Iraqi Journal of Intelligent Computing and Informatics (IJICI)
Vol. 1, 1. 1, June 2022, pp. 34~40
ISSN: 2791-2868, DOI: 10.52940/ijici.v1i1.7

Determination of multi-modal dimension metric learning with application to web image retrieval

Khalid Hameed Zaboon[1]
[1]Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq

Article Info

Article history:

Received: Mar,10, 2022
Revised: Apr,14, 2022
Accepted: May,5, 2022

Keywords:

Multi-modal
Metric learning
Image Retrieval
Online Learning

ABSTRACT

Many real-world applications, like multimedia retrieval, confront the difficulty of determining the distance between any two items on Multi-modal data. According to most current Dimension Metrics Learning (DML) techniques, distance metrics may be learned using just one feature type or an aggregated feature space where many features are simply connected. Even though DML has been extensively researched. This study proposes a new framework for online learning, as well as a new classroom learning system for online Multi-modal dimension metric learning (OMDML), that is both efficient and scalable. This paper proposes a low-rank OMDML calculation to reduce the expensive cost of DML on high-dimensional component space, which reduces the computational cost while maintaining extremely competitive or much higher learning accuracy. With the purpose of determining whether or not multi-modular image recovery calculations can be successfully implemented, a large number of experiments are carried out. In most datasets tested, the suggested approach consistently outperforms alternative state-of-the-art algorithms, according to extensive experimental results.

Corresponding Author:

Khalid Hameed Zaboon
Department of Computer Science, Shatt Al-Arab University College, Basra, Iraq
Email: khalid.hameed@sa-uc.edu.iq

1. INTRODUCTION

The estimation of the distance or similarity between objects in many machine learning and pattern recognition algorithms requires a measure. Content-Based Information Retrieval (CBIR) systems, for example, use a similarity function to rank items based on their resemblance to a query picture [1][2]. The purpose of CBIR is to search photographs by assessing the image's actual content rather than information such as keywords, title, and author, thus a great deal of work has been devoted to the exploration of various low-level feature descriptors for picture representation [2].

To handle picture classification and object identification difficulties, a multi-class classifier is widely utilized. Finding the association between data characteristics and their classes is the first step in establishing a classification model from a database [3]. To begin, pair or triplet side information is used to learn a metric. The data is subsequently classified using a metric-based method. Metric learning may be especially beneficial in determining an invariant space for these adjustments since images are very vulnerable to changes in light and angle. For content-based multimedia retrieval tasks, an ideal measure or function is still being sought.

Dimension metric learning (DML) has recently been studied as a potential solution to this issue [4][5], this means using machine learning techniques to find the best way to measure distances using training data or other information, like logs of user relevance feedback from CBIR systems from the previous [6]. In the fields of machine learning and pattern recognition, metric learning is a hot topic. Through surveys, DML methodologies and applications are thoroughly examined [7].

One of the most difficult challenges to solve is content-based picture retrieval [8], [9]. For example, researchers have spent years examining global picture representation characteristics including color features, edge features, and texture features [10]. Local feature-based representation, such as bag-of-words models, has seen a rush of study in recent years [11], [12], and Using descriptors for local features [13]. The inflexible similarity/distance function may not always be ideal due to the complexity of visual picture representation and the semantic gap between low-level visual information obtained by computers and high-level human perception and interpretation.

As a consequence, there has been a surge in active research efforts in recent years to use machine learning methods to generate multiple distance/similarity measures on certain low-level variable [14], [15]. Some of the studies focus on learning to hash for compact codes [16]–[18]. Image classification and object identification have attracted a great deal of interest in recent years due to multimodal/multiview investigations [19], [20]. Nevertheless, applying these approaches directly to CBIR is usually challenging since image classes are not explicitly offered on CBIR tasks; even if they are, the number of classes would be enormous, as image datasets on CBIR tasks are far greater than on classification tasks.

Machine learning and multimedia retrieval fields have both done substantial research on distance metric learning. The primary idea is to discover a measurement that reduces the gap between similar/related pictures while increasing the distance between dissimilar/unrelated ones. Only a few recent DML research are exploring online learning techniques [21], [22], and they all employ batch learning approaches that need the whole training data collection to be provided before the learning assignment.

