Deep Learning for Image Search and Retrieval in Large Remote Sensing Archives

Gencer Sumbul, Jian Kang, and Begüm Demir

Abstract

This chapter presents recent advances in content based image search and retrieval (CBIR) systems in remote sensing (RS) for fast and accurate information discovery from massive data archives. Initially, we analyze the limitations of the traditional CBIR systems that rely on the hand-crafted RS image descriptors applied to exhaustive search and retrieval problems. Then, we focus our attention on the advances in RS CBIR systems for which the deep learning (DL) models are at the forefront. In particular, we present the theoretical properties of the most recent DL based CBIR systems for the characterization of the complex semantic content of RS images. After discussing their strengths and limitations, we present the deep hashing based CBIR systems that have high time-efficient search capability within huge data archives. Finally, the most promising research directions in RS CBIR are discussed.

1 Introduction

With the unprecedented advances in the satellite technology, recent years have witnessed a significant increase in the volume of remote sensing (RS) image archives. Thus, the development of efficient and accurate content based image retrieval (CBIR) systems in massive archives of RS images is a growing research interest in RS. CBIR aims to search for RS images of the similar information content within a large archive with respect to a query image. To this end, CBIR systems are defined based on two main steps: i) image description step (which characterizes the spatial and spectral information content of RS images); and ii) image retrieval step (which evaluates the similarity among the considered descriptors and then retrieve images similar to a query image in the order of similarity). A general block scheme of a CBIR system is shown in Figure 1.

Traditional CBIR systems extract and exploit hand-crafted features to describe the content of RS images. As an example, bag of-visual-words representations built upon the scale invariant feature transform features are introduced in [1]. In [2], a bag-of-morphological-words representation of the local morphological texture descriptors is proposed in the context of CBIR. In [3], a comparative study that analyzes and compares advanced local binary patterns (LBPs) in RS CBIR problems is presented. To define the spectral information content of high dimensional RS images, the bag of spectral values descriptors are presented in [4]. Graph-based image representations, where the nodes describe the image region properties and the edges represent the spatial relationships among the regions, are presented in [5], [6]. In [7], [8] image representations through binary hash codes are introduced for large-scale CBIR problems in RS. In detail, in [7] kernel-based hashing methods are presented, whereas in [8] a partial randomness hashing method is introduced.

Once image descriptors are obtained, one can use the k-nearest neighbor (k-NN) algorithm, which computes the similarity between the query image and all archive images to find the k images most similar to the query. If the images are represented by graphs, graph matching techniques can be used. As an example, in [6] an inexact graph matching approach, which is based on the sub-graph isomorphism and spectral embedding algorithms, is presented. If the images are represented by binary hash codes, image
retrieval can be employed based on either hash lookup or hamming ranking that allow time-efficient search capability [2]. However, these unsupervised systems do not always result in satisfactory query responses due to the semantic gap, which is occurred among the low-level features and the high-level semantic content of RS images [3]. To overcome this limitation and improve the performance of CBIR systems, semi-supervised and fully supervised systems, which require user feedback in terms of RS image annotations, are introduced [9]. Most of these systems depend on the availability of training images, each of which is annotated with a single high-level land-use or land-cover category label that is associated to the most significant content of the image. However, RS images typically contain multiple classes and thus can simultaneously be associated with different class labels. Thus, supervised CBIR methods that properly exploit training images annotated by multi-labels are recently found very promising in RS. As an example, in [4] a CBIR system that exploits a measure of label likelihood based on a sparse reconstruction-based classifier is presented in the framework of multi-label RS CBIR problems. Semi-supervised CBIR systems based on graph matching algorithms are proposed in [10], [11]. In detail, in [10] a three-layer framework in the context graph-based learning is proposed for query expansion and fusion of global and local features by using the label information of query images. In [11] a correlated label propagation algorithm, which operates on a neighborhood graph for automatic labeling of images by using a small number of training images, is proposed.

