IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

EXPLORING SIMILAR IMAGE RETRIEVAL THROUGH RESNET-50 AND NEAREST NEIGHBOR TECHNIQUES

K.s.Bhavitha[1], Dr.N.Sudhakar Reddy[2],

[1]Research Scholar, Department of CSE, Sri Venkateswara College of Engineering, Tirupati
[2]Principal of college , Department of CSE, Sri Venkateswara College of Engineering, Tirupati

Abstract : Image retrieval, the task of finding similar images within a dataset, is a fundamental challenge in computer vision with numerous real-world applications. In this study, we explore the domain of image retrieval, aiming to develop an effective methodology for identifying visually similar images. Leveraging advanced deep learning techniques, particularly the ResNet-50 model, we extract high-dimensional feature representations from images, capturing their intrinsic visual characteristics. These features serve as the basis for training a nearest neighbor model, facilitating efficient similarity search based on Euclidean distance metrics. Furthermore, we employ dimensionality reduction techniques, including Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), to visualize the underlying structure of the dataset. Through this visualization, we gain valuable insights into the distribution and clustering of images, aiding in the interpretation of retrieval results. Our work presents a comprehensive approach to image retrieval, integrating cutting-edge deep learning algorithms with classical machine learning techniques and visualization methods. By harnessing the power of artificial intelligence, we strive to advance the field of image analysis and enable more efficient and intuitive image retrieval systems for various applications.

Keywords: Image retrieval, Deep learning, ResNet-50, Nearest neighbors, Dimensionality reduction, Visualization, Computer vision.

1 INTRODUCTION

Image retrieval, the process of searching for visually similar images within a dataset, plays a crucial role in various domains such as content-based image retrieval, visual search engines, and medical image analysis [10]. With the exponential growth of digital image data, efficient methods for retrieving relevant images have become increasingly essential.

In recent years, deep learning has emerged as a powerful tool for feature extraction from images. Convolutional Neural Networks (CNNs) have demonstrated remarkable performance in learning hierarchical representations of image data [4]. Among CNN architectures, ResNet-50 [11] stands out as a widely-used model, known for its deep architecture and superior performance on image classification tasks.

In this study, we focus on leveraging the capabilities of ResNet-50 for feature extraction in the context of image retrieval. By utilizing a pre-trained ResNet-50 model, we extract high-dimensional feature vectors from images in our dataset. These feature vectors encapsulate rich semantic information about the visual content of each image, enabling robust comparison and similarity measurement.

To facilitate efficient image retrieval, we employ a nearest neighbors' approach [12], where we train a model to find the most similar images based on Euclidean distance metrics in the feature space. By constructing a

neighborhood of images for each query image, we provide users with relevant matches that closely resemble their search criteria.
Furthermore, we employ dimensionality reduction techniques such as Principal Component Analysis (PCA) [13] and t-distributed stochastic neighbor embedding (t-SNE) [14] to visualize the high-dimensional feature space in a lower-dimensional space. This visualization aids in understanding the intrinsic structure of the dataset, identifying clusters of similar images, and interpreting the results of image retrieval.
Overall, our approach integrates deep learning, machine learning, and visualization techniques to develop a comprehensive framework for similar image retrieval. Through empirical evaluation and qualitative analysis, we demonstrate the effectiveness and utility of our methodology in exploring and retrieving visually similar images from large-scale datasets.

