

Neighbor Embedding Projection and Rank-based Manifold Learning for Image Retrieval

Vinicius Atsushi Sato Kawai, Gustavo Rosseto Leticio, Lucas Pascotti Valem, Daniel Carlos Guimarães Pedronette
Department of Statistics, Applied Mathematics and Computing
State University of São Paulo (UNESP), Rio Claro, Brazil
vinicius.kawai@unesp.br, gustavo.leticio@unesp.br, lucas.valem@unesp.br, daniel.pedronette@unesp.br

Abstract—Despite the impressive advances in image understanding approaches, defining similarity among images remains a challenging task, crucial for many applications such as classification and retrieval. Mainly supported by Convolution Neural Networks (CNNs) and Transformer-based models, image representation techniques are the main reason for the advances. On the other hand, comparisons are mostly computed based on traditional pairwise measures, such as the Euclidean distance, while contextual similarity approaches can lead to effective results in defining similarity between points in high-dimensional spaces. This paper introduces a novel approach to contextual similarity by combining two techniques: neighbor embedding projection methods and rank-based manifold learning. High-dimensional features are projected in a 2D space used for efficiently ranking computation. Subsequently, manifold learning methods are exploited for a re-ranking step. An experimental evaluation conducted on different datasets and visual features indicates that the proposed approach leads to significant gains in comparison to the original feature representations and the neighbor embedding method in isolation.

I. INTRODUCTION

Manifold learning is a pivotal research area in computer science and data analysis, focusing on uncovering the underlying geometric structure of high-dimensional data. This technique is fundamental in reducing dimensionality while preserving essential data characteristics, enabling more efficient and accurate data processing and analysis. Manifold learning methods have found extensive applications in fields such as image processing and computer vision where high-dimensional data is prevalent [1]–[3].

The main idea of manifold learning is the assumption that high-dimensional data points lie on or near a lower-dimensional manifold. Different methods can achieve significant dimensionality reduction representations by identifying and mapping this manifold, thus facilitating tasks like visualization, clustering, and retrieval. Traditional methods, such as Principal Component Analysis (PCA) [4] and Multi-Dimensional Scaling (MDS) [5], provide linear approaches to dimensionality reduction. However, real-world data often exhibit non-linear structures that require more sophisticated techniques.

In this scenario, unlike linear dimensionality reduction methods, manifold learning techniques can provide a good approach for non-linear problems by preserving the local structure of the input data in the low-dimensional space. Several manifold learning methods have been proposed to address these challenges, each with unique strengths and applications.

One of the most well-known non-linear manifold learning methods is Locally Linear Embedding (LLE) [6]. LLE works by preserving local relationships among data points, assuming

that each data point and its neighbors lie on or near a locally linear patch of the manifold. Another notable method is Isomap (Isometric Mapping) [7], which extends Multi-Dimensional Scaling (MDS) [5] by incorporating geodesic distances instead of Euclidean distances. By approximating the manifold’s intrinsic geometry through these geodesic distances, Isomap can effectively handle non-linear structures in high-dimensional data, providing a more accurate low-dimensional representation.

Recently, Uniform Manifold Approximation and Projection (UMAP) [8] has gained attention for its scalability and efficiency. UMAP constructs a high-dimensional graph representation of the data and then optimizes a low-dimensional graph to be as structurally similar as possible. This approach balances the preservation of local and global data structure, making UMAP a versatile tool for dimensionality reduction and visualization.

Due to UMAP’s remarkable ability to preserve both local and global structures while enhancing the quality of the representations in a lower-dimensional space, this dimensionality reduction method has been widely applied across various fields of knowledge. Some examples include biological applications, such as single-cell RNA sequencing [9], understanding microbial diversity [10], and genomic studies [11], [12]. Similarly, other dimensionality reduction methods have also been utilized effectively in these areas. Beyond biology, UMAP and other techniques have been employed as preprocessing steps in computational methods. For instance, they have significantly improved clustering algorithms [13], [14] and Content Based Image Retrieval tasks [15], [16].

Given these advancements, recent literature has introduced novel approaches to image retrieval that leverage dimensionality reduction techniques [15], specifically t-SNE [17] and UMAP [8]. These strategies utilize the spatial relationships defined by neighbor embedding methods to compute more effective distance/similarity measures between images. Experiments demonstrate significant gains in retrieval effectiveness compared to original feature representations and competitive results against state-of-the-art methods. This highlights the potential of visualization techniques in improving image retrieval by better exploiting the manifold structure of image datasets.

