Proposal: Learning Hash for Image Retrieval

1st Ba Cong Nguyen
Faculty of Information Technology
University of Science, VNU-HCM
Ho Chi Minh City, Vietnam
nbcong22@clc.fitus.edu.vn

2nd Tran Anh Khoa Dang
Faculty of Information Technology
University of Science, VNU-HCM
Ho Chi Minh City, Vietnam
dtakhoa22@clc.fitus.edu.vn

***Abstract*—The rapid growth of image datasets has made traditional image retrieval methods inefficient in terms of accuracy and scalability. Hashing-based techniques address this by encoding high-dimensional image data into compact binary codes, enabling fast retrieval with minimal storage. However, Deep Supervised Hashing (DSH) methods often struggle with quantization loss and suboptimal similarity preservation.**

This study explores a two-stage retrieval approach, where an initial fast search using hashing is followed by a refinement stage to improve ranking precision. We compare various DSH methods under this framework, analyzing their trade-offs between retrieval speed, accuracy, and computational efficiency. By standardizing evaluation metrics and testing across multiple datasets, we aim to provide clear insights into optimizing hashing-based image retrieval.

***Index Terms*—hashing, image retrieval, deep learning**

I. Introduction

Efficient image retrieval relies on encoding high-dimensional visual features into lower-dimensional representations while preserving semantic relationships. Hashing techniques have been widely adopted due to their ability to enable fast comparisons using Hamming distance. However, challenges such as semantic loss during hashing and sensitivity to data distribution limit their real-world effectiveness.

A two-stage retrieval approach can mitigate these issues by leveraging hashing for an initial fast search, followed by a refinement stage that improves ranking precision. This study explores:

- The strengths and weaknesses of various DSH methods under this two-stage framework.
- The impact of different ranking refinement strategies on retrieval accuracy.
- A fair and standardized evaluation of these techniques across diverse datasets.

A. Problem Statement

With the exponential growth of image datasets, traditional image retrieval methods have become increasingly inefficient in terms of both accuracy and scalability. Conventional approaches struggle to handle the high-dimensional nature of image data, leading to slow retrieval speeds and excessive storage requirements. Hashing-based techniques offer a promising alternative by encoding images into compact binary codes, allowing for faster search operations with reduced memory consumption.

Despite the advancements in hashing-based retrieval, existing methods, particularly Deep Supervised Hashing, still face challenges. These include suboptimal hash code learning, difficulty in preserving semantic similarities, sensitivity to noise and variations in images, and computational overhead during training. Furthermore, many current models lack the ability to generalize well across diverse datasets, limiting their applicability in real-world scenarios.

Therefore, there is a need to explore and develop improved Deep Supervised Hashing techniques that enhance retrieval accuracy, efficiency, and robustness.

II. Related Work

Deep Supervised Hashing (DSH)

Approach

The approach begins by designing a CNN model that processes image pairs alongside labels indicating whether the two images are similar or not, and outputs binary codes. To increase efficiency, image pairs are generated dynamically during training, enabling the utilization of a significantly larger number of pairs. The loss function is crafted to bring the outputs of similar image pairs closer together while pushing the outputs of dissimilar pairs further apart. This ensures that the resulting Hamming space effectively reflects the semantic structure of the images. Since optimizing directly in Hamming space is nondifferentiable, the network outputs are relaxed to continuous values during training. At the same time, a regularization term is introduced to drive the continuous outputs toward discrete binary values. Using this framework, images are encoded by passing them through the network, followed by converting the network's outputs into binary code representations via quantization.

Image retrieval process within DSH

During the training phase, the DSH model learns to map images into binary hash codes due to the CNN. The network is trained with pairs of images labeled as similar or dissimilar, ensuring that semantically similar images are mapped to similar binary codes. And when a query image is provided, it is passed through the CNN to generate its binary hash code. That code shall then be compared with other binary hash codes within the dataset(s). Calculating Hamming distance will provide the number of differing bits between the codes.

After all the calculations, the image with hash codes that have the smallest Hamming distances to the query's hash code are retrieved as results.