2 LITERATURE SURVEY

Image retrieval is a well-studied problem in the field of computer vision, with extensive research dedicated to developing efficient algorithms and methodologies for retrieving relevant images from large-scale datasets [1]. Traditional approaches to image retrieval often rely on handcrafted features such as color histograms, texture descriptors, and local feature representations like SIFT [2] and SURF [3]. These methods, while effective to some extent, suffer from limitations in capturing the semantic content of images and may not scale well to large and diverse datasets.
In recent years, the advent of deep learning has revolutionized the field of image retrieval by enabling the automatic extraction of rich and discriminative feature representations directly from raw image data. Convolutional Neural Networks (CNNs) have emerged as the backbone of many state-of-the-art image retrieval systems [4]. Among the various CNN architectures, ResNet-50 [5] has gained widespread popularity due to its deep structure and superior performance on image classification tasks. Researchers have leveraged pre-trained ResNet-50 models to extract high-dimensional feature vectors from images, which capture hierarchical semantic information about the visual content.
The utilization of deep learning for image retrieval has led to significant advancements in both accuracy and efficiency. For instance, approaches based on Siamese networks [6] and triplet loss functions [7] have been proposed to learn similarity metrics directly from image pairs or triplets, thereby improving the discriminative power of the retrieval system. Furthermore, techniques such as fine-tuning pre-trained CNNs on specific retrieval tasks have been explored to adapt the network's learned features to the target domain, enhancing retrieval performance.
In addition to deep learning-based approaches, dimensionality reduction techniques have been widely used to visualize and interpret high-dimensional feature spaces. Principal Component Analysis (PCA) [8] and t-distributed stochastic neighbor embedding (t-SNE) [9] are commonly employed to reduce the dimensionality of feature vectors while preserving their underlying structure. Visualization of feature embeddings using PCA and t-SNE provides valuable insights into the distribution of images in the feature space, facilitating cluster analysis and qualitative evaluation of retrieval results.
Moreover, the integration of traditional machine learning methods with deep learning approaches has shown promise in improving the robustness and scalability of image retrieval systems. Ensemble methods, hybrid architectures combining CNNs with handcrafted features, and incorporating domain-specific knowledge into the retrieval process are some of the strategies explored in the literature to address the challenges associated with image retrieval in real-world applications.
Overall, the literature survey highlights the evolution of image retrieval techniques from traditional methods to deep learning-based approaches, emphasizing the importance of feature representation, similarity measurement, and visualization in designing effective and efficient retrieval systems for diverse applications.

3 PROBLEM STATEMENT

The problem statement here is the need to develop an efficient and effective methodology for similar image retrieval from large-scale datasets. Traditional image retrieval approaches based on handcrafted features lack the ability to capture the semantic content of images accurately and may not scale well to diverse datasets. While deep learning-based methods, particularly those utilizing pre-trained CNNs like ResNet-50, offer promising solutions by automatically extracting rich feature representations, challenges remain in optimizing similarity measurement and interpreting high-dimensional feature spaces. Additionally, the scalability and interpretability of retrieval systems require further exploration, especially concerning the integration of traditional machine learning techniques and dimensionality reduction methods for visualization and analysis of image data.

