♦ Recepción/ 27 junio 2019
♦ Aceptación/ 25 agosto 2019

Image Retrieval using Neural Network based Hash Encoding:A Survey

Recuperación de imágenes utilizando la codificación de hash basada en redes neuronales: una encuesta

Sameer I. Ali Al-Janabi
University of Anbar, Iraq
isl.samir.ia2012@uoanbar.edu.iq

Sufyan T. Faraj Al-Janabi
University of Anbar, Iraq
sufyan.aljanabi@uoanbar.edu.iq

Belal Al-Khateeb
University of Anbar, Iraq
belal@computer-college.org

ABSTRACT/ Analysis of image contents has become one of the important subjects in modern life. In order to recognize the images with efficient way, several techniques have appeared and periodically enhanced by the developers. Image retrieval becomes one of the main problems that face the computer society inside the revolution of technology. To increase the effectiveness of computing similarities between images, hashing approaches became the focusing of the programmers. Indeed, deep learning in the past few years has been considered the backbone of image analysis using a convolutional neural network (CNN). The paper is providing a survey of the latest work carried out in the field of image retrieval. Several techniques have appeared in this field. However, the most common of these techniques are using neural network-based hash encoding, which can be categorized into three main classes: Supervised, unsupervised, and semi-supervised techniques according to each technique's learning method. The most important related works appeared in the literature are reviewed and constructive comparisons have been done to show the strengths and limitations of various techniques. Keywords: Image Retrieval, Deep Learning, Convolutional Neural Network (CNN), Hashing Techniques. **RESUMEN** / El análisis del contenido de la imagen se ha convertido en uno de los temas importantes en la vida moderna. Para reconocer las imágenes de manera eficiente, han aparecido varias técnicas que los desarrolladores han mejorado periódicamente. La recuperación de imágenes se convierte en uno de los principales problemas que enfrenta la sociedad informática dentro de la revolución de la tecnología. Para aumentar la efectividad de las similitudes informáticas entre imágenes, los enfoques de hash se convirtieron en el foco de los programadores. De hecho, el aprendizaje profundo en los últimos años se ha considerado la columna vertebral del análisis de imágenes utilizando una red neuronal convolucional (CNN). El documento proporciona una encuesta sobre el último trabajo realizado en el campo de la recuperación de imágenes. Varias técnicas han aparecido en este campo. Sin embargo, la más común de estas técnicas es utilizar la codificación hash basada en redes neuronales, que se puede clasificar en tres clases principales: técnicas supervisadas, no supervisadas y semi-supervisadas de acuerdo con el método de aprendizaje de cada técnica. Se revisan los trabajos relacionados más importantes que aparecen en la literatura y se han realizado comparaciones constructivas para mostrar las fortalezas y limitaciones de varias técnicas.
Palabras clave: recuperación de imágenes, aprendizaje profundo, red neuronal convolucional (CNN), técnicas de hash.

1. Introduction

Computer vision has recently turned out to be essential technology because of its wide-going applications in various areas including diverse and smart monitoring, health and medicine, sports and entertainment, robotics, drones, and self-driving cars. Visual recognition errands, for example, image classification, localization, and detection are the center building blocks of a significant number of these applications. Ongoing improvements in convolutional neural network (CNN) have prompted the remarkable execution in these state-of-the-art visual recognition tasks and systems [1], [2].
On the other hand, the advances in Deep Learning (DL) over the most recent years have made special infiltration in a few zones,

especially in computer vision, in which machine insight has gone past human execution. The deep architecture joins the low-level features to abstract high-level characteristics with nonlinear transform. This results in the required power and ability to take in the semantic representation from images [3]. For example, deep reinforcement learning has been utilized in self-driving vehicles to solve in choices by utilizing criticism from numerous sorts of sensors around the vehicle [4].

Another strategy that is utilized in the image retrieval framework is a hash function, or compression function, which means the output is being shorter than the input. Hash functions are already utilized in many applications including numerous cryptographic tasks [5]. There are three fundamental types of the cryptographic hash functions that are used in the security field: preimage resistance, second preimage resistance, and collision resistance[6]. Another type of hash function is dynamic hash functions. This type contains two parameters instead of one; the extra parameter is called security parameter. This parameter is responsible for specifying the interior working and size of message digest produced [7].

This paper includes a survey of some common image retrieval architectures with a briefing about the method of working for each technique. There are several techniques used in the system of image retrieval. Those methods may be classified according to the learning method into three categories: Supervised, unsupervised, and semi-supervised techniques, as shown in Figure 1.

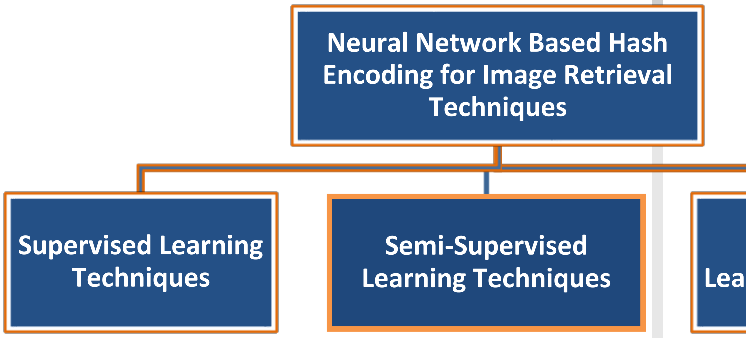

Figure 1. Main Classification of Neural Network Based Hash Encoding for Image Retrieval Techniques.

In general, there are some challenging limitations of most image retrieval techniques. These include the following [8]:

- Prediction of semantic features from primitive features in the image or video.
- Reaching to the relationship between semantic features and primitive features.
- Indexing the contents of multimedia by reducing the needing to the human interaction.
- Enabling the system to understand and interpret the user request semantically.

Furthermore, it is possible to recognize several advantages and disadvantages of neural network based hash encoding, as summarized in Table 1.

Table 1. Advantages and disadvantages of neural network based hash encoding.

Advantages	**Disadvantages**
Security in hash function generation [9]. Speed of hash function generation [9]. Extracting patterns and detecting behavior that are complex for noticing by humans or another computer technique [10]. The ability to emulate highly complex computational machines [10].	The neural network need for training to work. High processing time also needed for large size neural networks [11].

In the remaining of this paper, each of the supervised, unsupervised, and semi-supervised categories is reviewed in a subsequent section (Sections 2, 3, and 4), respectively. The emphasis is mainly on supervised techniques due to their importance and wide-spreading use in many applications. Finally, Section 5 contains the most important conclusions of this work.