There have been a lot of DML techniques in the literature [23], [24], but most of them have a place in single-modular DML because they take a separation measure either on one type of highlight or on a joined component space by connecting different types of different elements together. In practice, such strategies may be limited by a few opportune impediments [25] . Some parts may fundamentally dominate the others in the DML job, hindering the capacity to harness the capabilities of all components. Also, an accidental connection approach could lead to a connected high-dimensional component space, which would make the DML process hard to calculate. To address these restrictions, this study examines a new method of Online Multi-modular Dimension Metric Learning (OMDML), which involves obtaining separation measures from multi-modular data or different sorts of components using a competent and extendable web-based learning strategy.

In contrast to the above connection technique, the suggested OMDML system has two main concepts: (i) It determines how to enhance each methodology's separation metric (i.e., each kind of highlight space), and (ii) It figures out how to achieve the optimal mix of different separation metrics across a variety of modalities. In addition, the recommended OMDML graphic combines the points of interest of web-based learning systems for high productivity and adaptability in large-scale learning assignments.

This study also presents a Low-rank Online Multimodal DML (LOMDML) computation, which avoids the requirement for improved positive semi-distinct (PSD) projections and so saves considerable time when running DML on high-dimensional data [26], [27]. Furthermore, a theoretical study of the OMDML approach is offered, and several tests were conducted to evaluate the performance of the suggested methodologies for CBIR tasks utilizing a variety of criteria.

2. METHOD

2.1 Overview of Online Multi-Modal Distance Metric Learning

In the literature, several ways to improve CBIR performance have been presented. Some previous research has looked at developing unique low-level feature descriptors to better define image visual information, the development or learning of effective distance/similarity measures based on extracted low-level characteristics has been the focus of several researchers. In reality, it is impossible to find a single best low-level feature representation that consistently outperforms the competition in all situations. Therefore, the use of machine learning algorithms to automatically include a wide range of attributes and their related distances is a highly sought-after metric. This open research subject is known as a multi-modal distance metric learning problem, and this section offers unique solutions to it.