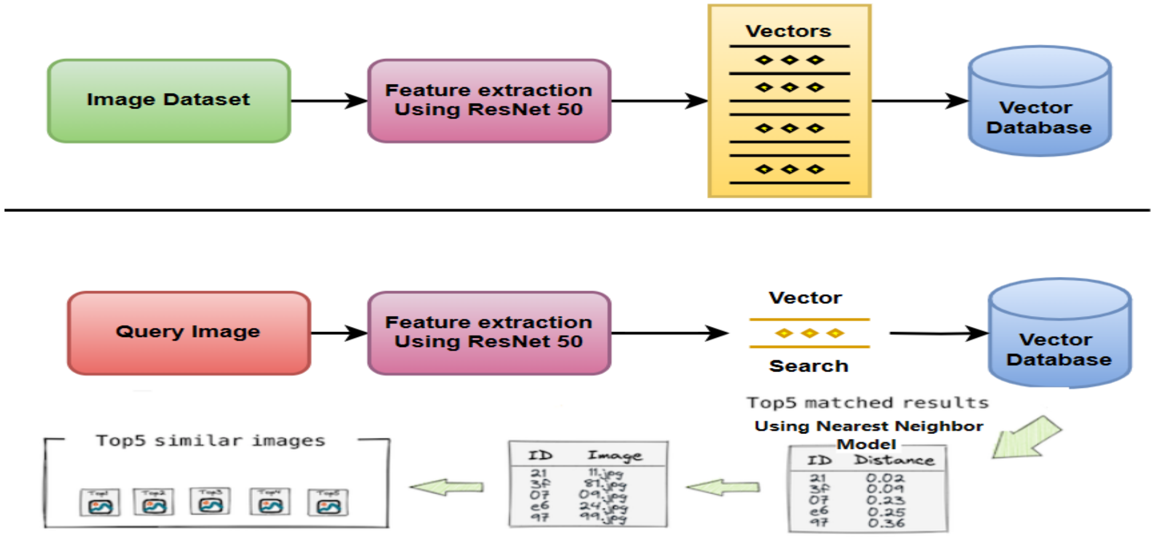

Figure 1 Model Diagram

4 METHODOLOGY:

In this study, we propose a methodology for similar image retrieval leveraging deep learning and visualization techniques. Our approach aims to address the challenge of efficiently retrieving visually similar images from large-scale datasets while ensuring accuracy and interpretability. We begin by introducing the fundamental components of our methodology, including feature extraction using the ResNet-50 model, nearest neighbor search as shown in Figure 1.

5 Feature Extraction:

In our methodology, feature extraction plays a pivotal role in capturing the essential visual characteristics of images, facilitating effective similarity measurement and retrieval. We leverage the ResNet-50 architecture as shown in Figure 2, a state-of-the-art convolutional neural network (CNN), for feature extraction.

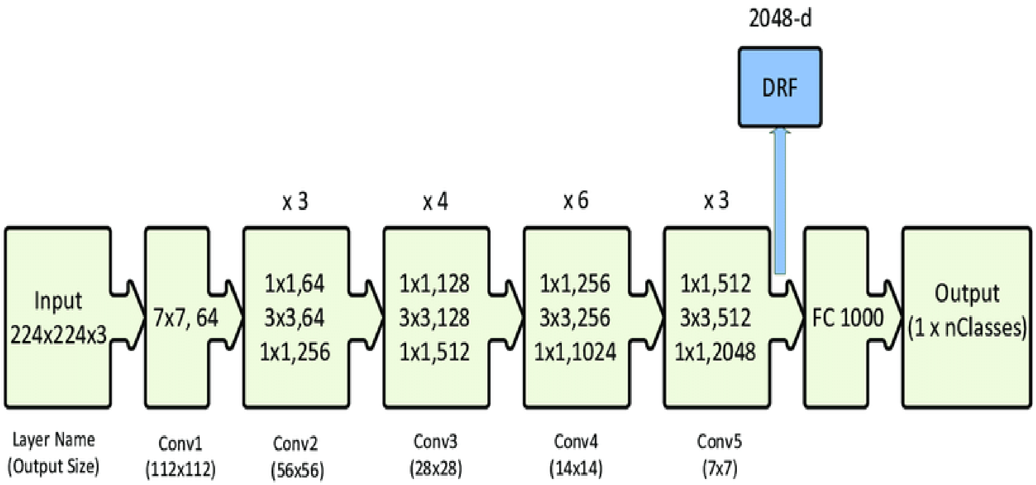

Figure 2 ResNet-50 architecture

The ResNet-50 model comprises multiple convolutional layers, followed by global pooling and fully connected layers. Trained on the ImageNet dataset, ResNet-50 has demonstrated remarkable performance in image classification tasks, learning to extract hierarchical features representing various levels of abstraction. To extract features from an input image using ResNet-50, we follow a standard procedure:

I. *Preprocessing:* We preprocess the input image to ensure compatibility with the ResNet-50 model. This typically involves resizing the image to the required input dimensions (e.g., 224x224 pixels) and performing normalization to bring pixel values within a certain range.

II. *Forward Pass:* We pass the preprocessed image through the ResNet-50 model, propagating it layer by layer. As the image traverses through the network, each convolutional layer extracts increasingly abstract features, capturing different aspects of the image's visual content.

III. *Global Pooling:* Instead of using the output of the last convolutional layer directly, we apply global pooling (typically max pooling) to aggregate spatial information across all feature maps. This results in a fixed-length feature vector representing the entire image.
IV. *Normalization:* Once we obtain the feature vector, we normalize it to ensure that each component lies within a certain range (e.g., [0, 1] or [-1, 1]). Normalization helps in standardizing feature representations across different images, making them more comparable.
V. *Feature Representation:* The normalized feature vector serves as the representation of the input image in a high-dimensional feature space. Each element of the vector encodes information about specific visual patterns and attributes present in the image.