2. Supervised Learning Techniques

The approaches of supervised traffic classification must have a pre-labelled data-set for training. A classifier is trained in the feature space via the training dataset and will be applied for classifying new network traffic. Numerous researches were performed for solving a variety of traffic classification issues with the use of supervised approaches[1]. In this section, we will describe the common architectures for supervised learning methods of image retrieval with examples for each architecture.

2.1 One-Stage Supervised Deep Hashing Framework (SDHP)

Qi Li *et al* [3] suggested a technique that includes developing a deep supervised discrete hashing approach which is modelled on the basis of the hypothesis that the learned binary codes must be optimal for classification. Each of the classification information and the pairwise label information are utilized for learning the hashing codes within one stream model. This was achieved by constraining the results of the past layer to be directly binary codes, and that has been seldom researched in the deep hashing approach. Because hash codes have discrete nature, an alternating minimizing approach is utilized for optimizing the objective function. This might be quite a useful approach for applications of image or video search because it is computationally inexpensive and storage efficient. Comparing this method to conventional hashing approaches utilizing deeply learned properties that have been obtained by the CNN-F network, interesting results could be obtained. The average MAP outputs of SDH+CNN and FastH+CNN on CIFAR10 data-set are 0.553 and 0.604, respectively. The average MAP output of this approach on CIFAR10 data-set is 0.787, and that performs better than previous conventional techniques of hashing with deeply learned features. Indeed, the suggested approach accomplishes a comparable efficiency to the optimal conventional approaches of hashing on NUS-WIDE data-set under the 1st experimental setting.

Another method proposed by Dongbao Yang *et al* [4] was called a one-stage supervised hash approach for learning high-quality binary codes. Those researchers executed a deep CNN and enforced the learned codes for satisfying the following points: (a) the binary codes must be distributed uniformly; (b) images that are similar must be encoded to similar binary codes, and the other way around; (c) quantization loss must be minimized. Experimental comparisons between this approach and previous algorithms had been conducted on NUS-WIDE and CIFAR-I0 data-sets. The MAP of this approach reached to 87.67% and 77.48% on CIFAR-I0 and NUS-WIDE datasets respectively with 48-bit. Thus, this method improved the search accuracy. The contribution of this work mainly focused on four aspects: At first, the GoogleNet architecture gives to the developers the ability to put their own equations in the classification layer because of its highly efficient feature representation power. Secondly, a pairwise loss function has been devised to maintain the semantic similarity of the original data. Thirdly, enforcing the binary codes uniformly spread for carrying more information. Fourth, the loss of quantizing from Euclidean to Hamming space has been diminished.

Also, Chenggang Yan *et al* [5] developed the previous architecture for learning high-quality binary codes, and in the same framework, the method was extended into what they called SDHP+ for improving the discriminative power of binary codes. Implementation of this approach used a deep CNN, and the learned codes were enforced for satisfying the following criteria: Similar binary codes must result from similar images, and vice versa, minimizing the loss of quantization from Euclidean to Hamming spaces, and the learned codes must be distributed in a uniform manner. The application of this approach was for the efficient recognition of the on-road environment based on the analysis of the contents of the images from the large-scale scene database. Experimental comparisons of this method with some previous hashing algorithms had been performed on NUS-WIDE and CIFAR-10. The MAP of SDHP reached 87.67% and 77.48% with 48 b (binary code), respectively, and the MAP of SDHP+ reaches 91.16%, 81.08% with 12 b, 48 b on CIFAR10 and NUS-WIDE, respectively. It can be concluded the fact that the latter approach described in [5]outperformed the others because the enhancement in this method called (SDPH+) had improved the search accuracy.

2.2 Nearest Neighbour Search

For the majority of the available hashing approaches, an image is initially encoded in a form of a vector of hand-engineering visual characteristics, which is followed by one more step of separate projection or quantizing which produces binary codes.

Hanjiang Lai *et al* [6] proposed a single-stage supervised hash approach for retrieving images that generated bitwise hash codes for images by a thoroughly designed deep model. The suggested deep architecture used a triplet ranking loss designed in order to keep relative similarities. In this approach, the input images were converted to unified representations of image through a shared subnetwork of stacked convolution layers. After that, those intermediate image representations were encoded to hash codes via "divide-and-encode" modules, as shown in Figure 2. Empirical evaluations in image retrieval showed the fact that the suggested approach had achieved better performance gains compared to some previous supervised or unsupervised hashing approaches. The method achieved better search accuracies (accuracy within Hamming distance 2, MAP, accuracy with varying size of top returned samples, and precision-recall) than the ones that are considered baseline approaches utilizing conventional hand-crafted visual features. For instance, in comparison with the optimal competitor KSH, the MAP results of the suggested approach refer to a relative increasing of 58.8%-90.6 %, 61.3%-82.2 %, and 21.2%-22.7% on SVHN, CIFAR10, and NUS-WIDE datasets, respectively.

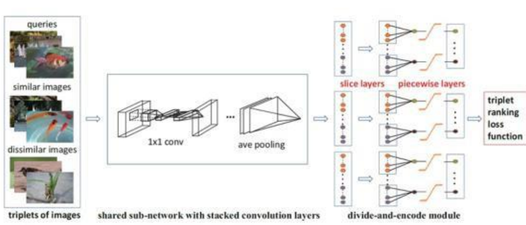

Figure 2. One-stage supervised hashing method architecture [6]

Manish Sapkota *et al* [7] designed a Deep Convolutional Hashing (DCH) approach which may be trained "point-wise" for a simultaneous learning of each of binary and semantic representations of histo-pathological images. Particularly, the researchers proposed a CNN that introduced a latent binary encoding (LBE) layer for low dimensional embedding of features for learning binary codes. This method included designing a joint optimizing objective function which encourages the network for learning discriminative representations from the label data, and diminish the gap between the desired binary values and the real-valued low dimensional embedded features. The binary encoding for new images could be obtained through forward propagation via the NN and the quantization of the result of the LBE layer. Results of experimentations on a large-scale histo-pathological image data-set demonstrated the efficiency of the suggested approach. This architecture provided a fast image query and retrieval of related cases, this way could be helpful for specialists in the evidence-based research of the diseases for diagnosing. In addition to that, they could research the retrieved similar instances for understanding the biological and morphological properties of an illness. A better ranking performance has been noticed for DCH with MAP that ranges between 0.94 and 0.96 utilizing various numbers of bits, and performing batter that the other considered approaches by 2%-4%. However, the results can be further enhanced by enhancing the loss of quantization with a more sufficient optimization formulation. In addition to that, this work had the assumption that image labels have no noise, in other words, data annotation has been consistent. Label noise would negatively influence the learning of the suggested model and after that in the final diagnose of the disease. This is why, the research can be enhanced by developing noise-insensitive learning algorithms for training the network in addition to the statistical evaluation of the robustness of the suggested approach on the diagnosis of the disease.