The above-mentioned CBIR systems rely on shallow learning architectures and hand-crafted features. Thus, they can not simultaneously optimize feature learning and image retrieval, resulting in compromised search and retrieval performance in practice and limited representation capability for the high-level semantic content of RS images. Recent advances in deep neural networks (DNNs) have triggered substantial performance gain for image retrieval due to their high capability to encode higher level semantics present in RS images. Differently from conventional CBIR systems, deep learning (DL) based CBIR systems learn image descriptors in such a way that feature representations are optimized during the image retrieval process. In order words, DNNs eliminate the need for human effort to design discriminative and descriptive image descriptors for the retrieval problems. Most of the existing RS CBIR systems based on DNNs attempt to improve image retrieval performance by: 1) learning discriminative image descriptors; and 2) achieving scalable image search and retrieval. The aim of this chapter is to present different DNNs proposed in the literature for the retrieval of RS images. The rest of this chapter is organized as follows. Section II reviews the DNNs proposed in the literature for the description of the complex information content of RS images in the framework of CBIR. Section III presents recent progress on the scalable CBIR systems defined based on DNNs in RS. Finally, Section IV draws the conclusion of this chapter.
2 Description of RS Images Based on Deep Learning for CBIR

The DL based CBIR systems in RS differ from each other in terms of: i) the strategies considered for the mini-batch sampling; ii) the approaches used for the initialization of the parameters of the considered DNN model; iii) the type of the considered DNN; and iv) the strategies used for image representation learning. Figure 2 illustrates the main approaches utilized in DL based CBIR systems in RS. In detail, a set of training images is initially selected from the considered archive to train a DNN. Then, the selected training images are divided into mini-batches and fed into the considered DNN. After initializing the model parameters of the network, the training phase is conducted with an iterative estimation of the model parameters based on a loss function. The loss function is selected on the basis of the characteristics of the adapted learning strategy.

During the last years, several DL based CBIR systems that adapt different strategies for the above-mentioned factors are presented. As an example, in [12] an unsupervised feature learning framework that learns image descriptors from a set of unlabeled RS images based on an auto-encoder (AE) is introduced. After random selection of mini-batches and initialization of the model parameters, SIFT based image descriptors are encoded into sparse descriptors by learning the reconstruction of the descriptors. The learning strategy relies on minimization of the reconstruction loss between the SIFT descriptors and the reconstructed image descriptors in the framework of the AE. A CBIR system that applies a multiple feature representation learning and a collaborative affinity metric fusion is presented in [13]. This system randomly selects RS images for mini-batches and initializes the model parameters of a Convolutional Neural Network (CNN). Then, it employs the CNN for k-means clustering (instead of classification). To this end, a reconstruction loss is utilized to minimize the error induced between the CNN results and the cluster assignments. Collaborative affinity metric fusion is employed to incorporate the traditional image descriptors (e.g., SIFT, LBP) with those extracted from different layers of the CNN. A CBIR system with deep bag-of-words is proposed in [14]. This system employs a convolutional auto-encoder (CAE) for extracting image descriptors in an unsupervised manner. The method first encodes the local areas of randomly selected RS images into a descriptor space and then decodes from descriptors to image space. Since encoding and decoding steps are based on convolutional layers, a reconstruction loss is
directly applied to reduce the error between the input and constructed local areas for the unsupervised reconstruction based learning. Since this system operates on local areas of the images, bag-of-words approach with k-means clustering is applied to define the global image descriptor from local areas. Although this system has the same learning strategy as [12], its advantages are twofold compared to [12]. First, model parameters are initialized with greedy layer-wise pre-training that allows more effective learning procedure with respect to the random initialization approach. Second, the CAE model has better capability to characterize the semantic content of images since it considers the neighborhood relationship through the convolution operations.