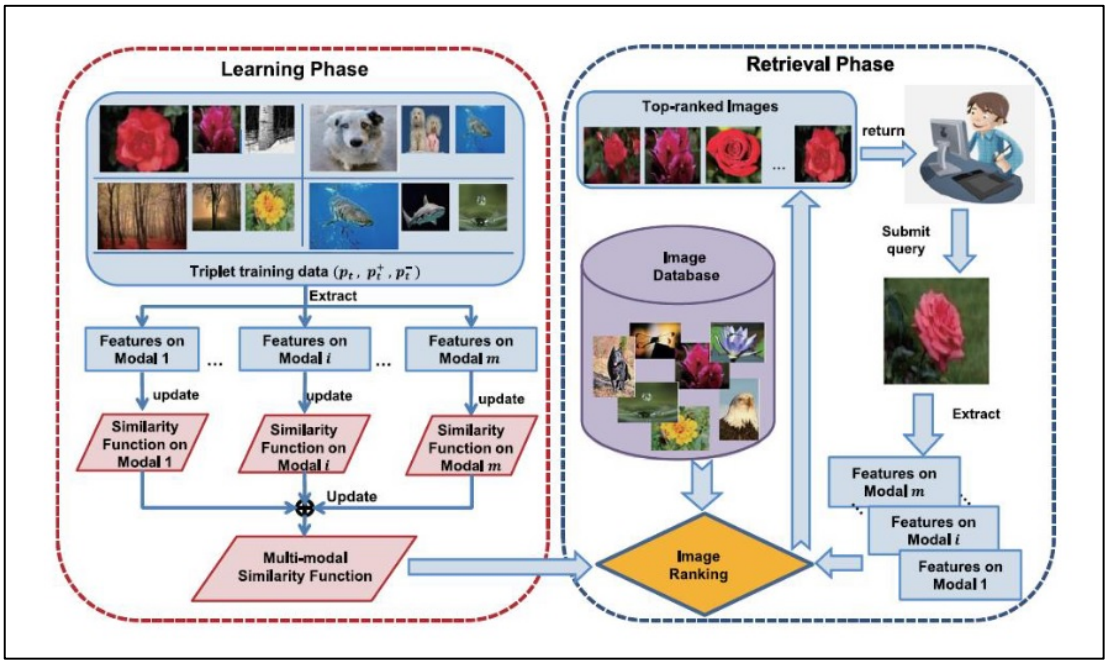

Figure 1. Shows the system flow of the proposed method

The suggested multi-modal distance metric learning technique for content-based picture retrieval is shown in Figure 1, which is divided into two phases: learning and retrieval. For the retrieval phase's image rating, the learning phase's purpose is to memorize the distance measurements. In fact, these two stages may continue indefinitely, with the learning phase learning from an endless supply of training data.

May assume that throughout the learning phase, triplet training data instances come in a sequential order, which is acceptable for a real-life CBIR system. For instance, in online relevance feedback, users are commonly requested to offer input on whether a returned picture is relevant or unrelated to a query; as a consequence, users' relevance feedback log data may be collected sequentially to provide training data for the learning task [28]. Once a triplet of photos is received, extract various low-level feature descriptors across many modalities. Extract multiple low-level feature descriptors from various modalities after receiving a triplet of pictures. These characteristics and labels may then be used to update the distance function on a single modality.

Extraction of many low-level feature descriptors across several modalities after receiving a triplet of photographs Using the associated features and label information, each distance function on a single modality may then be revised. Similar techniques are used by CBIR systems to extract low-level feature descriptors on several modalities, rank the pictures in their database using an optimum distance function, and lastly show a list of top-ranking images to the viewer.

2.2 Low-Rank Online Multi-modal Distance Metric Learning Algorithm

The objective is to leverage side information to discover a distance function for content-based picture retrieval; the discussion will be limited to the Mahalanobis distances family. For the purpose of computing the separation between any two images, the following distance function is used in order to generate an ideal distance metric M: One dimension of the feature space is Rn, which is the number of dimensions of p1, p2. Assume that a set of training data examples is given (one after the other) in the form of triplet constraints. Each triplet shows how three pictures are related.

The proposed OMDML Method has a significant issue in the PSD projection stage, which may be computationally costly when some feature space has many dimensions. Present a low-rank learning technique to dramatically increase OMDML's efficiency and scalability in this part. The PSD projection step of the

proposed OMDML Method is a critical problem, as it may be computationally challenging when feature spaces have large dimensions. The goal is to learn a low-rank decomposition rather than a full-rank matrix. There are methods of solving this problem that may be found online. The preferred Low-rank Online Multi-modal Metric Learning technique is described in depth in this approach (LOMDML). This method clearly retains the PSD property of the final distance measure and removes the requirement for a lengthy PSD projection.

3. RESULTS AND DISCUSSION

Examining the theoretical performance of the recommended Methods. To keep things simple, it is possible to demonstrate a bound of errors theorem for a technique for assessing the relative similarity of a series of triplet training examples. Where (x) is a true/false indicator function. In addition, a collection of training instances may determine the optimal margin similarity function error.

In general, combining the Hedge algorithm's findings with PA online learning, which is comparable to the approach used, is a straightforward way to prove the assumption. More detail on the evidence may be found in the companion file. When compared to the best single measure, the suggested algorithm's total number of mistakes is limited, according to the above theorem. The ssbsection conducts a series of experiments to evaluate the effectiveness of the offered techniques for visual feature similarity search.

3.1 Experimental Testbeds

In the research, four picture data sets that are available to the public have been used a lot as benchmarks for content-based image retrieval, image classification, and image recognition tasks. TABLE 1 outlines the statistics for various data sources.

TABLE 1- Image datasets used in our testbed

Dataset	size	Classes #	Avg # pre class
Caltech101	8,677	101	85.91
Indoor	15,620	67	233.14
Image CLEF	7,157	20	367.85
Corel	5,000	50	100
ImageCLEFFlickr	1,007,157	21	47959.86

The "Caltech101"1 is the first and most widely used testbed for picture retrieval and object recognition [29]. This collection includes 101 item categories and 8,677 pictures. The second testbed, "indoor" dataset 2, was utilized to detect interior scenes [30]. There are 15,620 pictures in this collection, divided into 67 interior categories. Each category comprises at least 100 photos, despite the fact that the number of photographs in each category varies. Retail, home, public areas, leisure, and the workplace are further subdivided. It is conceivable to treat it as a 67-category dataset and run several algorithms over the whole indoor collection.