Approximate nearest neighbour search could be an effective approach for large-scale retrieval of images. Jun-Yi Li and Jian-hua Li [12]developed a sufficient deep learning framework approach for generating binary hash codes for time efficient image retrieval based on CNN. The adopted concept was that it is possible learning binary codes with the use of a hidden layer for presenting the latent concepts that dominate the class labels in which the data labels can be used. In addition to that, CNN may be utilized for learning image representations. Other supervised approaches need pair-wised binary code learning inputs. Nevertheless, this approach

may be utilized for learning image representations and hash codes in a point-by-point way, for this reason, it is appropriate for large-scale data-sets. Experimental results have shown the fact that this method is more efficient than a number of other hashing approaches on the MNIST and CIFAR10 data-sets. The work can be extended by investigating its efficiency and scalability on a data-set of a larger scale. The experimental results which indicated that this approach can enhance some earlier best retrieval results with 30% and 1% retrieval accuracy on the CIFAR-10 and MNIST data-sets respectively with only a simple alteration of the deep CNN. This method provided 83.75% precision (which was obtained by the last layer) on the task of 116 classes of clothing classification.
Another structure proposed by Siying Zhu *et al*[13] had merged the process of generating binary codes within deep NNs for sufficient retrieval of images. The suggested model included two main blocks. The stacked convolution layers of Network-In-Network with global average pooling for the calculation of the sufficient representation of images and the embedded latent layer with binary activation functions learn binary hash codes in a simultaneous manner. Experiments have shown that the suggested approach has gained improvements over a number of earlier hashing approaches on two well-known deep learning data-sets, which are MNIST and CIFAR-10. These researchers had implemented their suggested structure on the NN toolkit Caffe1. In all of the experiments, the networks were trained via stochastic gradient descent with a learning rate of 0.01 and weight decay of 0.0005. The training iterations were 10,000 for each one of the datasets. The experimental results showed that the precision of the approach while using 48 bits is 0.98 and 0.587 on MNIST and CIFAR10 respectively.
Another joint binary code learning approach has been suggested by Xuelong Li *et al* [14] for combining image feature to latent semantic feature with as little encoding loss as possible. This was known as Latent Semantic Minimal Hashing. The latent semantic feature has been learned according to matrix decomposition for refining the initial feature, this way it makes the learned feature more discriminative. In addition to that, a minimum encoding loss had been combined with the process of latent semantic feature learning in a simultaneous way, in order to ensure that the obtained binary codes are also discriminative. Extensive experiments on numerous common large data-bases have shown that the suggested method performed more efficiently compared to numerous other approaches of hashing. The conclusion is that this latter approach that has been described [14] is, in general, more efficient compared to the rest of the approaches of this sub-section due to the fact that its retrieval precision has reached up to 0.98%.

2.3 Content-Based Image Retrieval (CBIR)

Huafeng Wang *et al.*[15] suggested an architecture based on content-based image retrieval systems (CBIRs) which consists of three parts: feature extraction, processing and indexing, as shown in Figure 3. Via selecting the intermediate model layers like feature representation, and pre-processing the data with some important approaches, the CBIR task based on CNN could noticeably be enhanced. Retrieval performance was between 34.40% and 53.41%.

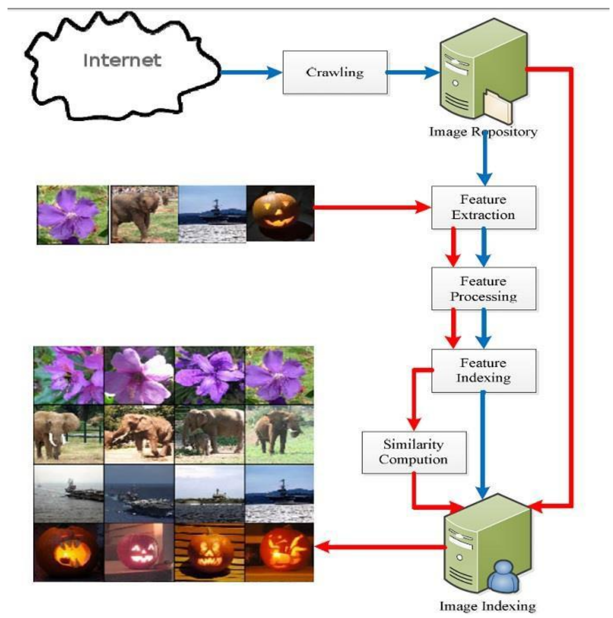

Figure 3. Simplified industrial CBIR architecture[15]

A second method designed by Yang Li *et al* [16] called a deep feature hash codes model for CBIRs, where they initially extracted features of the image by a pretrained CNN model. Then, they used various hashing approaches for binary feature extraction. Lastly, they used the optimal binary encoding features for building a CBIRs. The experimental results demonstrated that with decreasing feature dimension, the approach did not decrease the accuracy of retrieval and could as well improve the precision of retrieval in some of the cases. The precision of retrieval of 256 bits binary characteristics might exceed the conventional approach of 256 dimensional (4096 bits) features. As soon as the feature bits were 16 times lower, the space of the storage had decreased 16 times and the performance of retrieval has increased. Which is why, this approach could sufficiently enhance the speed and accuracy of CBIRs. The efficiency of retrieval was in the range from 50.08 to 77.55. The work can be additionally improved with the addition of the approach of matrix learning in this framework and enhance the prcision of the CBIR retrieval.