Reconstruction based unsupervised learning of RS image descriptors is found effective particularly when the annotated training images are not existing. However, minimizing reconstruction loss on a small amount of unlabeled data with a shallow neural network limits the accurate description of the high-level information content of RS images. This problem can be addressed by supervised DL-based CBIR systems that require a training set of a high number of annotated images to learn effective models with several different parameters. The availability and quality of such data determine the feasibility and success of the supervised DL models. However, annotating RS images at large-scale to define sufficiently large high-quality training sets is time consuming, complex, and costly in operational applications. To overcome this problem, a common approach is to exploit DL models with proven architectures (such as ResNet or VGG), which are pre-trained on publicly available general purpose computer vision (CV) datasets (e.g., ImageNet). The existing models are then fine-tuned on a small set of annotated RS images to calibrate the final scales (which is achieved based on data augmentation). In the last step, the system transforms VLAD representations with Memory Vector (MV) construction (which produces the expanded query descriptor) to make the CBIR system sensitive to the selected query images. In this system, the query expansion strategy is applied after obtaining all the local descriptors. This query-sensitive CBIR approach can provide more discriminative image descriptors, since it adapts the overall learning procedure of DNNs based on
Fig. 3: The intuition behind the triplet loss: after training, a positive sample is moved closer to the anchor sample than the negative samples of the other classes.

the selected queries. Thus, it has a huge potential for RS CBIR problems.

Most of the above-mentioned DL based supervised CBIR systems learn an image feature space directly optimized for a classification task by considering entropy-based losses. Thus, the image descriptors are designed to discriminate the pre-defined classes by taking into account the class-based similarities rather than the image-based similarities during the training stage of the DL models. The absence of positive and negative images with respect to the selected query image during the training phase can lead to a poor CBIR performance. To overcome this limitation, metric learning is recently introduced in RS to take into account image similarities within DNNs. Accordingly, a Siamese graph convolutional network is introduced in [20] to model the weighted region adjacency graph (RAG) based image descriptors by a metric learning strategy. To this end, mini-batches are first constructed to include either similar or dissimilar RS images (Siamese pairs). If a pair of images belongs to the same class, they are assumed as similar images, and vice versa. Then, RAGs are fed into two graph convolutional networks with shared parameters to model image similarities with a contrastive loss. Due to the considered metric learning strategy (which is guided by the contrastive loss) the distance between the descriptors of similar images is decreased, while that between dissimilar images is increased. The contrastive loss only considers the similarity estimated among image pairs, i.e., similarities among multiple images are not evaluated, which can limit the success of similarity learning for CBIR problems.

To partially address this limitation, a triplet deep metric learning network (TDMLN) is proposed in [21]. TDMLN employs three CNNs with shared model parameters for the similarity learning through image triplets in the content of metric learning. Model parameters of the TDMLN are initialized with a state-of-the-art CNN model pre-trained on ImageNet. For the mini-batch sampling, TDMLN considers an anchor image together with the similar (i.e., positive) image and dissimilar (i.e., negative) image to the anchor image at a time. Image triplets are constructed based on the annotated training images [20]. While anchor and positive images belong to the same class, negative image is associated to a different class. Then, similarity learning of the triplets is achieved based on the triplet loss. By the use of a triplet loss, the distance estimated between the anchor and positive images in the descriptor (i.e., feature) space is minimized, whereas that computed between the anchor and negative images is separated by a certain margin. Figure 3 illustrates intuition behind the triplet loss. The metric learning guided by a triplet loss learns similarity based on the image triplets and thus provides highly discriminative image descriptors in the framework of CBIR. However, how to define and select image triplets is still an open question. Current methods rely on the image-level annotations based on the land-cover land-use class labels, which do not directly represent the similarity of RS images. Thus, metric learning based CBIR systems need further improvements to characterize retrieval specific image descriptors. One possible way to overcome this limitation can be an identification of image triplets through visual interpretation instead of defining triplets based on the class labels. Tabular overview of the recent DL based CBIR systems in RS is presented in Table 1.