The "Image CLEF" dataset3, which was utilized in [31], is the third testbed. This medical picture collection contains 7,157 photographs divided into 20 categories. The "Corel" dataset, which comprises of images from COREL CDs, is the fourth testbed. It contains 50 categories, each with precisely 100 images picked at random from COREL image CDs from related cases. To make "ImageCLEFFlickr," combine "Image CLEF" with a library of one million social pictures taken from Flickr. Which considers Flickr images as a subset of noisy backdrop shots that are primarily utilized to assess the scalability of the proposed method.

3.2 Experimental Setup

There may be three distinct subsets for each database: one for training, one for testing, and one for validation. The research randomly selected 500 pictures for the test set and another 500 for the validation set. The rest of the photos are used to make a training set for learning how to find similar functions. Can produce

side information in the form of triplet instances for learning ranking functions, as well as sample triplet strains based on their ground truth labels from images in the training set.

In order to create a triplet instance, randomly choose two images from one class and one picture from a separate class. Each typical dataset generates 100K triplets (except for the small-scale and large-scale experiments). To compare competing approaches equally, can apply the same cross validation methodology to choose their parameters. May experimentally set ri = r = 50 for the i-th modality in the LOMDML algorithm and the maximum iteration to 500 for LMNN. The mean Average Precision (mAP) and top-K retrieval precision are two metrics used to evaluate retrieval performance. The Average Precision (AP) value of all queries, which measures the area under a query's precision-recall curve, is averaged as a common IR metric. An example's precision is calculated by dividing the number of related instances found in a search by how many total related examples were found in the database; this is known as the recall value. A Linux PC with an Intel Xeon CPU operating at 2.33 GHz and 16GB of RAM was used for all of the tests.

OMDML and LOMDML should be compared against a number of existing representative DML algorithms such as RCA, LMNN, and OASIS in order to thoroughly evaluate the proposed method's efficacy. As a heuristic baseline metric, utilize the square Euclidean distance, abbreviated "EUCL-*." Using DML to each modality independently and then choosing the best one is what "best" implies. These algorithms are identified by the suffix "-B," such as RCA-B, in which learn metrics for each modality individually on the Relevance Component Analysis training set (RCA). A validation set is used to test the retrieval performance of all measures on each modality, and the modality with the greatest mAP is declared the best. RCA allows us to determine which scoring method performed best on the test set. Before using DML, "concatenation" mixes characteristics from all modalities.

Concatenating numerous features together, using LMNN to train an ideal metric on this combined feature space, and then assessing mAP scores on this best-fit metric are all examples of these "-C" techniques. "Uniform combination" is a late fusion approach that combines all modalities evenly. An "-U" suffix denotes that these methods are used to find the best metrics for each modality using OA-SIS, and then to aggregate all distance functions to get a final ranking.

A CBIR model's performance is evaluated using the MAP measure. The identical experimental set-up is used for both the CBIR and picture categorization tasks. CBIR's performance employing previously learned distance and similarity functions was examined on the tested dataset. Results reveal that multi-modal techniques outperform single-modal ones in CBIR applications.