From another perspective, manifold analysis also has been exploited by several different approaches in order to improve the effectiveness of image retrieval tasks. Rank-based manifold learning techniques focus on encoding the more global structure of ranked lists to generate a new, improved ranking result. These methods leverage similarity information encoded

within a set of ranked lists to better capture the underlying structure of the data. For a ranked list of images, evaluating the top positions in the ranking, i.e., the most similar images, is equivalent to assessing a subset of the entire ranking. This approach aligns with the concept of manifolds, as it focuses on the local neighborhood structure within the ranked list.

By analyzing and refining these top-ranked subsets, rank-based manifold learning methods can effectively improve the overall ranking quality, enhancing tasks such as image retrieval. In this way, recent studies have modeled visualization tasks as retrieval tasks [18], [19], thereby employing similarity learning strategies to enhance the effectiveness of visualizations. By utilizing graph-based approaches [20]–[22], diffusion processes [23], affinity learning [24], representation learning [25], clustering-based re-ranking [26], contextual measures [27] and scalable manifold ranking [28], post-processing methods based on manifold analysis have demonstrated significant improvements in the final retrieval results.

This work proposes a novel approach that combines recent rank-based manifold learning methods with neighbor embedding projection techniques, specifically UMAP, to achieve improved image retrieval results. The embeddings provided by UMAP offer a better representation of features in the reduced-dimensional space, enabling an efficient construction of rankings that encode similarity information. Subsequently, rank-based manifold learning methods perform re-ranking tasks, further enhancing the final effectiveness of image retrieval. By integrating these techniques, our approach seeks to leverage the strengths of both methodologies, resulting in a more effective retrieval system. Experiments considering CNN and Transformer-based features indicated that the proposed approach can improve the effectiveness of image retrieval tasks compared with both the original feature and the UMAP results in isolation.

The remainder of the paper is organized as follows: Section II presents a detailed description of the proposed approach. Section III discusses the experimental evaluation and Section IV presents the conclusions.

II. PROPOSED APPROACH: COMBINED MANIFOLD ANALYSIS BY NEIGHBOR EMBEDDING AND RANKING

A. Overview

The proposed approach integrates rank-based manifold learning methods with a recent dimensionality reduction technique to enhance content-based image retrieval tasks. The strategy involves several key steps, as illustrated in Figure 1 and detailed below:

- **Feature Extraction [A-C]:** High-dimensional features are extracted from images using deep learning models such as CNNs [29] and Transformers [30], [31]. This step provides a robust feature representation of the images.
- **Dimensionality Reduction [D-E]:** Uniform Manifold Approximation and Projection (UMAP) is applied to the high-dimensional features to project them into a lower-dimensional space. UMAP preserves the local and global structures of the data, resulting in a compact and meaningful representation of the image features.
- **Initial Ranking [F]:** In the reduced-dimensional space, initial similarity rankings are generated based on the Euclidean distance between image features via a tree-based indexing

method (Balltree). This step establishes a baseline ranking of images according to their similarity.

- **Rank-based Manifold Analysis and Re-ranking [G]:** The rank-based manifold methods are applied to the initial rankings to improve the quality and accuracy of the retrieval results. By analyzing and refining the initial ranking, these methods ensure that the most relevant images are prioritized in the re-ranking step.

- **Final Retrieval Result [H]:** The refined rankings are used to retrieve the most similar images to a given query, achieving enhanced retrieval performance by leveraging the strengths of both dimensionality reduction and rank-based manifold learning techniques.

B. Formal Definitions

• Image Retrieval Task

The image retrieval task aims to identify and return images from a collection \mathcal{C} that are most similar to a given query image x_q . Typically, retrieval tasks are based on features extracted from the images, which represent their content in a high-dimensional space. In this way, a feature descriptor can be formally defined as a function $f: \mathcal{C} \rightarrow \mathbb{R}^d$ that computes a d -dimensional vector for an input image, such that $\mathbf{x}_i = f(x_i)$ and $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{id}]$. Here, $x_{ij} \in \mathbb{R}$ denotes the j -th feature for the image x_i .

In this manner, the collection represented as $\mathcal{C} = \{x_1, x_2, \dots, x_n\}$, consists of images where each image x is defined by a d -dimensional feature vector \mathbf{x} . Based on these features, $\mathcal{X} \subset \mathbb{R}^d$ denotes a set of n points in a d -dimensional Euclidean space \mathbb{R}^d , such that $\mathbf{x}_i \in \mathcal{X}$.

The similarity between two images is defined based on the distance between their respective feature vectors. The distance function can be defined as $\rho: \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^+$, typically given by the Euclidean distance. Thus, the distance between two images x_i and x_j can be calculated as $\rho(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{x} \in \mathbb{R}^d$ is the feature representation of image x .