A third method is large-scale remote sensing (RS) retrieval of images that had become a valuable research problem in geo-sciences. This method was proposed by Peng Li and Peng Ren [17]. It is compatible with the fast evolution of the technologies of satellite and aerial vehicle. Methods of Hashing-based searching have been commonly utilized in the tasks of content-based retrieval of images. On the other hand, the majority of hash approaches make a trade-off between the precision of retrieval accuracy and the effectiveness of learning, and therefore can hardly meet the exact RS data analysis requirements. For the sake of addressing those drawbacks, they have proposed an approach for partial randomness in order to learn hash functions, which had been called as partial randomness hashing (PRH). Particularly, for the construction of hash functions, a part of model parameter values were arbitrarily produced and the rest of them have been trained on the basis of the RS images. Experiments on 2 large public RS image datasets had proven the fact that this PRH approach had outperformed several other related algorithms according to each of precision of retrieval and effectiveness of learning. Retrieval performance was between 0.4138 to 0.5202 using MAP. Finally, comparing these three methods to each other, one generally may conclude that **t**he second method [16] is superior to the other methods because it produced the best retrieval performance that reached to 77.55.

2.4 Neural Network based Hash Function Method for Authentication and Retrieval of Color Images

In this approach, we have noticed two interesting methods. The first method was proposed by Yakup Kutlu and Abdullah Yayik [18]. The main idea of this method was re-sizing input colour image to a constant size, after that, generating hash values with the use of NN one-way feature and nonlinear approaches, for three dimensions respectively. This work proposed a NN-based hash function architecture for colour image authentication. The presented system presented binary and hexadecimal sensitivity of hash value. The binary sensitivity was nearly 50% that satisfies diffusion of the hash value. Also, almost 100% hexadecimal sensitivity obtained meaning that the algorithm has very high ability of robustness. The binary sensevity reached 51.36%

The second method designed by Yang Li and Zhuang Miao [19], which included combining non-linear reduction of dimension and hashing approach for effective retrieval of images. They firstly extracted 4096-dimension characteristics via a pretrained CNN model. After that, they have used t-Distributed Stochastic Neighbour Embedding (t-SNE) for the reduction of the deep characteristics to 1024-dimension. Lastly, they have utilized Sparse Projection (SP) for building 256 bits binary encoding characteristics for retrieving images. They assessed the efficiency of their approach with the Oxford-5k, Paris-6k and Holidays data-sets. Experiments on those three datasets showed that the performance was reached to 81.35 using MAP. The retrieval performance of this latter method was higher than the previous one as it reached to 81.35.

2.5 Other Supervised Learning Techniques

In this subsection, we will describe some other interesting techniques for supervised learning.

2.5.1 Siamese Architecture

This approach was proposed by Abin Jose *et al*[20]. for learning binary codes for fast retrieval of images. A Siamese model was used with 2 parallel feed-forward branches but with a shared weight for generating binary codes. The training data was split to similar and different pairs. The NN attempts learning the weights in a way that it decreases the distance between similar pairs of images and increases the distance between pairs of images which are not similar. The binary codes had been formed via squashing the output of the NN through a sigmoid function of activation, as shown in Figure 4. The training with sigmoid hash constrained the result of every node in the final fully connected layer into either 1 or 0. The input was fed to the NN in the form of image pairs. The result of the fully connected layer was transmitted to the sigmoid function. The distance between the result feature vectors had been computed inside the function of loss. Retrieval performance reached to 67.19 on CIFAR-10.

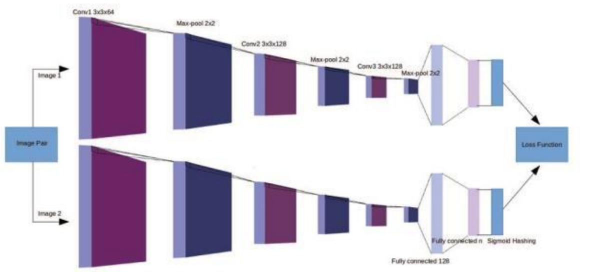

Figure 4: Architecture of the Siamese NN.[20]

2.5.2 Deep-Networks based Hashing for Multi-label Images

Another approach designed by Hanjiang Lai *et al* [21] focused on deep-networks-based hashing for images that have multi-label; every one of these images may contain objects of multiple categories, as shown in Figure 5. Also, in the most common hashing methods, the representation of each image consists of one piece of hash code, which is called semantic hashing. The performance of this approach reached to 0.8830 using MAP.

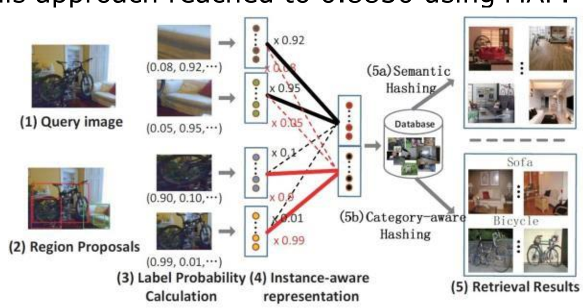

Figure 5: Illustration of Instance-Aware Image Retrieval.[21]

2.5.3 The Well-Known Model of Bag-of-Features (BoFs) & A Symmetry-Aware Spatial Segmentation Approach

Nikolaos Passalis and Anastasios Tefas [22] proposed the bag-of-features (BoFs) approach which consists of 3 layers: A radial basis function (RBF) layer, fully connected layer, and accumulation layer, as shown in Figure 6. This module allows for split the size of the representation from the number of utilized code-words and also for a more sufficient formulation the scattering of features with the use of an independent trainable for every one of the RBF neurons. The final network, referred to as the "retrieval oriented neural BoF" (RN-BoF) was trained by the use of the regular backpropagation and takes into consideration the quick extraction of conservative picture representations. The RN-BoF approach was capable of accelerating the speed of retrieval and encoding of objects, minimizing the size of obtained representation, and increasing the accuracy of retrieving. Another technique called "A symmetry-aware spatial segmentation" was also proposed for minimizing the storage needing and time of encoding which make this approach capable of scaling large datasets in an efficient way. The performance of this latter approach reached to 97.87 % using MAP. This work can be further enhanced by combining the RN-BoF model with extreme learning techniques for reducing the time of training.

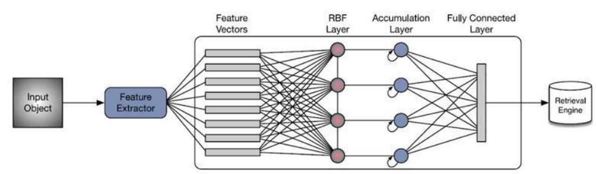

Figure 6. Proposed RN-BoF model.[22]

2.5.4 Supervised Semantics-Preserving Deep Hashing (SSDH)

Huei-Fang Yang *et al*[23]. designed an approach called effective supervised deep hash approach, where the constructing of the binary hash codes was from labelled data for image search which is of large-scale. Those researchers assumed the fact that the semantic labels are administrated via multiple interior specifications with every specification On or Off, and that the classification depends on those specifications. Constructing hash functions was done as a latent layer in a deep NN. The binary codes were trained by reducing an objective function which has been defined over classification error and other wanted hash codes characteristics, as shown in Figure 7. With this model, SSDH achieved a good attribute where a single learning model contains both classification and retrieval parts [23][41][42]. The performance of this approach reached to 91.45 % and 99.39% on MNIST and CIFAR10 datasets respectively using MAP.