Training Methodology

Traditional Deep Supervised Hashing (DSH) methods often utilize a Siamese network structure for generating image pairs offline. However, this approach is constrained by storage limitations, as a dataset of n images results in storing $n/2$ valid pairs. To improve scalability, modern methods generate image pairs dynamically within mini-batches, ensuring diverse pairs across iterations without requiring pre-stored similarity matrices. This enhances computational efficiency, making it viable for large-scale datasets. Additionally, reusing early network layers across different hash code lengths optimizes memory usage, but increasing code length adds model complexity, necessitating careful regularization to prevent overfitting. [3]

Feature Learning-Based Deep Supervised Hashing with Pairwise Labels

Training Methodology

DPSH employs pairwise similarity labels to learn hash codes that align closely for semantically related images. The training set is composed of a small labeled subset (seed set), a large unlabeled pool, and a query set whose labels are dynamically acquired. Active learning is employed iteratively: (1) training on the labeled set, (2) selecting the most informative samples from the unlabeled pool, and (3) acquiring labels from an oracle. Hash codes are generated by projecting images through the network, where binarization is performed using thresholding. To optimize retrieval efficiency, a loss function enforces compactness for similar pairs and divergence for dissimilar ones. Due to the inherent difficulty in direct binary optimization, the model relaxes constraints using continuous approximations while balancing objectives like pairwise similarity and quantization error. Semi-supervised learning strategies further enhance performance by incorporating both labeled and unlabeled data. [2]

Deep Supervised Discrete Hashing (DSDH or DDSH)

Core Contributions and Training Strategy

DSDH explicitly optimizes the final layer of the model to output discrete binary hash codes, ensuring consistency with pairwise labels and class information. This marks one of the first attempts to integrate classification and pairwise similarity learning into a unified hashing framework. Quantization errors, a common challenge in hashing, are mitigated through an alternating minimization approach, employing discrete cyclic coordinate descent for efficient optimization.

The binary hash learning process iteratively updates hash bits while ensuring similarity preservation. The k-th bit of each data point is optimized by minimizing an objective function that enforces binary constraints (-1 or $+1$ values). To address the non-differentiability of the sign function, activation functions like hyperbolic tangent provide smooth approximations, enabling gradient-based updates. The training alternates between optimizing hash codes and refining neural network parameters via backpropagation. Experiments demonstrate that DSDH surpasses prior approaches in retrieval efficiency, particularly in balancing discrete optimization with deep learning. [1]

However, existing DSH models still face three key challenges: (1) semantic preservation—many methods struggle to retain fine-grained relationships between images; (2) scalability—computational costs remain high for large-scale datasets; and (3) robustness—hashing methods are sensitive to variations in image quality and distribution.

III. Methodology

We aim to implement the two-stage retrieval approach, along with compare various deep learning based hashing models (or methods) in a multitude of aspects (such as training, runtime and general efficiency of resources) to hopefully find out the most effective model in general. The models will all be fed the same dataset and operated on the same device to ensure that the comparison is as impartial as possible, as different machines' processing power and datasets certainly will make or break some parts of the algorithms.

1) Dataset Selection: Use diverse datasets such as CIFAR-10, NUS-WIDE to test generalization
2) Model Implementation and Training: Train DSH models using PyTorch. Ensure consistent training settings across models.
3) Two-Stage Retrieval Process:
    - Step 1 - Fast Approximate Search via Hashing: Compute binary hash codes for query and database images
    - Step 2 - Refinement and Re-ranking: Apply Similarity re-ranking using deep features embedding or graph-based methods, then compare different refinement techniques with analysis.
4) Evaluation Metrics: Analyze retrieval performance under varying code lengths (e.g., 16-bit, 32-bit, 64-bit).

IV. Expected Outcomes

1) A clear comparative analysis of DSH methods under the two-stage retrieval framework.
2) Identification of the best-performing refinement techniques for improving accuracy.
3) Practical insights into balancing retrieval speed and ranking precision.
4) Comprehensive benchmarking results that position the proposed method as a leader in learning-based hashing for image retrieval.

Acknowledgment

References

[1] Qi Li, Zhenan Sun, Ran He, and Tieniu Tan. Deep supervised discrete hashing. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 2479–2488, Red Hook, NY, USA, 2017. Curran Associates Inc.

[2] Wu-Jun Li, Sheng Wang, and Wang-Cheng Kang. Feature learning based deep supervised hashing with pairwise labels, 2016.

[3] Haomiao Liu, Ruiping Wang, Shiguang Shan, and Xilin Chen. Deep supervised hashing for fast image retrieval. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.