The nearest neighbor search problem aims to find the element $\mathcal{N}(\mathbf{x}_q)$ in the set \mathcal{X} with the smallest distance ρ to the query \mathbf{x}_q . This definition can be extended to consider a set of k nearest neighbors $\mathcal{N}(\mathbf{x}_q, k)$, which includes the k closest elements to the query. In this way, the ranked list τ_q can be defined as a permutation of (x_1, x_2, \dots, x_n) from $\mathcal{N}(\mathbf{x}_q, k)$, where $\tau_q(x_i)$ is interpreted as the position (or rank) of image x_i in the ranked list τ_q . If x_i is ranked before x_j in the ranked list of x_q , i.e., $\tau_q(x_i) < \tau_q(x_j)$, then $\rho(\mathbf{x}_q, \mathbf{x}_i) \leq \rho(\mathbf{x}_q, \mathbf{x}_j)$.

• Rank-based Manifold Learning Methods

Rank-based manifold learning methods aim to utilize similarity information encoded in a set of ranked lists $\mathbf{T} = [\tau_1, \tau_2, \dots, \tau_n]$ to better capture the underlying data structure. The goal is to compute a new, more effective set of ranked lists \mathbf{T}_m in an unsupervised manner, thereby enhancing the effectiveness of the retrieval results. Formally, these methods can be described as a function f_m :

$$\mathbf{T}_m = f_m(\mathbf{T}) \quad (1)$$

By generating a new set of ranked lists, usually these methods are referred to as re-ranking methods.

• Neighbor Embedding Projection

Pairwise comparisons between images, such as those using Euclidean distance on original high-dimensional features,

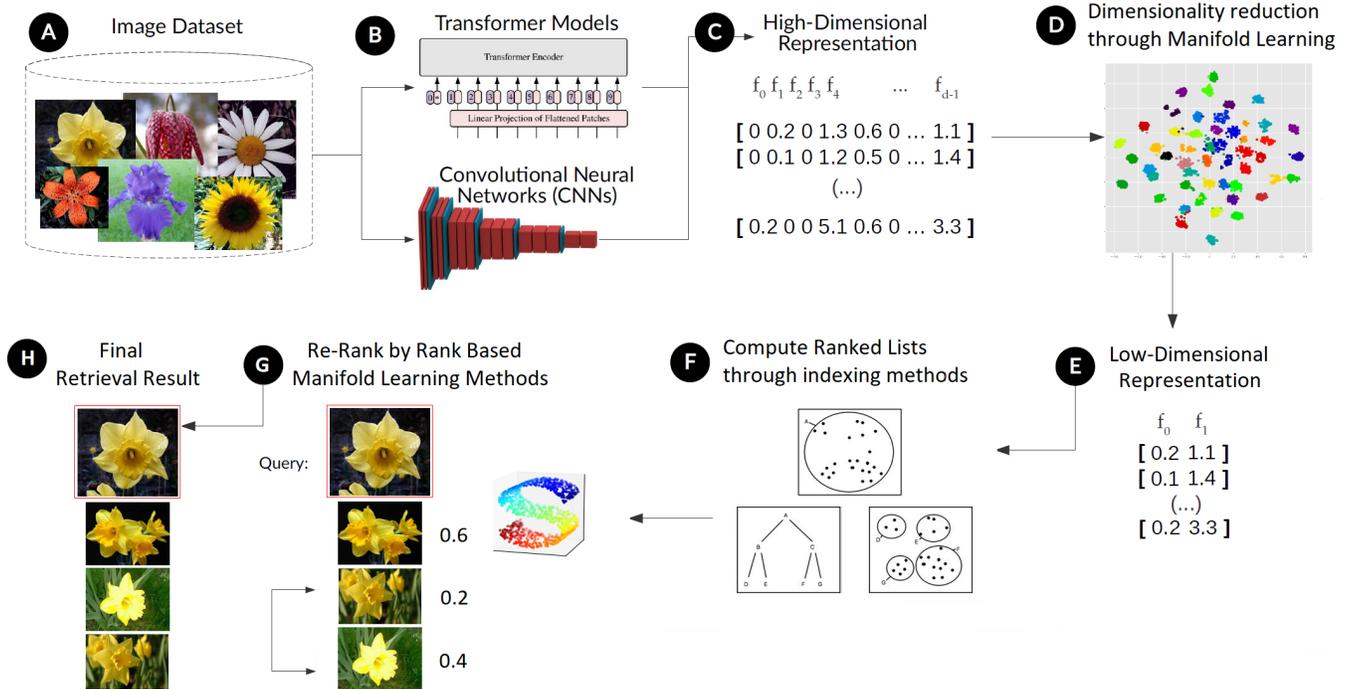


Fig. 1. Proposed Approach: Combined Manifold by Ranking and Projection for Content Based Image Retrieval Tasks

often fail to capture the global similarity information inherent in the dataset’s manifold structure. Additionally, high-dimensional spaces pose significant challenges related to efficiency, such as increased processing time and storage requirements, as well as effectiveness, including the quality of results.