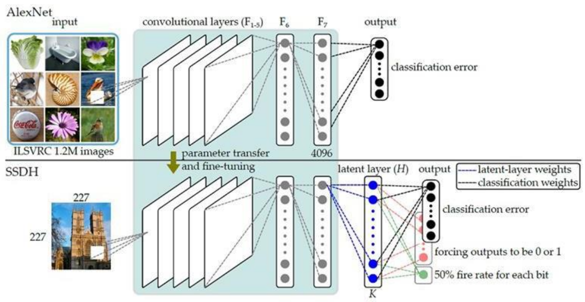

Figure 7: Supervised Semantic-Preserving Deep Hashing (SSDH).[23]

2.5.5 Bit-Scalable Hashing Approach

Another supervised learning architecture was proposed by Ruimao Zhang *et al.*[24], where the raw images used to directly create compact and bit-scalable hashing codes, as shown in Figure 8. Firstly, the triplet samples should be generated by organising the training images into a batch, and every sample includes 3 images where two of those images are in the same category while the third one is different. Then in these samples, the margin between the identical pairs and mismatch pairs should be maximized in the Hamming space. The images of similar components must have similar codes, also in the resulting hashing code, each bit should be weighted individually. The performance of this approach reached to 63.26 %, 98.09% and 64.14 on MNIST, CIFAR10 and NUS-WIDE data-sets respectively using MAP. This framework can be further enhanced via leveraging more semantics (*for instance,* numerous attributes) of images and/or introducing feedback learning in the framework such as to make it more powerful in practice.

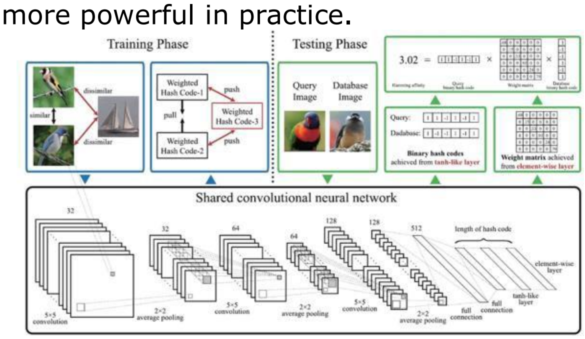

Figure 8: The Bit-Scalable Deep Hashing Learning Framework.[24]

2.5.6 Deep Hashing (DH) Method for Learning Compact Binary Codes for Scalable Image Search

This architecture was designed by Jiwen Lu *et al* [25]. It differs from most other learning techniques that it has multiple linear projection of mapping each image into binary feature vector, as shown in Figure 9. Those researchers built up a deep NN to look for numerous hierarchical nonlinear transformations for learning the binary codes, with the goal that the non-linear correlation of tests may be all around misused. They pointed out two main headings for future work: The first one is to stretch out this approach to deal with adaptable video search. And the second one is to stretch out this approach to deal with the cross-modular search for scalable multimedia search. The Hamming ranking of this approach using MAP reached to 0.72 on CIFAR10 data-set.

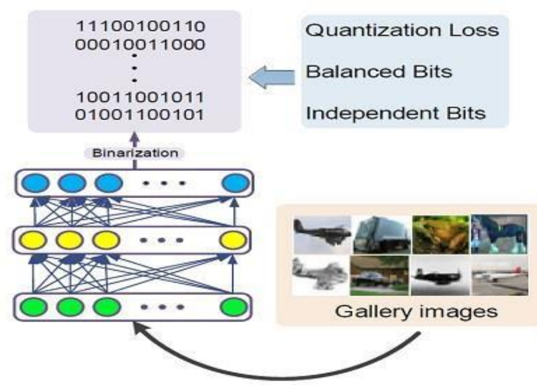

Figure 9: The Main Idea of Deep Hashing Method for Compact Binary Codes Learning.[25]

2.5.7 The Method of Query-adaptive Deep Weighted Hashing (QaDWH)

The query-adaptive deep weighted hashing (QaDWH) approach was proposed by Jian Zhang and Yuxin Peng [26][377][38]. This approach is capable of performing fine-grained ranking for various queries by weighted Hamming distance (See Figure 10). Here the researchers used an example of 6-bits hashing codes. Supposing that the query image has a hash code of 000000, there are six images within Hamming radius 1 with query image, however, they are different in various bits. [39][40]Conventional Hamming distance is not capable of performing fine-grained ranking amongst them. The proposed approach consists of two parts: First, designing a new deep hash network, consisting of 2 streams: the hashing stream learns the compact hashing codes and corresponding class wise hash bit weights in a simultaneous manner, whereas the classification stream maintains the semantic info and enhances the efficiency of the hash. After that, designing a sufficient method of query-adaptive retrieval of images that initially rapidly creates the query-adaptive hashing weights based on the class-wise weights and the projected semantic probability of the query, and after that performs sufficient image retrieval via weighted Hamming distance. As a future work, the researchers suggested extending the deep weighted hashing approach to a multi-table framework of deep hash, where various weights have been learned for various functions of hashing map. The Hamming ranking of this approach using MAP reached to 0.884 on CIFAR-10 dataset.

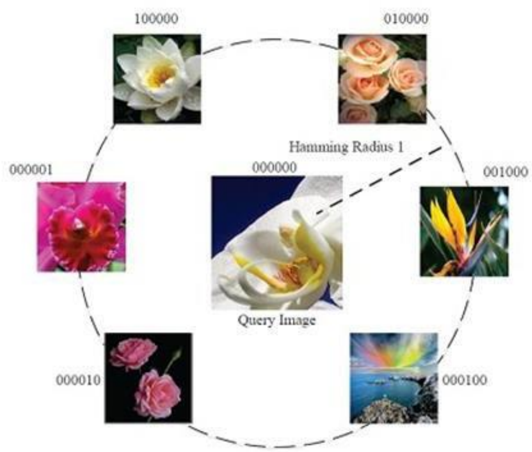

Figure 10: The Discrete Hamming Distance.[26]

2.5.8 Jointly Sparse Regression (JSH) Model

The purpose of this image retrieval model was minimizing the loss of locality information and get jointly sparse hash approach. The presented approach by Zhihui Lai *et al* [27] integrated joint scarcity, locality and process of rotation together in a smooth construction. The presented JSH could learn the best jointly sparse matrix of projection for low-dimensional feature extracting and selecting. With the orthogonal constraints in the form of a rotational operation between the binary codes and the low-dimensional characteristics, the suggested model could additionally reduce the loss of information and get the binary solutions in a direct manner The Hamming ranking of this approach using MAP reached to 0.1913 on CIFAR-10 dataset.