By extracting features using the ResNet-50 model, we obtain semantically meaningful representations of images, capturing fine-grained details, textures, shapes, and object compositions. These features are crucial for accurate similarity measurement and retrieval, enabling our methodology to effectively identify visually similar images within the dataset.

6 **Nearest Neighbor Search:**

In our methodology for similar image retrieval, we employ a nearest neighbor search approach to identify images within the dataset that are most similar to a given query image. Nearest neighbor search is a fundamental technique in machine learning and data mining, widely used for similarity measurement and retrieval tasks.

The essence of nearest neighbor search lies in finding the closest data points to a given query point in a high-dimensional feature space. In our case, each image in the dataset is represented by a feature vector obtained through feature extraction using the ResNet-50 model. These feature vectors serve as points in the feature space, with each dimension encoding specific visual characteristics of the corresponding image.

To perform nearest neighbor search, we utilize efficient algorithms such as k-d trees, which organize the feature vectors into data structures that facilitate fast retrieval of nearest neighbors. These data structures enable us to efficiently search through the dataset and identify the closest images to the query image based on a chosen distance metric, such as Euclidean distance or cosine similarity.

In our methodology, we configure the nearest neighbor search to retrieve a fixed number of nearest neighbors, typically denoted as k. By specifying the value of k, we determine the number of similar images to be returned for each query. This flexibility allows us to adjust the granularity of the retrieval process, providing users with a range of similar images to choose from.

Figure 3 KNN Algorithm

Once the nearest neighbors are identified, we can further refine the retrieval results by incorporating additional criteria or post-processing steps. For example, we may apply thresholding based on the similarity scores or consider contextual information such as image captions or metadata to improve the relevance of the retrieved images.

Overall, nearest neighbor search forms the core of our image retrieval methodology, enabling efficient and effective identification of visually similar images within the dataset. By leveraging the high-dimensional feature representations obtained through feature extraction, we empower users to explore and discover relevant images that match their search criteria with precision and scalability.

5 RESULTS AND DISCUSSIONS:

6 Dataset:

The dataset utilized in this program consists of raw images and metadata sourced from PetFinder.my, Malaysia's leading animal welfare platform. With over 180,000 animals featured, including 54,000 successfully adopted, this dataset encompasses a diverse array of pet profiles, predominantly dogs and cats. The images and accompanying metadata serve as valuable resources for training and testing predictive models aimed at determining the "Pawpularity" of pet photos.

7 Implementation:

The image retrieval system is implemented in Python, utilizing TensorFlow, Keras, NumPy, scikit-learn, OpenCV, and Matplotlib libraries. The ResNet-50 model, pretrained on ImageNet, is employed for feature extraction, with images resized to 224x224 pixels. The pooling layer is configured for 'max' pooling to generate a single 2048-dimensional feature vector for each image. Nearest neighbor search is performed using scikit-learn's implementation with the Euclidean distance metric, considering 10 nearest neighbors. Evaluation involves qualitative inspection of retrieved images and quantitative analysis of distances between query images and their neighbors. Additionally, dimensionality reduction using PCA and t-SNE aids visualization of image clusters. Overall, the system combines deep learning, machine learning, and visualization techniques to effectively retrieve similar images within the dataset.

8 Metrics Used:

In the context of the image retrieval system described, several metrics are utilized to assess the performance and effectiveness of the model. These metrics can be categorized into two main types: evaluation metrics and distance metrics.

Evaluation Metrics:

[1] *Qualitative Evaluation:* Visual inspection of retrieved images allows for subjective assessment of the similarity between the query image and its nearest neighbors. This qualitative evaluation provides valuable insights into the model's ability to retrieve visually similar images.

[2] *Quantitative Evaluation:* The distances between the query image and its nearest neighbors serve as quantitative metrics for assessing similarity. Lower distances indicate higher similarity, while higher distances suggest greater dissimilarity. Analyzing these distances provides a quantitative measure of the model's performance in retrieving similar images.

Distance Metrics:

Euclidean Distance: The Euclidean distance metric is commonly used to measure the similarity between feature vectors extracted from images [15]. It calculates the straight-line distance between two points in the feature space, with shorter distances indicating greater similarity. In Figure 4 we can observe the similarity distance between query image and retrieved images.