In this work, we employ a projection technique based on the neighbor embedding framework to leverage the spatial relationships defined in low-dimensional spaces. A projection technique or algorithm can be formally defined as a function $P : \mathbb{R}^d \rightarrow \mathbb{R}^q$, where $q \ll d$, typically with $q = 2$. This projection function transforms a set \mathcal{X} in the high-dimensional space to a new set \mathcal{X}' in the low-dimensional space, such that $\mathcal{X}' = P(\mathcal{X})$.

C. Manifold Analysis by Neighbor Embedding Projection

Uniform Manifold Approximation and Projection (UMAP) [8] is a powerful dimensionality reduction technique that effectively balances the preservation of both local and global structures within data. By constructing a high-dimensional k-nearest neighbor (k-NN) graph and optimizing it to create a lower-dimensional representation, UMAP ensures that the intrinsic geometry of the data is maintained. In this manner, this method transforms complex high-dimensional data into a more manageable form while retaining essential characteristics needed for accurate analysis.

UMAP’s process involves the creation of a fuzzy simplicial set to represent the data, followed by an optimization step that minimizes the cross-entropy between the high-dimensional and low-dimensional fuzzy sets. The fuzzy simplicial set introduces a degree of uncertainty to the connections between data points, allowing for a more flexible representation of data relationships. This results in compact and meaningful embeddings that capture the underlying structure of the data,

making UMAP a useful tool for various applications in data analysis, pre-processing and visualization.

However, one of the primary limitations of UMAP is its sensitivity to hyperparameter settings, such as the number of neighbors ($n_neighbors$) and the minimum distance (min_dist). If these hyperparameters are not carefully tuned, UMAP may produce embeddings that do not accurately capture the intrinsic geometry of the data, leading to suboptimal retrieval performance. Additionally, UMAP can struggle with datasets that exhibit significant variability in local density, which can result in distortions in the low-dimensional representation and affect the quality of the rankings generated.

D. Rank-based Manifold Analysis

Considering recent advances in rank-based manifold learning approaches, pyUDLF (Python framework for unsupervised distance learning methods) [32] was proposed in order to provide easy access to different unsupervised distance learning approaches. In this way, we evaluate the proposed combination between UMAP and three different ranking methods: LHRR, CPRR, and RFE, described as follows:

- **Log-based Hypergraph of Ranking References (LHRR)** [22]: An approach that uses hypergraphs structures for data representation to explore similarity relationships between elements;
- **Cartesian Product of Ranking References (CPRR)** [33]: Uses the Cartesian product of rankings and neighborhood sets to calculate a new similarity measure, considering the pairs of relationships weighted by the ranking information;
- **Rank Flow Embedding (RFE)** [34]: This method combines Cartesian Product strategies and hypergraph structures to build connected components capable of generating new embeddings as well as performing re-ranking tasks.

III. EXPERIMENTAL EVALUATION

A. Datasets and Features

The experimental evaluation utilized several well-known public datasets, which are outlined below:

- **Flowers** [35]: This dataset consists of 1,360 images across 17 different flower species, with each species represented by 80 images. The dataset is provided by the University of Oxford.
- **Corel5k** [36]: Featuring a diverse range of scenes, this dataset includes images of fireworks, vehicles, microscopy images, tiles, trees, and more. It comprises 50 categories, each containing 100 images.
- **Dogs** [37]: This dataset contains a total of 20,580 images, showcasing 120 different breeds of dogs from around the world. Unlike the Flowers and Corel5k datasets, which are balanced in terms of the number of images per class, the Dogs dataset has slight variations in the number of images across different classes.

For the experimental evaluation, various deep learning-based features were employed. Results were obtained using features from three different models: one Convolutional Neural Network (CNN) model (ResNet152 [29]) and two Transformer-based models (Swin-TF [30] and Dinov2 [31]). All features were pre-trained on the ImageNet [38] dataset using transfer learning techniques.

B. Experimental Protocol

The experimental protocol follows an unsupervised setting, where all images are taken as queries for Flowers, Corel5k, and Dogs datasets. The Euclidean distance was used in all experiments. The Ball Tree [39] is used as the indexing method for all datasets, using the size of ranked lists as 1K.