Comparing the eight latter techniques mentioned in this subsection, it can be noticed that the approach proposed by Nikolaos Passalis and Anastasios Tefas[22] achieved the best image retrieval ration of 97.87 %.

3. Semi-Supervised Techniques

Usually, it is hard to obtain labelled data. In the case of using the label data alone for training, the training dataset would be excessively small for entirely reflecting the characteristics of the traffic. New applications of network are continuously developed, which results in traffic data with no label. Each one of the situations results in issues for the conventional supervised ML approaches. For addressing this issue, the semi-supervised ML approach has emerged, because of its capability of combining unsupervised and supervised learning. In semi-supervised learning, the training dataset includes each of samples that are labelled or unlabelled [1]. In this section, some of the most interesting semi-supervised learning proposals are reviewed.

3.1 Semi-Supervised Hashing for Scalable Image Retrieval

J. Wang *et al* [28] suggested a semi-supervised hashing method which was formulated in a form of a minimization of empirical error on the labeled data whereas increasing variance and independence of hash bits over the labeled and unlabeled data. The suggested approach was capable of handling both metric in addition to the semantic similarity. This method can be enhanced via relaxing the widely utilized constraints of orthogonality in a way that one is capable of achieving better results, particularly for larger number of bits.

3.2 Semi-supervised Kernel Hyperplane Learning (SKHL)

Meina Kan *et al* [29] proposed a hashing approach that has been referred to as the Semi-supervised Kernel Hyper-plane Learning (SKHL) for semantic image retrieving via the modeling of every hashing function as a non-linear kernel hyper-plane that has been constructed from an unlabeled data-set. In addition to that, a Fisher-like criterion had been suggested for learning the best kernel hyperplane and hashing function, with the use of only weakly labeled samples of training with side information. For the sake of additionally integrating various feature types, they have incorporated multiple kernel learning (MKL) in the suggested SKHL (referred to as the SKHL-MKL), that led to a more sufficient hashing function. The work can be extended by studying how to update them jointly. It is also justifiable to apply this method for the retrieval of videos with the use of the video key frames as the input.

3.3 Semi-Supervised composite Multi-view Discrete Hash (SSMDH) Model

At the time where the majority of researches on the subject of hash models have been focused on single-view data, lately the multiview methods with a majority of unsupervised multi-view hashing models were taken under consideration. In addition to the incorporation of a part of label data in the model, the suggested multiview model that was designed by Wei-Shi Zheng and C. Zhang [30] is different from the available multi-view hashing models in thrree-fold: 1) a composite discrete hashing learning modelling which is capable of jointly minimizing the loss on multi-view characteristics in the case of utilizing relaxation on learning hash codes, 2) a composite locality preserving modelling for locally compact coding, 3) the exploration of statistically uncorrelated multiview characteristics for the generation of hashing codes. A future extension of this work can be investigating if an entirely end-to-end multi-view semi-supervised architecture would get considerably more enhancement.

3.4 Semi-Supervised Deep Hashing (SSDH) Method

This approach has been presented by Jian Zhang and Yuxin Peng [31] for performing better hash function learning via a simultaneous preservation of the semantic similarity and underlying data structures. The basic contributions were the following: 1) a semi-supervised loss for jointly minimizing the empirical error on labeled data, in addition to the embedding error on each of the unlabeled and labeled data that might maintain the semantic similarity and record the meaningful neighbours on the underlying data structures for sufficient hashing. 2) a SSDH network has been proposed for broadly exploiting each of the labeled and unlabeled data. The suggested deep NN performed learning of hashing code and learning of features in a simultaneous manner in a semi-supervised manner. As a future work, it is possible to discover a variety of semi-supervised embedding methods which might take more advantage of the unlabeled data, and more improved strategies of graph construction may be used. Indeed, this framework can be extended to an unsupervised scenario, in which clustering methods are used for obtaining virtual image labels.

3.5 The Approach of Semi-Supervised Metric Learning-based Anchor Graph Hashing (MLAGH)

Another approach was presented by Haifeng Hu *et al.* [32]. This method may be split to 3 parts. 1) utilizing a transform matrix for the construction of the anchor-based graph of similarity of the training dataset. 2) proposing the objective function which is based on the triplet correlation, where the best transform matrix may be learnt with the use of the label smoothness and the margin hinge loss which is incurred by the triplet constraint. In addition to that, the approach of stochastic gradient descent (SGD) can leverage the gradient on every one of the triplets for updating the transform matrix. 3) designing a penalty factor for accelerating the speed of execution for the SGD.

3.6 Comparison of the Semi-Supervised Methods

Comparing the semi-supervised learning based image retrieval methods mentioned in the present section, a preference can be assigned to the second method (SKHL) [29] in general broad terms. This can be for the following reasons: First, every one of the hashing functions is modeled in a form of a nonlinear kernel hyper-plane that has been constructed from an unlabeled data-set. By the maximization of a Fisher-like criterion on a weakly labeled data-set only with side information, one can get a collection of hashing functions and optimal kernel hyper-planes. In addition to that, every one of the hashing functions is independently updated in every one of the iterations. This approach is applicable as well for the retrieval of videos with the use of video key frames as the input.

4. Unsupervised Learning Techniques

Unsupervised approaches detect internal correlations in the unlabeled input data. One of the main unsupervised approaches is the clustering. Even though clustering requires no class labels, classifiers may be derived in the case where the traffic clusters are corresponding to various applications of the network [1]. In this section, some of the most interesting unsupervised learning proposals are considered.

