 | 40 | 100 | 20 | 400 | 200 | 800 | 10 | 400 | 80 | 800 | 10 |

これらのデータセットはそれぞれ 70 個のクエリがあり，それぞれに異なる数の検索対象画像が割り当てられている．データベースに含まれる画像の枚数は，それぞれ 4,933 枚と 6,322 枚である．各クエリに対して，対応するデータベース画像には，検索の難易度に応じて *unclear*，*easy*，*hard* の三つのラベルが付与されている．そして，使用するラベルに基づいて三つ（Easy，Medium，Hard）の評価プロトコルに分かれる．

- Easy プロトコル： *easy* ラベルのみ
- Medium プロトコル： *easy* と *hard* ラベル
- Hard プロトコル： *hard* ラベルのみ

これまでの研究では，Easy プロトコルは既に十二分な性能が達成されており，検索手法の評価に適さないと言われているため，本研究では，Medium プロトコルと Hard プロトコルで実験を行なう．加えて，大規模なデータセットにおける検索性能を評価するために，クエリ画像に含まれない建造物や風景などの約 100 万枚の類似ドメインの画像からなる Distractor セット（$\mathcal{R}$1M）も用意されている．本実験では，Distractor セットを追加した実験も行なう．

評価指標には mAP（mean Average Precision）を用いる．

### 4.3 実装の詳細

本実験では，提案手法と全ての比較手法において，SuperGlobal Pooling によって抽出された 2,048 次元の大域特徴を用いる．また，提案手法における類似度再計算モデルには SuperGlobal Reranking を用いる．$k$ 近傍検索には，Offline Diffusion 同様に Facebook AI Similarity Search（FAISS）ツールキット [15] を用いる．

各手法について，実験で用いた変数を以下に示す．

- **R-DiP**

提案手法の変数 $k_m$，$k_d$，$k_t$, $k_q$，$\gamma$ は次のように設定した．

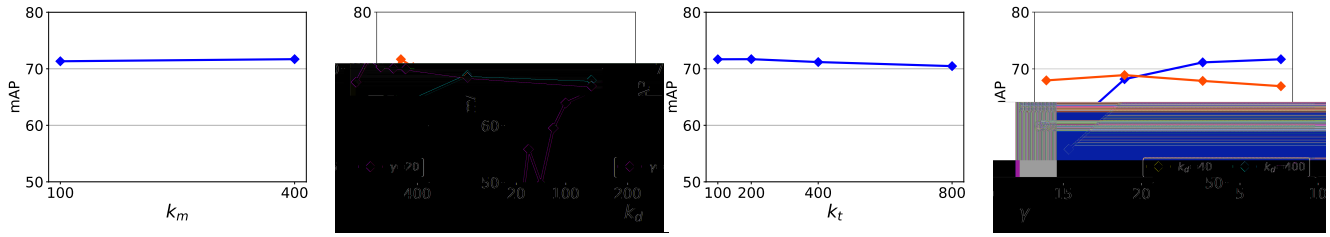
– オフライン過程における $k$ 近傍検索による初期類似画像検索で取得する件数（Offline $k$-NN Size）： $k_m \in \{100, 400\}$

– オフライン過程における近傍グラフ構築時の近傍数（Offline Graph Degree）： $k_d \in \{20, 40, 60, 80, 100, 200, 400\}$

– オフライン過程における Late Truncation のための $k$ 近傍検索での画像インデックス取得件数（Offline Truncation Size）： $k_t \in \{100, 200, 400, 800\}$

– オンライン検索時の $k$ 近傍検索により取得する画像件数（Online $k$-NN Size）： $k_q \in \{100, 400\}$

– オンライン検索時の類似度に対する重み変数（Online




(a) $k_m$ (b) $k_d$ (c) $k_t$ (d) $\gamma$

**図 4: 各パラメータが与える影響.**

**表 3: クエリごとのオンライン検索時間の平均［ms］.**

| 手法 | $k_q$=100 | | $k_q$=400 | |
|---|---|---|---|---|
| | $\mathcal{R}$Oxf | +$\mathcal{R}$1M | $\mathcal{R}$Oxf | +$\mathcal{R}$1M |
| Offline Diffusion [44] | **1.18** | 37.21 | 1.31 | 40.18 |
| SuperGlobal [34] | 17.74 | 41.43 | 139.34 | 162.86 |
| R-DiP | 1.19 | **37.14** | **1.30** | **40.01** |