The effectiveness measures considered are the Recall, Precision (at different depths defined according to the dataset) and the Mean Average Precision (MAP). For each experiment, the effectiveness measures are reported for the original features, UMAP, UMAP + RFE, UMAP + CPRR and UMAP + LHRR.

Regarding the UMAP parameters, $n_components = 2$ and $n_neighbors = 15$ were used. For the RFE, CPRR, and LHRR methods, the default parameters of the pyUDLF framework (from which their implementations were used) were employed, except for the parameter K , which varied according to the dataset. Specifically, $K = 80$ was used for the three methods on the Flowers dataset, $K = 100$ for the Corel5k dataset, and $K = 120$ for the Dogs dataset.

C. Retrieval Results

In this section, we present the effectiveness results of our proposed approach evaluated on three different datasets. Tables I, II and III provides Precision, Recall and MAP values for Flowers, Corel5k and Dogs datasets respectively.

The results indicate that the baseline effectiveness was significantly improved by the application of UMAP. Further enhancements were observed when UMAP was combined with rank-based methods (RFE, CPRR, and LHRR). For example, on the Flowers dataset, the combination of UMAP with RFE resulted in notable improvements in MAP, with values increasing from 92.91% (baseline SwinTF) to 99.53%. Similarly, for the SwinTF Corel5k dataset, combining UMAP with LHRR achieved a MAP of 92.69%, up from the baseline

MAP of 73.92%. On the Dogs dataset, the combination of UMAP with LHRR improved the MAP from 55.18% (baseline DinoV2) to 69.58%.

Additionally, Figure 2 showcases visual examples of the ranking results for the same query image using different methods. These visual comparisons highlight the qualitative improvements in retrieval quality achieved by our approach, demonstrating more accurate and relevant image retrieval results.

The observed differences in effectiveness across datasets can be attributed to their inherent characteristics. For instance, the Flowers dataset, with its relatively lower complexity and more homogeneous classes, benefited greatly from the refined rankings, achieving near-perfect precision in the top-ranked results. In contrast, the Dogs dataset presented a more challenging scenario. With its imbalanced class distribution, higher complexity, and greater class diversity, less precise manifold information is available to the projection approach. Nevertheless, the proposed approach still managed to outperform the baseline significantly.

These findings underscore the effectiveness of integrating UMAP with rank-based manifold learning methods, particularly in contexts where capturing complex data structures and improving retrieval accuracy are critical. The consistent improvements across different datasets validate the robustness of the proposed method, highlighting its potential for broader applications in image retrieval.

TABLE I
RETRIEVAL RESULTS FOR OXFORD 17 FLOWERS DATASET.

Method	Measure	Resnet152	SwinTF	DinoV2
None - Baseline	P@10	81.47 %	98.73 %	99.96 %
	P@20	74.41 %	97.87 %	99.82 %
	P@30	68.88 %	97.16 %	99.64 %
	P@50	60.28 %	96.21 %	99.30 %
	P@80	49.19 %	90.59 %	97.27 %
	Recall@40	32.18 %	48.32 %	49.75 %
	MAP	51.62 %	92.91 %	98.77 %
UMAP	P@10	84.38 %	99.49 %	100.00 %
	P@20	81.85 %	99.40 %	99.64 %
	P@30	80.20 %	99.36 %	99.49 %
	P@50	77.38 %	99.28 %	99.37 %
	P@80	69.45 %	98.88 %	98.60 %
	Recall@40	39.45 %	49.67 %	49.71 %
	MAP	73.39 %	99.25 %	98.84 %
UMAP + CPRR	P@10	84.28 %	99.57 %	99.37 %
	P@20	82.50 %	99.54 %	99.43 %
	P@30	81.28 %	99.49 %	99.46 %
	P@50	78.29 %	99.35 %	99.23 %
	P@80	70.47 %	99.04 %	98.26 %
	Recall@40	39.95 %	49.69 %	49.70 %
	MAP	74.42 %	99.43 %	99.00 %
UMAP + LHRR	P@10	85.33 %	99.62 %	99.46 %
	P@20	83.04 %	99.57 %	99.49 %
	P@30	81.87 %	99.42 %	99.47 %
	P@50	78.90 %	99.36 %	99.29 %
	P@80	71.07 %	99.06 %	98.31 %
	Recall@40	40.21 %	49.70 %	49.70 %
	MAP	75.83 %	99.42 %	98.87 %
UMAP + RFE	P@10	84.29 %	99.65 %	100.00 %
	P@20	82.92 %	99.64 %	99.64 %
	P@30	81.73 %	99.55 %	99.49 %
	P@50	79.00 %	99.51 %	99.37 %
	P@80	71.78 %	99.14 %	98.60 %
	Recall@40	40.32 %	49.78 %	49.71 %
	MAP	76.47 %	99.53 %	98.84 %

IV. CONCLUSION

In this work, we have proposed a novel approach that integrates UMAP, a powerful dimensionality reduction tech-



Fig. 2. Visual Ranking from same query image by different approaches: (i) the original feature, given by CNN Resnet; (ii) rank-based manifold learning by RFE [34]; (iii) UMAP in isolation and; (iv) proposed combined approach considering UMAP+RFE.