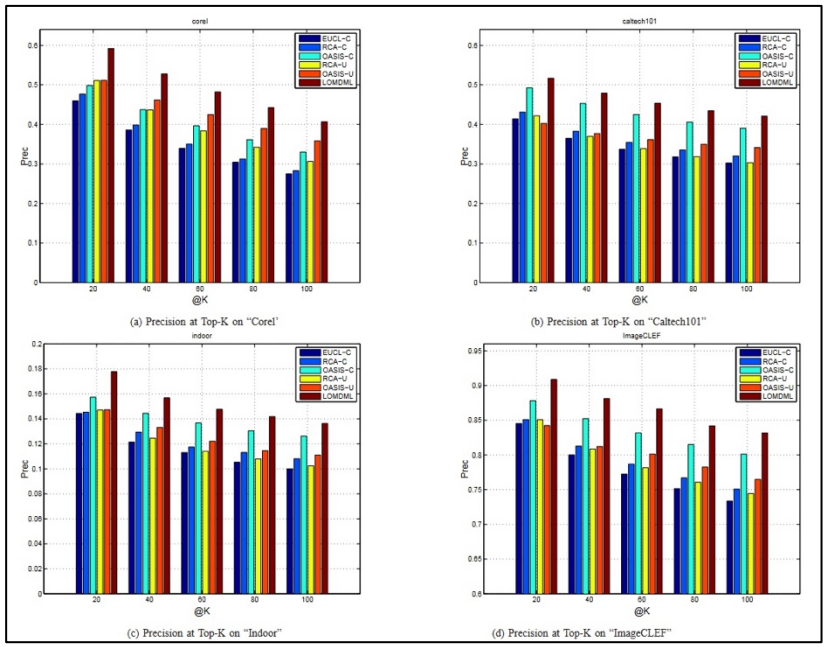

Figure 2. The evaluation of the convergence rates of the competing algorithms on the datasets analyzed.

In terms of MAP scores, LOMDML regularly beats OMDML in the majority of the datasets studied, comparable to the findings of image classification. The data support our method's capacity to aggregate numerous measure outcomes in multi-modal scenarios.

4 CONCLUSION

The LOMDML metric learning technique for multimodal data is introduced in this paper. In a multi-modal dataset, LOMDML generates a metric for each feature type. This method's adaptive factor modifies each metric's coefficient, resulting in a faster convergence rate. In addition, the suggested technique offers a method to online learning utilizing a multi-modal dataset, which minimizes the learning time while preserving the discriminatory capacity of the learned measure. Research shows that the suggested method consistently outperforms other current algorithms in most datasets investigated. LOMDML has been shown to be scalable in terms of both the dimensions and quantity of training data using online learning.

Future prospects for research comprise: (1) Examining various approaches to combine learnt metrics in multimodal data, (2) The application of the suggested approach to imbalanced datasets, and (3) Evaluating the effectiveness of our approaches in diverse applications, including re-identification and recommender systems.

REFERENCES

[1] J. Li, C. Xu, W. Yang, C. Sun, J. Xu, and H. Zhang, "Discriminative multi-view privileged information learning for image re-ranking," *IEEE Trans. Image Process.*, vol. 29, pp. 3490–3505, 2020.

[2] P. Wu, S. C. H. Hoi, P. Zhao, C. Miao, and Z.-Y. Liu, "Online multi-modal distance metric learning with application to image retrieval," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 454–467, 2015.

[3] H. N. K. Al-behadili, K. R. Ku-Mahamud, and R. Sagban, "Annealing strategy for an enhance rule pruning technique in ACO-based rule classification," *Indones. J. Electr. Eng. Comput. Sci*, vol. 16, no. 3, pp. 1499–1507, 2019.

[4] K. Roth, T. Milbich, S. Sinha, P. Gupta, B. Ommer, and J. P. Cohen, "Revisiting training strategies and generalization performance in deep metric learning," in *International Conference on Machine Learning*, 2020, pp. 8242–8252.

[5] G. Cheng, P. Zhou, and J. Han, "Duplex metric learning for image set classification," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 281–292, 2017.

[6] L. Zhang, H. P. H. Shum, and L. Shao, "Discriminative semantic subspace analysis for relevance feedback," *IEEE Trans. image Process.*, vol. 25, no. 3, pp. 1275–1287, 2016.

[7] Y. Luo, Y. Wen, L.-Y. Duan, and D. Tao, "Transfer metric learning: Algorithms, applications and outlooks," *arXiv Prepr. arXiv1810.03944*, 2018.

[8] J. Choe *et al.*, "Content-based Image Retrieval by Using Deep Learning for Interstitial Lung Disease Diagnosis with Chest CT," *Radiology*, vol. 302, no. 1, pp. 187–197, 2022.

[9] M. Garg and G. Dhiman, "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Comput. Appl.*, vol. 33, no. 4, pp. 1311–1328, 2021.