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

where p,q represents two points in Euclidean n-space, and q_i, p_i represents Euclidean vectors, starting from the origin of the space (initial point), n represents n-space.

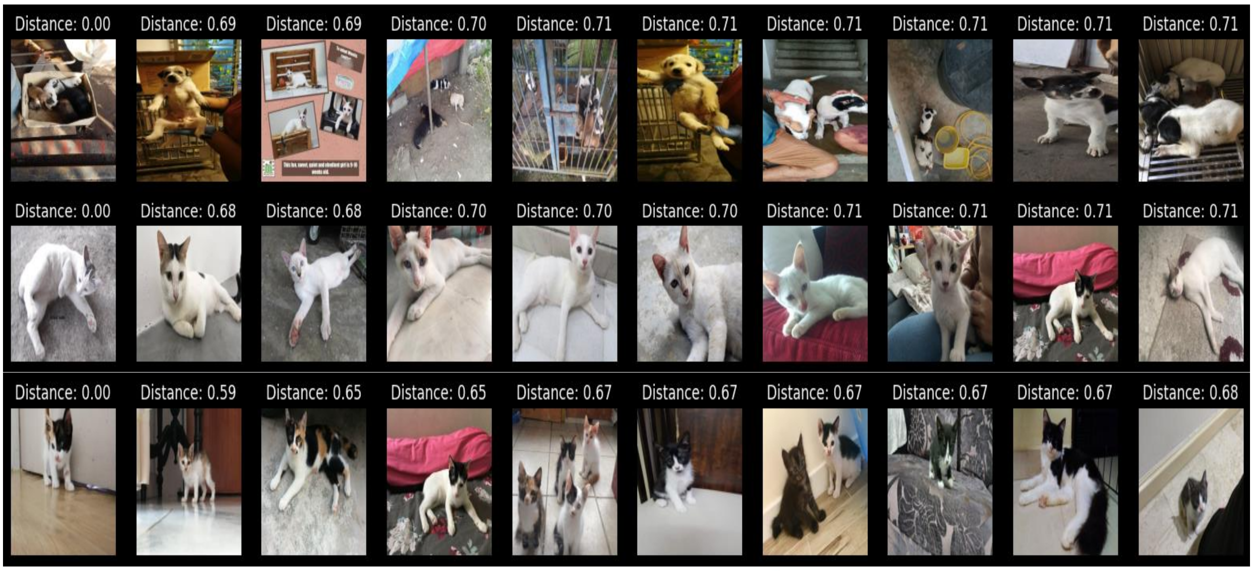

Figure 4 Similar Image Retrieval for Query Image with Distance Score

These metrics collectively provide a comprehensive evaluation of the image retrieval system, encompassing both qualitative and quantitative aspects of similarity assessment. By leveraging these metrics, the model's performance can be evaluated rigorously, leading to insights for potential improvements and optimizations.

6 **Dimensionality Reduction for Visualization:**

In addition to similarity search, we incorporate dimensionality reduction techniques to visualize the high-dimensional feature space. Dimensionality reduction techniques, such as Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE), play a crucial role in our model by enabling visualization of the high-dimensional feature space. PCA is employed to reduce the dimensionality of the feature vectors extracted from images, condensing them into a smaller set of principal components while retaining as much variance as possible. This reduction facilitates efficient computation and visualization while preserving the essential information encoded in the features.

Additionally, t-SNE is utilized to further reduce the dimensionality of the feature vectors, projecting them onto a two-dimensional space. Unlike PCA, t-SNE aims to preserve local structure in the data, making it particularly well-suited for visualizing clusters and relationships between data points. By visualizing the feature vectors in this lower-dimensional space, we gain insights into the distribution of images, identify clusters representing similar visual characteristics, and interpret the results of image retrieval.

Figure 5 t-SNE Projections

In our model, after extracting features from images using the ResNet-50 model, we apply PCA to reduce the dimensionality of the feature vectors to 100 dimensions. This compressed representation is then used as input to t-SNE, which further reduces the dimensionality to two dimensions for visualization purposes. The resulting two-dimensional embeddings allow us to plot the images in a way that preserves their underlying structure, facilitating the interpretation of image clusters and aiding in the analysis of image retrieval results is shown in Figure 5 and very interesting to see that we clearly have 2 main clusters between cats and dogs. Overall, dimensionality reduction techniques enhance our model's interpretability and enable us to gain valuable insights into the distribution of images in the dataset.