TABLE II
RETRIEVAL RESULTS FOR COREL5K DATASET.

Method	Measure	Resnet152	SwinTF	DinoV2
None - Baseline	P@10	89.97 %	96.15 %	95.81 %
	P@20	85.75 %	93.68 %	93.66 %
	P@30	82.56 %	91.46 %	92.01 %
	P@50	76.66 %	87.22 %	88.91 %
	P@100	61.29 %	71.27 %	77.84 %
	Recall@50	38.33 %	43.61 %	44.46 %
	Recall@100	61.29 %	71.27 %	77.84 %
	MAP	64.50 %	73.92 %	81.27 %
	UMAP	P@10	92.21 %	97.46 %
P@20		91.27 %	97.08 %	95.56 %
P@30		90.70 %	96.86 %	95.04 %
P@50		89.59 %	95.86 %	93.57 %
P@100		84.54 %	91.59 %	86.21 %
Recall@50		44.80 %	47.93 %	46.79 %
Recall@100		84.54 %	91.59 %	86.21 %
MAP		86.84 %	92.43 %	88.22 %
UMAP + CPRR		P@10	92.30 %	96.94 %
	P@20	91.53 %	96.88 %	95.43 %
	P@30	91.00 %	96.85 %	95.06 %
	P@50	89.91 %	96.13 %	93.72 %
	P@100	85.65 %	91.96 %	86.47 %
	Recall@50	44.95 %	48.06 %	46.86 %
	Recall@100	85.65 %	91.96 %	86.47 %
	MAP	87.89 %	92.65 %	88.62 %
	UMAP + LHRR	P@10	92.42 %	97.30 %
P@20		91.71 %	97.01 %	95.57 %
P@30		91.14 %	96.92 %	95.20 %
P@50		89.98 %	96.08 %	93.74 %
P@100		85.76 %	91.94 %	86.47 %
Recall@50		44.99 %	48.04 %	46.87 %
Recall@100		85.76 %	91.94 %	86.47 %
MAP		87.97 %	92.69 %	88.58 %
UMAP + RFE		P@10	92.46 %	97.60 %
	P@20	91.76 %	97.25 %	95.83 %
	P@30	91.27 %	97.08 %	95.30 %
	P@50	90.40 %	96.18 %	93.92 %
	P@100	85.80 %	92.18 %	87.54 %
	Recall@50	45.20 %	48.09 %	46.96 %
	Recall@100	85.80 %	92.18 %	87.54 %
	MAP	88.09 %	92.92 %	89.09 %

TABLE III
RETRIEVAL RESULTS FOR DOGS DATASET.

Method	Measure	Resnet152	SwinTF	DinoV2
None - Baseline	P@10	86.05 %	79.39 %	80.96 %
	P@20	83.53 %	75.63 %	76.98 %
	P@30	81.81 %	73.14 %	74.48 %
	P@50	79.13 %	69.04 %	70.72 %
	P@100	72.74 %	59.83 %	63.57 %
	Recall@50	23.23 %	20.29 %	20.74 %
	Recall@100	42.51 %	34.92 %	37.09 %
	MAP	63.73 %	45.53 %	55.18 %
	UMAP	P@10	86.84 %	79.01 %
P@20		85.93 %	77.55 %	77.35 %
P@30		85.47 %	76.86 %	76.16 %
P@50		84.92 %	75.99 %	74.54 %
P@100		83.79 %	74.24 %	71.74 %
Recall@50		25.03 %	22.45 %	21.94 %
Recall@100		49.35 %	43.81 %	42.25 %
MAP		80.54 %	69.70 %	66.54 %
UMAP + CPRR		P@10	87.04 %	79.07 %
	P@20	86.05 %	77.65 %	76.99 %
	P@30	85.55 %	76.95 %	75.94 %
	P@50	84.76 %	75.85 %	74.19 %
	P@100	82.72 %	73.09 %	70.72 %
	Recall@50	24.99 %	22.41 %	21.83 %
	Recall@100	48.75 %	43.13 %	41.66 %
	MAP	79.87 %	68.75 %	65.80 %
	UMAP + LHRR	P@10	87.24 %	79.25 %
P@20		86.16 %	77.79 %	77.76 %
P@30		85.58 %	77.01 %	76.52 %
P@50		84.77 %	75.86 %	74.58 %
P@100		82.95 %	73.36 %	71.27 %
Recall@50		25.00 %	22.41 %	21.94 %
Recall@100		48.89 %	43.29 %	41.99 %
MAP		80.10 %	68.97 %	69.58 %
UMAP + RFE		P@10	87.09 %	79.07 %
	P@20	86.11 %	77.82 %	76.58 %
	P@30	85.60 %	77.18 %	75.61 %
	P@50	84.84 %	76.18 %	74.10 %
	P@100	83.22 %	73.82 %	70.91 %
	Recall@50	25.01 %	22.50 %	21.78 %
	Recall@100	49.05 %	43.58 %	41.72 %
	MAP	80.20 %	69.43 %	66.20 %