3 Deep Learning for Scalable RS CBIR

Due to the significant growth of RS image archives, an image search and retrieval through linear scan (which exhaustively compares the query image with each image in the archive) is computationally expensive and thus impractical. This problem is also known as large-scale CBIR problem. In large-scale CBIR, the storage of the data is also challenging as RS image contents are often represented in high-dimensional
features. Accordingly, in addition to the scalability problem, the storage of the image descriptors also becomes a critical bottleneck. To address these problems, approximate nearest neighbor (ANN) search has attracted extensive research attention in RS. In particular, hashing based ANN search schemes have become a cutting-edge research topic for large-scale RS image retrieval due to their high efficiency in both storage cost and search/retrieval speed. Hashing methods encode high-dimensional image descriptors into a low-dimensional Hamming space where the image descriptors are represented by binary hash codes. By this way, the (approximate) nearest neighbors among the images can be efficiently identified based on the Hamming distance with simple bit-wise operations. In addition, the binary codes can significantly reduce the amount of memory required for storing the content of images. Traditional hashing-based RS CBIR systems initially extract hand-crafted image descriptors and then generate hash functions that map the original high-dimensional representations into low-dimensional binary codes, such that the similarity to the original space can be well preserved. Thus, descriptor generation and hashing processes are independently applied, resulting in sub-optimal hash codes. Success of DNNs in image feature learning has inspired research on developing DL based hashing methods, which can simultaneously learn the image representation and the hash function with proper loss functions.

Recently, several deep hashing based CBIR systems that simultaneously achieve image representation and hash function learning are introduced (see Table 2). As an example, a supervised deep hashing neural network (DHNN) introduced in [22] learns deep features and binary hash codes by using a pairwise similarity loss and a quantization loss in an end-to-end manner. The pairwise loss can also be considered as a binary cross entropy loss, which is optimized to classify whether an input image pair is similar or not. One advantage of the pairwise loss is its capability of similarity learning, where similar images can be grouped together, while moving away dissimilar images from each other in the feature space. Due to the ill-posed gradient problem, the standard back-propagation for directly optimizing hash codes based on deep models

---

**Table 1: Main characteristics of the DL based CBIR systems in RS.**

| List of Works | Mini-batch Sampling | Network Initialization | DNN Type          | Learning Strategy | Considered Losses |
|---------------|---------------------|-----------------------|-------------------|-------------------|-------------------|
| [12]          | Random selection    | Random initialization  | Auto-encoder      | Reconstruction (unsupervised) | Reconstruction loss |
| [15]          | Data augmentation   | Pre-trained network weights | Convolutional neural network | Classification (supervised) | Cross-entropy loss |
| [13]          | Random selection    | Random initialization  | Convolutional neural network | Clustering (unsupervised) | Reconstruction loss |
| [16]          | Random selection    | Pre-trained network weights | Convolutional neural network | Classification (unsupervised) | Cross-entropy loss |
| [17]          | Random selection    | Pre-trained network weights | Convolutional neural network | Classification (supervised) | Cross-entropy loss |
| [14]          | Random selection    | Greedy layer-wise pre-training | Convolutional auto-encoder | Reconstruction (unsupervised) | Reconstruction loss |
| [18]          | Random Selection    | Pre-trained network weights | Convolutional neural network | Relevance feedback (supervised) | Cross-entropy loss |
| [19]          | Random selection    | Pre-trained network weights | Convolutional neural network | Classification (supervised) | Cross-entropy loss |
| [20]          | Siamese pairs       | Random initialization  | Graph convolutional network | Metric learning (supervised) | Contrastive loss |
| [21]          | Image triplets      | Pre-trained network weights | Convolutional neural network | Metric learning (supervised) | Triplet loss |
TABLE 2: Main characteristics of the state-of-the-art deep hashing based CBIR systems in RS.