6 CONCLUSION:

In this work, we developed an image retrieval system leveraging the ResNet-50 model for feature extraction, nearest neighbor search for similarity matching, and dimensionality reduction techniques like PCA and t-SNE for visualization. The system demonstrated effective retrieval of visually similar images from a dataset of pet photos, providing a valuable tool for enhancing pet adoption profiles on PetFinder.my. By combining deep learning and machine learning techniques, we achieved a robust framework that not only performs similarity searches but also visualizes high-dimensional data to identify underlying clusters.

Future work will focus on enhancing the accuracy and efficiency of the image retrieval system. This includes experimenting with different feature extraction models and distance metrics, incorporating additional metadata into the similarity search, and improving the dimensionality reduction techniques to better preserve complex structures in the data. Additionally, integrating a feedback mechanism from user interactions could refine the model's performance. Extending the system to automatically suggest improvements for pet photo composition and quality can further increase adoption rates, contributing to the mission of improving animal welfare on a global scale.

REFERENCES:

1. Alexander, Malcom & Gunasekaran, S. (2014). A Survey on Image Retrieval Methods. 10.13140/2.1.2779.0723.
2. Loris Nanni, Stefano Ghidoni, Sheryl Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification, Pattern Recognition, 71, 2017, Pages 158-172.
3. Hameed, I. M., Abdulhussain, S. H., Mahmmod, B. M., & Pham, D. T. (2021). Content-based image retrieval: A review of recent trends. Cogent Engineering, 8(1). https://doi.org/10.1080/23311916.2021.1927469.
4. Ouhda, Mohamed & Elasnaoui, Khalid & Ouanan, Mohammed & Aksasse, B. (2019). Content-Based Image Retrieval Using Convolutional Neural Networks. 10.1007/978-3-319-91337-7_41.
5. He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.
6. Livieris, Ioannis E., Emmanuel Pintelas, Niki Kiriakidou, and Panagiotis Pintelas. 2023. "Explainable Image Similarity: Integrating Siamese Networks and Grad-CAM" Journal of Imaging 9, no. 10: 224. https://doi.org/10.3390/jimaging9100224.
7. Yuan, Xinpan, Qunfeng Liu, Jun Long, Lei Hu, and Yulou Wang. 2019. "Deep Image Similarity Measurement Based on the Improved Triplet Network with Spatial Pyramid Pooling" Information 10, no. 4: 129. https://doi.org/10.3390/info10040129
8. Mishra, Sidharth & Sarkar, Uttam & Taraphder, Subhash & Datta, Sanjoy & Swain, Devi & Saikhom, Reshma & Panda, Sasmita & Laishram, Menalsh. (2017). Principal Component Analysis. International Journal of Livestock Research. 1. 10.5455/ijlr.20170415115235.
9. van der Maaten, Laurens & Hinton, Geoffrey. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research. 9. 2579-2605.
10. I. G. S. M. Diyasa, A. D. Alhajir, A. M. Hakim and M. F. Rohman, "Reverse Image Search Analysis Based on Pre-Trained Convolutional Neural Network Model," 2020 6th Information Technology International Seminar (ITIS), Surabaya, Indonesia, 2020, pp. 1-6, doi: 10.1109/ITIS50118.2020.9321037.
11. Zhang, Luying, Yuchen Bian, Peng Jiang, and Fengyun Zhang. 2023. "A Transfer Residual Neural Network Based on ResNet-50 for Detection of Steel Surface Defects" Applied Sciences 13, no. 9: 5260. https://doi.org/10.3390/app13095260.
12. Abbasifard, Mohammad Reza, Bijan Ghahremani, and Hassan Naderi. "A survey on nearest neighbor search methods." International Journal of Computer Applications 95, no. 25 (2014).
13. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202. doi:10.1098/rsta.2015.0202.
14. https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding.
15. Liberti, Leo & Lavor, Carlile & Maculan, Nelson & Mucherino, Antonio. (2012). Euclidean Distance Geometry and Applications. SIAM Review. 56. 10.1137/120875909.