nique, with rank-based manifold learning methods to enhance content-based image retrieval tasks. The experimental results demonstrate significant improvements in retrieval performance across multiple datasets, with notable gains in precision, recall, and mean average precision. By combining UMAP's low-dimensional features with rank-based manifold learning methods like RFE, CPRR, and LHRR, the proposed approach has proven to be effective in providing more accurate and

relevant retrieval outcomes.

These findings highlight the potential of leveraging the strengths of both dimensionality reduction and rank-based learning to advance the state of the art in image retrieval. Future work may explore further refinements and extensions of this approach to other types of data and more complex retrieval tasks. Furthermore, the proposed approach could be evaluated for its effectiveness by utilizing different indexing

and projection methods, as the dimensionality of the features needed for indexing data is significantly lower than that of the original feature space. Additionally, while re-ranking methods introduce a small efficiency cost, they are inherently fast, as they operate on small subsets within the data, adding minimal additional processing time.

ACKNOWLEDGMENT

The authors are grateful to São Paulo Research Foundation - FAPESP (grant #2018/15597-6), Brazilian National Council for Scientific and Technological Development - CNPq (grants #313193/2023-1, and #422667/2021-8), and Petrobras (grant #2023/00095-3) for financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES).

REFERENCES

- [1] Y. Han, Z. Xu, Z. Ma, and Z. Huang, "Image classification with manifold learning for out-of-sample data," *Signal Processing*, vol. 93, no. 8, pp. 2169–2177, 2013.
- [2] Y. Ma and Y. Fu, *Manifold learning theory and applications*. CRC press Boca Raton, 2012, vol. 434.
- [3] J. Zhang, S. Z. Li, and J. Wang, "Manifold learning and applications in recognition," *Intelligent multimedia processing with soft computing*, pp. 281–300, 2005.
- [4] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.
- [5] A. Mead, "Review of the development of multidimensional scaling methods," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 41, no. 1, pp. 27–39, 1992.
- [6] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000. [Online]. Available: <http://www.sciencemag.org/cgi/content/abstract/290/5500/2323>
- [7] J. B. Tenenbaum, V. d. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [8] L. McInnes, J. Healy, and J. Melville, "Umap: Uniform manifold approximation and projection for dimension reduction," *arXiv preprint arXiv:1802.03426*, 2018.
- [9] E. Becht, L. McInnes, J. Healy *et al.*, "Dimensionality reduction for visualizing single-cell data using umap," *Nature Biotechnology*, vol. 37, pp. 38–44, 2019. [Online]. Available: <https://doi.org/10.1038/nbt.4314>
- [10] M. W. Dorrity, L. M. Saunders, C. Queitsch, S. Fields, and C. Trapnell, "Dimensionality reduction by umap to visualize physical and genetic interactions," *Nature communications*, vol. 11, no. 1, p. 1537, 2020.
- [11] F. Trozzi, X. Wang, and P. Tao, "Umap as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: a comparison study," *The Journal of Physical Chemistry B*, vol. 125, no. 19, pp. 5022–5034, 2021.
- [12] T. A. Ujas, V. Obregon-Perko, and A. M. Stowe, "A guide on analyzing flow cytometry data using clustering methods and nonlinear dimensionality reduction (t-SNE or UMAP)," in *Neural Repair: Methods and Protocols*. Springer, 2023, pp. 231–249.
- [13] M. Sánchez-Rico and J. M. Alvarado, "Dimensionality reduction techniques as a preliminary step to cluster analysis: A comparison between PCA, t-SNE and UMAP," in *9th European Congress of Methodology*, 2020.
- [14] M. Allaoui, M. L. Kherfi, and A. Cheriet, "Considerably improving clustering algorithms using umap dimensionality reduction technique: a comparative study," in *International conference on image and signal processing*. Springer, 2020, pp. 317–325.
- [15] G. R. Leticio, V. S. Kawai, L. P. Valem, D. C. G. Pedronette, and R. d. S. Torres, "Manifold information through neighbor embedding projection for image retrieval," *Pattern Recognition Letters*, vol. 183, pp. 17–25, 2024.