| List of Works | Considered Losses | Type of Learning | Hash Layer |
|---------------|-------------------|-----------------|------------|
| [22]          | Pairwise loss, Quantization loss | supervised | linear |
| [23]          | Pairwise loss, Quantization loss | supervised | linear |
| [24]          | Triplet loss, Bit balance loss, Quantization loss | supervised | sigmoid |
| [25]          | Pairwise loss, Quantization loss, Cross entropy loss | supervised | linear |
| [26]          | Cross entropy loss, Pairwise loss, Reconstruction loss, Quantization loss, Bit balance loss | semi-supervised | linear |
| [27]          | Adversarial loss, Quantization loss, Pairwise loss, Cross entropy loss | supervised | sigmoid |
| [28]          | Graph loss, Nonlinear embedding loss, Quantization loss, Augmentation invariant loss | unsupervised | linear |

is infeasible. The use of the quantization loss mitigates the performance degradation of the generated hash codes through the binarization on the CNN outputs. As a follow-up work in [23] these two loss terms are exploited in source-invariant deep hashing CNNs for learning a cross-modality hashing system. Without introducing a margin threshold between the similar and dissimilar images, a high image retrieval performance based on the pairwise loss can not be achieved. To address this issue, a supervised deep hashing method, called as Metric-Learning based Deep Hashing Network (MiLaN), is recently introduced in [24]. MiLaN is trained by using three different losses: 1) a triplet loss for learning a metric space (where semantically similar images are close to each other and dissimilar images are separated); 2) a bit balance loss (which aims at forcing the hash codes to have a balanced number of binary values); and 3) a quantization loss.

As noted in [24], the learned hash codes based on the triplet loss can efficiently characterize the complex semantics in RS images with a small number of annotated training images. The bit balance loss makes each bit of hash codes to have a 50% chance of being activated, and different bits to be independent from each other. A supervised deep hashing CNN (DHCNN) is proposed in [25] in order to simultaneously retrieve the semantically similar images in an end-to-end manner. In detail, DHCNN utilizes a joint loss function composed of: 1) a pairwise loss; 2) a cross entropy loss (which aims at increasing the class discrimination capability of hash codes); and 3) a quantization loss. In order to predict the classes based on hash codes, a FC layer is connected to the hash layer in DHCNN. As mentioned above, one disadvantage of the cross entropy loss is its deficiency to define a metric space, where similar images are clustered together. To address this issue, the pairwise loss is jointly optimized with the cross entropy loss in DHCNN. A semi-supervised deep hashing method based on the adversarial autoencoder network (SSHAAE) is proposed in [26] for RS CBIR problems. In order to generate the discriminative and similarity preserved hash codes with low quantization errors, SSHAAE exploits a joint loss function composed of: 1) a cross entropy loss; 2) a reconstruction loss; 3) a pairwise loss; 4) a bit balance loss; and 5) a quantization loss. By minimizing the reconstruction loss, the label vectors and hash codes can be obtained as the latent outputs of the AEs. A supervised deep hashing method in the framework of generative adversarial networks is proposed in [27]. To sufficiently learn the compact hash codes, a joint loss function is introduced, which is composed of: 1) a cross entropy loss; 2) a pairwise loss; 3) an adversarial loss (which aims at enforcing the bit balance property of the produced hash codes); and 4) a quantization loss. The adversarial loss is trained in a min-max manner to restrict the learned hash codes following the uniform binary distribution so that the bit balance capability of hash codes can be achieved. The above-mentioned losses tend to preserve the discriminative capability and the semantic similarity of the hash codes in the Hamming space based on the image annotations. However, without such supervised information, it is challenging to explicitly preserve these characteristics in the Hamming space. To address this problem, we recently propose an Unsupervised Deep Hashing method driven by the Graph structure (UDHG) [28] that utilizes a joint loss function made up of: 1) a graph regularization term for producing hash codes (which can preserve the topology of the images in the feature space); 2) a nonlinear embedding term (which learns the hash codes based on a
nonlinear hash function; 3) an augmentation invariant term (which makes the learned hash codes robust to the data augmentation processing); and 4) a quantization term.