Weight)： $\gamma \in \{5, 10, 15, 20\}$

- **Offline Diffusion**

Offline Diffusion では，$k_d$，$k_t$，$k_q$，$\gamma$ に関して提案手法と同じ変数集合を用いる．なお，Offline Diffusion において類似度再計算過程は存在しないため，$k_m$ は存在しない．

- **SuperGlobal**

SuperGlobal では， $k_q$ に関して提案手法と同じ変数集合を用いる．提案手法におけるその他の変数は，オフライン過程における拡散のための変数であるため，存在しない．

### 4.4 実験結果

提案手法の性能を評価するための実験結果を示す．

#### 4.4.1 検索性能 (mAP) の比較

表 1 に比較手法と R-DiP の検索精度（mAP）を示す．また，各評価プロトコルにおいて，各手法で最良の結果を示した変数を表 2 に示す．提案手法は，$\mathcal{R}$Oxford5k データセットでは SuperGlobal に匹敵する性能を，$\mathcal{R}$Paris6k データセットでは SuperGlobal を上回る精度を示した．また，Offline Diffusion との比較では，ほとんどの結果において提案手法が優れた精度を示した．これは，提案手法における類似度再計算が Offline Diffusion における近傍グラフ構築を改善したことを示す．また，大規模データセットにおける実験では，提案手法はほとんどの場合で他の比較手法を上回る精度を示した．特に，$\mathcal{R}$Oxford5k+$\mathcal{R}$1M データセットにおける Hard プロトコルでは，2.0%程度の mAP の向上が見られた．Offline Diffusion もまた，大規模データセットにおける実験では，SuperGlobal を上回る精度を示しており，Offline Diffusion の良い特性を取り入れることができていることが示唆される．

#### 4.4.2 オンライン検索時間

次にオンライン検索の実行時間を評価する．$\mathcal{R}$Oxford5k データセットと+$\mathcal{R}$1M セットのクエリあたりの平均オンライン検索実行時間（単位 ms）を表 3 に示す．提案手法は SuperGlobal と比較して実行時間を大幅に削減し，Offline Diffusion と同等の実行時間で検索できることが示された．

#### 4.4.3 パラメータ感度分析

実験で用いる変数が提案手法に与える影響を調べるために，$\mathcal{R}$Oxford5k+$\mathcal{R}$1M データセットの Hard プロトコルにおいて，各変数に対する感度分析を行なった．

- **Offline $k$-NN Size（$k_m$）**：図 4(a) に示すように，$k_m = 100$ に比べて $k_m = 400$ の時にわずかに高い mAP を示した．理由として，$k_m$ の値が大きいほど構築される類似度グラフの質が向上し，拡散による類似度計算の精度が向上したためであると考えられる．ただし，事前計算で用いる類似度再計算モデルによる計算コストは $k_m$ の値に比例することをふまえると，実運用では $k_m = 100$ で十分効果的であると考えられる．
- **Offline Graph Degree（$k_d$）**：$k_d = 20$ の時のみほとんどの場合で最低の mAP を示した．また，$k_d$ の変化が mAP に与える影響は，$\gamma$ の値によって変わる傾向が見られた．例として，図 4(b) に $\gamma \in \{5, 20\}$ の場合の mAP を示す．$\gamma$ の値が小さいほど，$k_d$ の値が大きい場合に高い mAP を示し，$\gamma$ の値が大きいほど，$k_d$ の値が小さい場合に高い mAP を示した．
- **Offline Truncation Size（$k_t$）**：他の変数と比較して $k_t$ の値の変化は mAP の大きな影響を与えなかった．また，mAP が最良となる $k_t$ の値に対しても，特に規則性は見られなかった．例として，図 4(c) に実験で最良の mAP を示した際に $k_t$ を変化させたグラフを示す．
- **Online Weight（$\gamma$）**：前述の通り，mAP が高い変数集合では $\gamma$ の値が大きいものが多かった．また，$\gamma$ の値が mAP に与える影響として，$k_d$ の分析同様に，$k_d$ の値が大きいほど，$\gamma$ の値が小さい場合に高い mAP を示し，$k_d$ の値が小さいほど，$\gamma$ の値が大きい場合に高い mAP を示した．例として，図 4(d) に $k_d \in \{40, 400\}$ の場合の $\gamma$ と mAP を示す．

#### 4.4.4 考　　察

精度評価において，提案手法は概ね良好な結果を示し，ほとんどの場合において提案手法は Offline Diffusion を上回る精度を示した．このことは，Offline Diffusion に対して類似度再計算モデルを組み込むことの有効性を示している．