4.1 Cascaded Principal Component Analysis (PCA)

This architecture was proposed by Tsung-Han *et al* [33]. The main segment is cascaded principal component analysis (PCA), the second one is binary hashing, and the last one is block-wise histograms. For comparing and to present a good understanding, the paper also introduced and studied two simple PCA-Net variations, which are: Rand-Net and LDA-Net. Experimentations on other public data-sets also demonstrated the possibility of PCA-Net in serving as a simple yet very competitive base-line for classification of textures and recognition of objects. As soon as the parameters get fixed, the training of the PCA-Net is very simple and sufficient due to the fact that the filter learning in the PCA-Net involves no regularized parameters or requires numerical optimization solvers. In addition to that, the construction of the PCA-Net includes only a cascaded linear map, which is followed by a non-linear output phase. This level of simplicity presents an alternative, but at the same time refreshing potential on CNNs and might additionally simplify the mathematical justification and analysis of their efficiency. The two simple PCA-Net extensions, which are the Rand-Net and LDA-Net, have been tested in combination with PCA-Net on a wide range of tasks of image classification, which include faces, objects, handwritten digits, and textures,. The extensive experimental results showed that the PCA-Net performs batter than the Rand-Net and the LDA-Net. The bottleneck which might prevent PCA-Net from getting deeper (for instance to more than 2 stages) is the fact that the resultant feature dimension would exponentially maximize with the number of stages. Which can probably be fixed via the replacement of 2-D convolutional filters with tensor-like one as a future work.

4.2 Quaternion Orthogonal Matching Pursuit (Q-OMP)

V. Risojevic and Z. Babic [34] proposed unsupervised learning of quaternion feature filters using quaternion representation for color images, in addition to feature encoding with the use of a quaternion orthogonal matching pursuit (Q-OMP). With the use of quaternion representation, there was a possibility of jointly encoding color information and intensity in an image. Local descriptors have been obtained with the use of soft thresholding and the calculation of the absolute values of scalar and 3 vector parts of the quaternion valued distributed code. Local descriptors had been pooled, normalized, and power-law transformed, thereby resulting in the output image representation. The suggested approach for quaternion feature learning has been capable of adapting to the properties of the available data, and being entirely unsupervised It has appeared as a suitable substitute to each of convolutional NNs and hand-crafted representations, particularly in application scenarios that include scarce-labeled training data. This approach might as well be extended to hierarchical classification structure.

4.3 Siamese-Twin Random Projection Neural Network (ST-RPNN)

Mohamed Fahkr *et al* [35] designed ST-RPNN approach that comprises two similar random projection NNs with hard thresholding neurons in which the binary code is fed as the results for the neuron. The learning goal was producing similar binary codes for similar pairs of input image and dissimilar binary codes in the opposite case. The procedure of learning has been split to 2 stages. Initially, over-complete random projection has been utilized for producing a suitably long code, which has been followed by a fast sparse technique for neurons selection (FSNS). Bootstrap Aggregation Trees or Bagging Trees (BT) has been after that, utilized for making an enhanced compact code section. BT has as well been utilized as a fast retrieval tool which was used for ranking the data-base in terms of a query with no calculations of distance and with a considerably lower level of complexity compared to the method of Hamming distance.

4.4 Converting Unsupervised Hashing to Supervised Hashing using Pseudo Labels of Images

Haofeng Zhang *et al* [36] proposed an unsupervised model which had 2 main contributions: The first one is converting the unsupervised DH architecture to supervised via learning pseudo labels and the second one is the framework unifies probability maximizing, quantization error minimizing, and mutual information maximizing, in a way that the pseudo labels may maximally be preserved.

4.5 Comparison of Unsupervised Methods

Comparing the unsupervised learning based image retrieval methods mentioned in this section, the first method [33] seems to be the best one in general terms. It is the simplest unsupervised convolutional deep learning network that called cascaded principal component analysis due to the fact that the filter learning in the PCA-Net involves no regularized parameters or requires numerical enhancement solvers. In addition to that, constructing the PCA-Net comprises only a cascaded linear map, which is followed by a non-linear output phase. Also, this architecture exceptionally straightforward deep learning network for classifying images which depends on extremely basic data processing parts.

5. Conclusion

This paper includes a survey of some computer vision techniques that are used in image retrieval field. These techniques are commonly utilized in a wide range of applications like intelligent vehicles, medical application, etc. Based on the previous works, we conclude that the Nearest Neighbour method in the supervised manners generally worked better than the other considered supervised methods because it has higher retrieval accuracy. In addition, the SKHL method in semi-supervised approaches seems to be better than other semi-supervised ones for these reasons: every one of the hashing functions has been modeled as a non-linear kernel hyper-plane, which has been constructed from an unlabeled data-set. Via the maximization of a Fisher-like criterion on a data-set which is weakly labeled only with side information, obtaining a group of optimal kernel hyper-planes and hashing functions. In addition to that, every hashing function has been independently updated in every one of the iterations, and the presented approach can be applicable for retrieval of videos with the use of the video key frames as inputs.

For unsupervised scenarios, the Cascaded Principal Component Analysis (PCA) looks to be the best one in general because it is the simplest convolutional deep learning network. Indeed, this architecture exceptionally straightforward deep learning NN for image classification which depends on extremely basic parts of the data processing. In general terms, the supervised models usually are better comparing with other models because the supervised learning has a previous knowledge acquired by training (using class labels). Furthermore, supervised learning is typically faster than other learning techniques such that semi-supervised and un-supervised techniques because these techniques need to classify objects according to some specification using multi-stages classification that make these techniques very slow.

References

[1] G. Vargas-Solar, J. A. Espinosa-Oviedo, and J. L. Zechinelli-Martini, *Big continuous data: Dealing with velocity by composing event streams*. 2016.

[2] P. Tino, L. Benuskova, and A. Sperduti, "Artificial Neural Network Models," vol. 8, no. 3, pp. 455–472, 1997.

[3] Q. Li, Z. Sun, R. He, and T. Tan, "Deep Supervised Discrete Hashing," no. Nips, 2017.

[4] D. Yang, H. Xie, J. Yin, Y. Liu, and C. Yan, "Supervised deep quantization for efficient image search," *2017 IEEE Int. Conf. Multimed. Expo Work. ICMEW 2017*, no. July, pp. 525–530, 2017.

[5] C. Yan, H. Xie, D. Yang, J. Yin, Y. Zhang, and Q. Dai, "Supervised Hash Coding with Deep Neural Network for Environment Perception of Intelligent Vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 284–295, 2018.

[6] H. Lai, Y. Pan, Y. Liu, and S. Yan, "Simultaneous feature learning and hash coding with deep neural networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07–12–June, pp. 3270–3278, 2015.