[10] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognit.*, vol. 62, pp. 135–160, 2017.

[11] L. Cheng, J. Li, Y. Silva, D. Hall, and H. Liu, "PI-bully: Personalized cyberbullying detection with peer influence," 2019.

[12] S. Dabiri and K. Heaslip, "Developing a Twitter-based traffic event detection model using deep learning architectures," *Expert Syst. Appl.*, vol. 118, pp. 425–439, 2019.

[13] M. Dusmanu *et al.*, "D2-net: A trainable cnn for joint detection and description of local features," *arXiv Prepr. arXiv1905.03561*, 2019.

[14] F. Ekpenyong, "A Stochastic Framework to Fuzzy Environments Through Case-based Reasoning and the Inverse Problem Methodology: An Investigation of Financial Bubbles." University of Brighton, 2021.

[15] N. Radwan, "Leveraging sparse and dense features for reliable state estimation in urban environments." University of Freiburg, Freiburg im Breisgau, Germany, 2019.

[16] J. Chen *et al.*, "Deep local binary coding for person re-identification by delving into the details," in

Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 3034–3043.
[17] Z. Lu, Y. Hu, Y. Jiang, Y. Chen, and B. Zeng, "Learning binary code for personalized fashion recommendation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10562–10570.
[18] Y. Wang, X. Ou, J. Liang, and Z. Sun, "Deep semantic reconstruction hashing for similarity retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 387–400, 2020.
[19] X. Qian *et al.*, "Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning," *Nat. Biomed. Eng.*, vol. 5, no. 6, pp. 522–532, 2021.
[20] I. Jegham, A. Ben Khalifa, I. Alouani, and M. A. Mahjoub, "A novel public dataset for multimodal multiview and multispectral driver distraction analysis: 3MDAD," *Signal Process. Image Commun.*, vol. 88, p. 115960, 2020.
[21] S. Mystakidis, E. Berki, and J.-P. Valtanen, "Deep and meaningful e-learning with social virtual reality environments in higher education: a systematic literature review," *Appl. Sci.*, vol. 11, no. 5, p. 2412, 2021.
[22] H. N. K. Al-Behadili, K. R. Ku-Mahamud, and R. Sagban, "Hybrid ant colony optimization and genetic algorithm for rule induction," *J. Comput. Sci.*, vol. 16, no. 7, pp. 1019–1028, 2020.
[23] P. Singh, P. N. Hrisheekesha, and V. K. Singh, "CBIR-CNN: content-based image retrieval on celebrity data using deep convolution neural network," *Recent Adv. Comput. Sci. Commun. (Formerly Recent Patents Comput. Sci.*, vol. 14, no. 1, pp. 257–272, 2021.
[24] S. Hu, X. Chen, W. Ni, E. Hossain, and X. Wang, "Distributed machine learning for wireless communication networks: Techniques, architectures, and applications," *IEEE Commun. Surv. Tutorials*, vol. 23, no. 3, pp. 1458–1493, 2021.
[25] N. Berkeley, D. Jarvis, and A. Jones, "Analysing the take up of battery electric vehicles: An investigation of barriers amongst drivers in the UK," *Transp. Res. Part D Transp. Environ.*, vol. 63, pp. 466–481, 2018.
[26] A. M. E. B. Ali, "Reinforcement Learning based Evolutionary Metric Filtering in Clustering," 2021.
[27] B. Ali, K. Moriyama, M. Numao, and K. Fukui, "Reinforcement Learning based Evolutionary Metric Filtering for High Dimensional Problems," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 2020, pp. 226–233.
[28] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 140–150, 2019.
[29] F. Radenović, G. Tolias, and O. Chum, "Fine-tuning CNN image retrieval with no human annotation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1655–1668, 2018.
[30] M. Afif, R. Ayachi, Y. Said, and M. Atri, "Deep learning based application for indoor scene recognition," *Neural Process. Lett.*, vol. 51, no. 3, pp. 2827–2837, 2020.
[31] C. Ma, J. Lu, and J. Zhou, "Rank-consistency deep hashing for scalable multi-label image search," *IEEE Trans. Multimed.*, vol. 23, pp. 3943–3956, 2020.