また，提案手法は大規模データセットにおいて SuperGlobal の精度を大きく上回った．SuperGlobal のようなリランキングに基づく手法は，初期の $k$ 近傍検索における検索結果画像集合に大きく依存するのに対し，拡散に基づく手法はデータの大域

的な構造を考慮するため，画像枚数が多い場合に特にその有効性を示すという特性によるためと考えられる．

一方で，提案手法は特に $\mathcal{R}$Oxford5k データセットの Hard プロトコルにおいて，Offline Diffusion と同様に SuperGlobal に劣る精度を示した．この結果は，この状況では Offline Diffusion の有効性が低いことを示し，提案手法もその特性に影響を受けているためであると考えられる．この結果に対し，具体的に二つの理由が考えられる．一つ目は，$\mathcal{R}$Oxford5k データセットにおいて，データの枚数が少ないことにより，大域的構造がもつ情報量の少なさである．拡散は，その特性により大域的構造がもつ情報の質や量に対して大きく影響を受ける．このことにより，特に $\mathcal{R}$Oxford5k データセットの Hard プロトコルのような画像間類似度推定の難易度が高い場合において，画像間の類似度を直接計算するリランキングに基づく手法に比べて，提案手法のような拡散に基づく手法は有効性が低下したと考えられる．二つ目の理由として，Offline Diffusion の特性である，検索結果の初期の $k$ 近傍検索結果における画像間類似度への依存性に影響を受けていることが考えられる．Offline Diffusion は事前計算した類似度に対して初期の $k$ 近傍検索結果における画像間類似度を線型結合する．初期の $k$ 近傍検索結果における画像間類似度により，ランダムウォークが遷移する画像群は大きく変化し，検索結果に大きな影響を与える．これにより，特に高難易度な場面において，オンライン検索で使用される $k$ 近傍検索がボトルネックとなり，提案手法のような拡散に基づく手法の有効性が低下すると考えられる．

## 5 ま と め

本論文では，画像検索タスクにおける速度と精度のトレードオフに対処する，新しい画像検索フレームワークを提案した．これは，リランキング手法で用いられる高精度な類似度再計算を Offline Diffusion [44] に組み込むことで達成された．実験の結果，提案手法は $\mathcal{R}$Oxford5k [30] と $\mathcal{R}$Paris6k [30] の両データセットにおいて，Offline Diffusion と同等の検索速度を維持しながら，最先端の手法よりも優れている，もしくは同等であることが示された．大規模データセットにおける実験では，提案手法は全てのベースラインと比較して優れた結果を残し，これは，現実世界で大規模データを扱うシナリオに適しており，画像検索技術の進展に貢献することが期待される．

### 謝　辞

本研究の一部は JSPS 科研費 21H03555，NII との共同研究による．

### 文　献

[1] A. Babenko, A. Slesarev, A. Chigorin, and V. S. Lempitsky. Neural codes for image retrieval. In *Computer Vision —ECCV 2014— 13th European Conf., Zurich, Switzerland, September 6–12, 2014, Procs. Part I, Lecture Notes in Computer Science*, volume 8689, pages 584–599. Springer, 2014.

[2] B. Cao, A. Araujo, and J. Sim. Unifying deep local and global features for image search. In *Computer Vision — ECCV 2020— 16th European Conf., Glasgow, UK, August 23–28, 2020, Procs. Part XX, Lecture Notes in Computer Science*, volume 12365, pages 726–743. Springer, 2020.

[3] T. Dharani and I. L. Aroquiaraj. Content based image retrieval system using feature classification with modified kNN algorithm. *Computing Research Repository arXiv Preprint*, arXiv:1307.4717, 2013.

[4] M. Donoser and H. Bischof. Diffusion processes for retrieval revisited. In *Proc. 2013 IEEE Conf. on Computer Vision and Pattern Recognition, Portland, OR, USA*, pages 1320–1327, 2013.

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proc. 9th Int. Conf. on Learning Representations, Online*, 22 pages, 2021.

[6] S. R. Dubey, S. K. Singh, and W. Chu. Vision transformer hashing for image retrieval. In *Proc. 2022 IEEE Int. Conf. on Multimedia and Expo, Taipei, Taiwan*, 6 pages, 2022.

[7] A. El-Nouby, N. Neverova, I. Laptev, and H. Jégou. Training vision transformers for image retrieval. *Computing Research Repository arXiv Preprint*, arXiv:2102.05644, 2021.