- [16] P. Wu, B. Manjunath, and H. Shin, "Dimensionality reduction for image retrieval," in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101)*, vol. 3, 2000, pp. 726–729 vol.3.
- [17] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [18] J. Venna, J. Peltonen, K. Nybo, H. Aidos, and S. Kaski, "Information retrieval perspective to nonlinear dimensionality reduction for data visualization," *Journal of Machine Learning Research*, vol. 11, no. 2, 2010.
- [19] K. Bunte, M. Järvisalo, J. Berg, P. Myllymäki, J. Peltonen, and S. Kaski, "Optimal neighborhood preserving visualization by maximum satisfiability," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28, no. 1, 2014.
- [20] D. C. G. Pedronette, F. M. F. Gonçalves, and I. R. Guilherme, "Unsupervised manifold learning through reciprocal knn graph and connected components for image retrieval tasks," *Pattern Recognition*, vol. 75, pp. 161–174, 2018.
- [21] X. Yang, X. Bai, L. J. Latecki, and Z. Tu, "Improving shape retrieval by learning graph transduction," in *Computer Vision—ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, Proceedings, Part IV 10*. Springer, 2008, pp. 788–801.
- [22] D. C. G. Pedronette, L. P. Valem, J. Almeida, and R. d. S. Torres, "Multimedia retrieval through unsupervised hypergraph-based manifold ranking," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5824–5838, 2019.
- [23] X. Yang, S. Koknar-Tezel, and L. J. Latecki, "Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 357–364.
- [24] X. Yang and L. J. Latecki, "Affinity learning on a tensor product graph with applications to shape and image retrieval," in *CVPR 2011*. IEEE, 2011, pp. 2369–2376.
- [25] L. B. de Almeida, V. H. Pereira-Ferrero, L. P. Valem, J. Almeida, and D. C. G. Pedronette, "Representation learning for image retrieval through 3d cnn and manifold ranking," in *2021 34th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2021, pp. 417–424.
- [26] G. Park, Y. Baek, and H.-K. Lee, "Re-ranking algorithm using post-retrieval clustering for content-based image retrieval," *Information Processing & Management*, vol. 41, no. 2, pp. 177–194, 2005. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306457303000712>
- [27] H. Jegou, C. Schmid, H. Harzallah, and J. Verbeek, "Accurate image search using the contextual dissimilarity measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 2–11, 2008.
- [28] D. Lin, V. J. Wei, and R. C.-W. Wong, "Effective and scalable manifold ranking-based image retrieval with output bound," *ACM Trans. Knowl. Discov. Data*, vol. 17, no. 5, apr 2023. [Online]. Available: <https://doi.org/10.1145/3565574>
- [29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, June 2016, pp. 770–778.
- [30] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," *ICCV*, 2021.
- [31] M. Oquab, T. Darcet, T. Moutakanni *et al.*, "Dinov2: Learning robust visual features without supervision," 2023, arXiv preprint arXiv:2304.07193.
- [32] G. Leticio, L. P. Valem, L. T. Lopes, and D. C. G. a. Pedronette, "pyudf: A python framework for unsupervised distance learning tasks," in *Proceedings of the 31st ACM International Conference on Multimedia*, ser. MM '23. New York, NY, USA: Association for Computing Machinery, 2023, p. 9680–9684. [Online]. Available: <https://doi.org/10.1145/3581783.3613466>
- [33] L. P. Valem, D. C. G. Pedronette, and J. Almeida, "Unsupervised similarity learning through cartesian product of ranking references," *Pattern Recognition Letters*, vol. 114, pp. 41–52, 2018.
- [34] L. P. Valem, D. C. G. Pedronette, and L. J. Latecki, "Rank flow embedding for unsupervised and semi-supervised manifold learning," *IEEE Transactions on Image Processing*, 2023.
- [35] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1447–1454.
- [36] G.-H. Liu and J.-Y. Yang, "Content-based image retrieval using color difference histogram," *Pattern Recognition*, vol. 46, no. 1, pp. 188 – 198, 2013.
- [37] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei, "Novel dataset for fine-grained image categorization," in *Workshop on Fine-Grained Visual Categorization, CVPR*, June 2011.
- [38] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.
- [39] A. M. Kibriya and E. Frank, "An empirical comparison of exact nearest neighbour algorithms," in *11th European Conference on Principles and Practice of Knowledge Discovery in Databases*, ser. ECMLPKDD'07, 2007, p. 140–151.