TABLE 3: Comparison of the DL loss functions considered within the deep hashing based RS CBIR systems. Different marks are provided: “✓” (no) or ‘√’ (yes).

| Loss functions | Capability of similarity learning | Requirement on mini-batch sampling | Capability of bit balancing | Capability of binarization | Requirement on image annotations |
|----------------|-----------------------------------|-----------------------------------|-----------------------------|----------------------------|--------------------------------|
| Pairwise       | ✓                                 | pairs of images                   | x                           | x                          | ✓                              |
| Triplet        | ✓                                 | triplets of images                | x                           | x                          | ✓                              |
| Adversarial    | x                                 | randomly selected images          | ✓                           | x                          | x                              |
| Reconstruction | x                                 | randomly selected images          | x                           | x                          | x                              |
| Cross entropy  | x                                 | randomly selected images          | x                           | x                          | ✓                              |
| Bit balance    | x                                 | randomly selected images          | ✓                           | x                          | x                              |
| Quantization   | x                                 | randomly selected images          | x                           | ✓                          | x                              |
| Graph          | ✓                                 | subgraphs of images               | x                           | x                          | x                              |

In Table 3, we analyze and compare all the above-mentioned loss functions based on their: i) capability on similarity learning, ii) requirement on the mini-batch sampling; iii) capability of assessing the bit balance issues; iv) capability of binarization of the image descriptors; and v) requirement on the annotated images. For instance, the pairwise, triplet and graph losses have the capabilities to learn the relationship among the images in the feature space, where the semantic similarity of hash codes can be preserved. Regarding to the requirement of mini-batch sampling, pairs of images should be sampled for the pairwise loss, image triplets should be constructed for the triplet loss, and the subgraphs of images should be provided for the graph loss functions. The bit balance and adversarial loss functions are exploited for learning the hash codes with the uniform binary distribution. It is worth noting that the adversarial loss can be also exploited for other purposes, such as for image augmentation problems to avoid the overfitting issue [29]. The quantization loss enforces the produced low-dimensional features by the CNN models to approximate the binary hash codes. With regard to the requirement on image annotations, the pairwise and triplet losses require the semantic labels to construct the relationships among the images.

4 DISCUSSION AND CONCLUSION

In this chapter, we presented a literature survey on the most recent CBIR systems for efficient and accurate search and retrieval of RS images from massive archives. We focused our attention on the DL based CBIR systems in RS. We initially analyzed the recent DL based CBIR systems based on: i) the strategies considered for the mini-batch sampling; ii) the approaches used for the initialization of the parameters of the considered DNN models; iii) the type of the considered DNNs; and iv) the strategies used for image representation learning. Then, the most recent methodological developments in RS related to scalable image search and retrieval were discussed. In particular, we reviewed the deep hashing based CBIR systems and analyzed the loss functions considered within these systems based on their: i) capability of similarity learning, ii) requirement on the mini-batch sampling; iii) capability of assessing the bit balance issues; iv) capability of binarization; and v) requirement on the annotated images. Analysis of the loss functions under these factors provides a guideline to select the most appropriate loss function for large-scale RS CBIR. It is worth emphasizing that developing accurate and scalable CBIR systems is becoming more and more important due to the increased number of images in the RS data archives. In this context, the CBIR systems discussed in this chapter are very promising.

Despite the promising developments discussed in this chapter (e.g., metric learning, local feature aggregation and graph learning), it is still necessary to develop more advanced CBIR systems. For example,
most of the systems are based on the direct use of the CNNs for the retrieval tasks, whereas the adapted CNNs are mainly designed for learning a classification problem and thus model the discrimination of pre-defined classes. Thus, the image descriptors obtained through these networks can not learn a image feature space directly optimized for the retrieval problems. Siamese and triplet networks are defined in the context of metric learning in RS to address this problem. However, the image similarity information to train these networks is still provided based on the pre-defined classes, limiting to extract retrieval specific image descriptors. Thus, CBIR systems that can efficiently learn image features optimized for retrieval problems are needed. Furthermore, the existing supervised DL based CBIR systems require a balanced and complete training set with annotated image pairs or triplets, which is difficult to gather in RS. Learning an accurate CBIR model from imbalanced and incomplete training data is very crucial and thus there is a need for developing systems addressing this problem for operational CBIR applications. Furthermore, the availability of increased number of multi-source RS images (multispectral, hyperspectral and SAR) associated to the same geographical area motivates the need for effective CBIR systems, which can extract and exploit multi-source the image descriptors to define rich characterization of RS images (and thus to improve image retrieval performance). However, multi-source RS CBIR has not been explored yet (i.e., all the deep hashing based CBIR systems are defined for images acquired by single sensors). Thus, it is necessary to study CBIR systems that can mitigate the aforementioned problems.