[8] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27:2672–2680, 2014.

[9] A. Gordo, J. Almazán, J. Revaud, and D. Larlus. Deep image retrieval: Learning global representations for image search. In *Computer Vision —ECCV 2016— 14th European Conf., Amsterdam, the Netherlands, October 11–14, 2016, Procs. Part VI, Lecture Notes in Computer Science*, volume 9910, pages 241–257. Springer, 2016.

[10] A. Gordo, J. Almazán, J. Revaud, and D. Larlus. End-to-end learning of deep visual representations for image retrieval. *Int. J. Computer Vision*, 124(2):237–254, 2017.

[11] L. J. Grady. Random walks for image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, 2006.

[12] Ç. Gülçehre, K. Cho, R. Pascanu, and Y. Bengio. Learned-norm pooling for deep feedforward and recurrent neural networks. In *Machine Learning and Knowledge Discovery in Databases —European Conf., ECML PKDD 2014, Nancy, France, September 15–19, 2014. Proceedings, Part I, Lecture Notes in Computer Science*, volume 8724, pages 530–546. Springer, 2014.

[13] K. He, Y. Lu, and S. Sclaroff. Local descriptors optimized for average precision. In *Proc. 2018 IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA*, pages 596–605, 2018.

[14] A. Iscen, G. Tolias, Y. Avrithis, T. Furon, and O. Chum. Efficient diffusion on region manifolds: Recovering small objects with compact CNN representations. In *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA*, pages 926–935, 2017.

[15] J. Johnson, M. Douze, and H. Jégou. Billion-scale similarity search with GPUs. *IEEE Trans. Big Data*, 7(3):535–547, 2021.

[16] H. Jun, B. Ko, Y. Kim, I. Kim, and J. Kim. Combination of multiple global descriptors for image retrieval. *Computing Research Repository arXiv Preprint*, arXiv:1903.10663, 2019.

[17] T. Kato. Database architecture for content-based image retrieval. In *Image Storage and Retrieval Systems, Proc. SPIE*, volume 1662, pages 112–123, 1992.

[18] J. Kim and S. Yoon. Regional attention based deep feature for image retrieval. In *Proc. British Machine Vision*

*Conf. 2018, Newcastle-upon-Tyne, England, UK*, number 209, pages 1–13, 2018.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:1106–1114, 2012.

[20] A. B. Laguna, E. Riba, D. Ponsa, and K. Mikolajczyk. Key.Net: Keypoint detection by handcrafted and learned CNN filters. In *Proc. 17th IEEE/CVF Int. Conf. on Computer Vision, Seoul, Korea*, pages 5835–5843, 2019.

[21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proc. IEEE*, 86(11):2278–2324, 1998.

[22] S. Lee, H. Seong, S. Lee, and E. Kim. Correlation verification for image retrieval. In *Proc. 2022 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, New Orleans, LA, USA*, pages 5364–5374, 2022.

[23] F. Magliani and A. Prati. An accurate retrieval through R-MAC+ descriptors for landmark recognition. In *Proc. 12th Int. Conf. on Distributed Smart Cameras, Eindhoven, the Netherlands*, number 6, pages 1–6, 2018.

[24] H. Nezamabadi-pour and E. Kabir. Concept learning by fuzzy $k$-NN classification and relevance feedback for efficient image retrieval. *Expert Systems and Application*, 36(3):5948–5954, 2009.

[25] T. Ng, V. Balntas, Y. Tian, and K. Mikolajczyk. SOLAR: Second-Order Loss and Attention for image Retrieval. In *Computer Vision —ECCV 2020— 16th European Conf., Glasgow, UK, August 23–28, 2020, Procs. Part XXV, Lecture Notes in Computer Science*, volume 12370, pages 253–270, 2020.

[26] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han. Large-scale image retrieval with attentive deep local features. In *Proc. 16th IEEE Int. Conf. on Computer Vision, Venice, Veneto, Italy*, pages 3476–3485, 2017.

[27] Y. Ono, E. Trulls, P. Fua, and K. M. Yi. LF-Net: Learning local features from images. *Advances in Neural Information Processing Systems*, 31:6237–6247, 2018.

[28] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. 2007 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Minneapolis, MI, USA*, 8 pages, 2007.

[29] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *Proc. 2008 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Anchorage, AK, USA*, 8 pages, 2008.

[30] F. Radenovic, A. Iscen, G. Tolias, Y. Avrithis, and O. Chum. Revisiting Oxford and Paris: Large-scale image retrieval benchmarking. In *Proc. 2018 IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA*, pages 5706–5715, 2018.