[7] M. Sapkota, X. Shi, F. Xing, and L. Yang, "Deep Convolutional Hashing for Low Dimensional Binary Embedding of Histopathological Images," *IEEE J. Biomed. Heal. Informatics*, vol. 2194, no. c, pp. 1–12, 2018.

[8] N. Aslam, "Limitation and Challenges: Image/Video Search & Retrieval," *Int. J. Digit. Content Technol. its Appl.*, vol. 3, no. 1, 2009.

[9] M. Turcanik and M. Javurek, "Hash function generation by neural network," *NTSP 2016 - Proc. Int. Conf. New Trends Signal Process.*, no. December 2017, 2016.

[10] T. Komal, R. Ashutosh, R. Roshan, and *A.* M. Nalawade, "Encryption and Decryption using Artificial Neural Network," *Int. Adv. Res. J. Sci. Eng.*

Technol., vol. 2, no. 4, pp. 2393–2395, 2015.

[11] "Artificial Neural Networks (ANN) and their Types." [Online]. Available: https://www.elprocus.com/artificial-neural-networks-ann-and-their-types/. [Accessed: 20-Mar-2019].

[12] Jun-yi Li and J. Li, "Fast image search with deep convolutional neural networks and efficient hashing codes," *2015 12th Int. Conf. Fuzzy Syst. Knowl. Discov.*, pp. 1285–1290, 2015.

[13] S. Zhu, B. N. Kang, and D. Kim, "A deep neural network based hashing for efficient image retrieval," *2016 IEEE Int. Conf. Syst. Man, Cybern. SMC 2016 - Conf. Proc.*, pp. 2483–2488, 2017.

[14] X. Lu, X. Zheng, and X. Li, "Latent Semantic Minimal Hashing for Image Retrieval," *IEEE Trans. Image Process.*, vol. 26, no. 1, pp. 355–368, 2017.

[15] H. Wang, Y. Cai, Y. Zhang, H. Pan, W. Lv, and H. Han, "Deep Learning for Image Retrieval: What Works and What Doesn't," *Proc. - 15th IEEE Int. Conf. Data Min. Work. ICDMW 2015*, pp. 1576–1583, 2016.

[16] Y. Li, Y. Xu, Z. Miao, H. Li, J. Wang, and Y. Zhang, "Deep feature hash codes framework for content-based image retrieval," *2016 8th Int. Conf. Wirel. Commun. Signal Process. WCSP 2016*, 2016.

[17] P. Li and P. Ren, "Partial Randomness Hashing for Large-Scale Remote Sensing Image Retrieval," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 464–468, 2017.

[18] Y. Kutlu and A. Yayik, "Neural Network Approach for Secure Hash-Based Color Image Authentication and Analysis," *Int. J. Sci. Technol. Res.*, vol. 1, no. 7, pp. 27–37, 2015.

[19] Y. Z. Yang Li, Zhuang Miao, Yulong Xu, Hang Li, "Combining Nonlinear Dimension Reduction and Hashing Method for Efficient Image Retrieval," *Asian J. Pharm. Clin. Res.*, vol. 11, no. 10, pp. 238–341, 2018.

[20] A. Jose, S. Yan, and I. Heisterklaus, "Binary hashing using siamese neural networks," *Proc. - Int. Conf. Image Process. ICIP*, vol. 2017–Septe, pp. 2916–2920, 2018.

[21] H. Lai, P. Yan, X. Shu, Y. Wei, and S. Yan, "Instance-aware hashing for multi-label image retrieval," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2469–2479, 2016.

[22] N. Passalis and A. Tefas, "Learning Neural Bag-of-Features for Large Scale Image Retrieval," vol. 14, no. 8, pp. 1–12, 2015.

[23] H. F. Yang, K. Lin, and C. S. Chen, "Supervised Learning of Semantics-Preserving Hash via Deep Convolutional Neural Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 2, pp. 437–451, 2018.

[24] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Deep Hashing Learning with Regularized Pairwise Comparison for Image Retrieval," vol. 24, no. 12, pp. 4766–4779, 2015.

[25] J. Lu, V. E. Liong, and J. Zhou, "Deep Hashing for Scalable Image Search," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2352–2367, 2017.

[26] J. Zhang and Y. Peng, "Query-adaptive image retrieval by deep-weighted hashing," *IEEE Trans. Multimed.*, vol. 20, no. 9, pp. 2400–2414, 2018.

[27] Z. Lai, Y. Chen, J. Wu, W. K. Wong, and F. Shen, "Jointly Sparse Hashing for Image Retrieval," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 6147–6158, 2018.

[28] S.-F. C. Jun Wang, Sanjiv Kumar, "Semi-Supervised Hashing for Scalable Image Retrieval," *Int. J. Soc. Welf.*, vol. 14, no. 2, pp. 107–115, 2005.

[29] Meina Kan, Dong Xu, Shiguang Shan, and Xilin Chen, "Semisupervised Hashing via Kernel Hyperplane Learning for Scalable Image Search," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 4, pp. 704–713, 2014.

[30] C. Zhang and W. S. Zheng, "Semi-Supervised Multi-View Discrete Hashing for Fast Image Search," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2604–2617, 2017.

[31] J. Zhang and Y. Peng, "SSDH: Semi-Supervised Deep Hashing for Large Scale Image Retrieval," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 212–225, 2019.

[32] H. Hu, K. Wang, C. Lv, J. Wu, and Z. Yang, "Semi-Supervised Metric Learning-Based Anchor Graph Hashing for Large-Scale Image Retrieval," *IEEE*

Trans. Image Process., vol. 28, no. 2, pp. 739–754, 2019.

[33] Y. M. Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, "PCANet: A Simple Deep Learning Baseline for Image Classification," vol. 24, no. 12, pp. 5017–5032, 2015.

[34] V. Risojevic and Z. Babic, "Unsupervised Quaternion Feature Learning for Remote Sensing Image Classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 9, no. 4, pp. 1521–1531, 2016.

[35] M. W. Fahkr, M. M. Emara, and M. B. Abdelhalim, "Siamese-Twin random projection neural network with Bagging Trees tuning for unsupervised binary image hashing," *5th Int. Symp. Comput. Bus. Intell. ISCBI 2017*, pp. 14–19, 2017.

[36] H. Zhang, L. Liu, Y. Long, L. Shao, and S. Member, "Unsupervised Deep Hashing With Pseudo Labels for Scalable Image Retrieval," vol. 27, no. 4, pp. 1626–1638, 2018.