[31] F. Radenovic, G. Tolias, and O. Chum. CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples. In *Computer Vision —ECCV 2016— 14th European Conf., Amsterdam, the Netherlands, October 11–14, 2016, Procs. Part I, Lecture Notes in Computer Science*, volume 9905, pages 3–20. Springer, 2016.

[32] F. Radenovic, G. Tolias, and O. Chum. Fine-tuning CNN image retrieval with no human annotation. *IEEE Trans. Pattern Analysis and Machine Intelligence.*, 41(7):1655–1668, 2019.

[33] J. Revaud, C. R. de Souza, M. Humenberger, and P. Weinzaepfel. R2D2: Reliable and Repeatable Detector and Descriptor. *Advances in Neural Information Processing Systems*, 32:12405–12415, 2019.

[34] S. Shao, K. Chen, A. Karpur, Q. Cui, A. Araujo, and B. Cao. Global features are all you need for image retrieval and reranking. In *Proc. 19th IEEE/CVF Int. Conf. on Computer Vision, Paris, France*, pages 11036–11046, 2023.

[35] O. Siméoni, Y. Avrithis, and O. Chum. Local features and visual words emerge in activations. In *Proc. 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA*, pages 11651–11660, 2019.

[36] J. Song, T. He, L. Gao, X. Xu, A. Hanjalic, and H. T. Shen. Binary generative adversarial networks for image retrieval. In *Proc. 32nd AAAI Conf. on Artificial Intelligence, New Orleans, LA, USA*, pages 394–401, 2018.

[37] F. Tan, J. Yuan, and V. Ordonez. Instance-level image retrieval using reranking transformers. In *Proc. 18th IEEE/CVF Int. Conf. on Computer Vision, Montreal, QC, Canada*, pages 12085–12095, 2021.

[38] M. Teichmann, A. Araujo, M. Zhu, and J. Sim. Detect-to-retrieve: Efficient regional aggregation for image search. In *Proc. 2019 IEEE Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA*, pages 5109–5118, 2019.

[39] Y. Tian, V. Balntas, T. Ng, A. B. Laguna, Y. Demiris, and K. Mikolajczyk. D2D: Keypoint extraction with Describe to Detect approach. In *Computer Vision —ACCV 2020— 15th Asian Conf. on Computer Vision, Kyoto, Japan, November 30–December 4, 2020, Revised Selected Papers, Part III, Lecture Notes in Computer Science*, volume 12624, pages 223–240. Springer, 2020.

[40] Y. Tian, X. Yu, B. Fan, F. Wu, H. Heijnen, and V. Balntas. SOSNet: Second Order Similarity regularization for local descriptor learning. In *Proc. 2017 IEEE Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA*, pages 11016–11025, 2019.

[41] G. Tolias, T. Jenícek, and O. Chum. Learning and aggregating deep local descriptors for instance-level recognition. In *Computer Vision —ECCV 2020— 16th European Conf., Glasgow, UK, August 23–28, 2020, Procs. Part I, Lecture Notes in Computer Science*, volume 12346, pages 460–477. Springer, 2020.

[42] G. Tolias, R. Sicre, and H. Jégou. Particular object retrieval with integral max-pooling of CNN activations. In *Proc. 4th Int. Conf. on Learning Representations, San Juan, Puerto Rico*, 12 pages, 2016.

[43] P. Weinzaepfel, T. Lucas, D. Larlus, and Y. Kalantidis. Learning super-features for image retrieval. In *Proc. 10th Int. Conf. on Learning Representations*, 19 pages, 2022.

[44] F. Yang, R. Hinami, Y. Matsui, S. Ly, and S. Satoh. Efficient image retrieval via decoupling diffusion into online and offline processing. In *Proc. 33rd AAAI Conf. on Artificial Intelligence, Honolulu, HI, USA*, pages 9087–9094, 2019.

[45] T. Yang, D. Nguyen, H. Heijnen, and V. Balntas. UR2KiD: Unifying Retrieval, Keypoint Detection, and Keypoint Description without local correspondence supervision. *Computing Research Repository arXiv Preprint*, arXiv:2001.07252, 2020.

[46] S. Zhang, M. Yang, T. Cour, K. Yu, and D. N. Metaxas. Query specific rank fusion for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 37(4):803–815, 2015.

[47] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf. Ranking on data manifolds. *Advances in Neural Information Processing Systems*, 16:169–176, 2003.