REFERENCES

[1] Y. Yang and S. Newsam, “Geographic image retrieval using local invariant features,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 2, pp. 818–832, Feb 2013.
[2] E. Aptoula, “Remote sensing image retrieval with global morphological texture descriptors,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 5, pp. 3023–3034, May 2014.
[3] I. Tekeste and B. Demir, “Advanced local binary patterns for remote sensing image retrieval,” in IEEE Intl. Geosci. Remote Sens. Symp., July 2018, pp. 6855–6858.
[4] O. E. Dai, B. Demir, B. Sankur, and L. Bruzzone, “A novel system for content-based retrieval of single and multi-label high-dimensional remote sensing images,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 11, no. 7, pp. 2473–2490, July 2018.
[5] Y. Li and T. R. Bretschneider, “Semantic-sensitive satellite image retrieval,” IEEE Trans. Geosci. Remote Sens., vol. 45, no. 4, pp. 853–860, April 2007.
[6] B. Chaudhuri, B. Demir, L. Bruzzone, and S. Chaudhuri, “Region-based retrieval of remote sensing images using an unsupervised graph-theoretic approach,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 7, pp. 987–991, July 2016.
[7] B. Demir and L. Bruzzone, “Hashing-based scalable remote sensing image search and retrieval in large archives,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 2, pp. 892–904, Feb 2016.
[8] 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, March 2017.
[9] B. Demir and L. Bruzzone, “A novel active learning method in relevance feedback for content-based remote sensing image retrieval,” IEEE Trans. Geosci. Remote Sens., vol. 53, no. 5, pp. 2323–2334, May 2015.
[10] Y. Wang, L. Zhang, X. Tong, L. Zhang, Z. Zhang, H. Liu, X. Xing, and P. T. Mathiopoulos, “A three-layered graph-based learning approach for remote sensing image retrieval,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 10, pp. 6020–6034, Oct 2016.
[11] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, “Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 2, pp. 1144–1158, Feb 2018.
[12] W. Zhou, Z. Shao, C. Diao, and Q. Cheng, “High-resolution remote-sensing imagery retrieval using sparse features by auto-encoder,” Remote Sens. Lett., vol. 6, no. 10, pp. 775–783, 2015.
[13] Y. Li, Y. Zhang, C. Tao, and H. Zhu, “Content-based high-resolution remote sensing image retrieval via unsupervised feature learning and collaborative affinity metric fusion,” Remote Sens., vol. 8, no. 9, 2016.
[14] X. Tang, X. Zhang, F. Liu, and L. Jiao, “Unsupervised deep feature learning for remote sensing image retrieval,” Remote Sens., vol. 10, no. 8, 2018.
[15] F. Hu, X. Tong, G. Xia, and L. Zhang, “Delving into deep representations for remote sensing image retrieval,” in IEEE Int. Conf. Signal Process., Nov 2016, pp. 198–203.
[16] W. Zhou, S. Newsam, C. Li, and Z. Shao, “Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval,” Remote Sens., vol. 9, no. 5, 2017.
[17] F. Ye, H. Xiao, X. Zhao, M. Dong, W. Luo, and W. Min, “Remote sensing image retrieval using convolutional neural network features and weighted distance,” IEEE Geosci. Remote Sens. Lett., vol. 15, no. 10, pp. 1535–1539, Oct 2018.

[18] Y. Boualleg and M. Farah, “Enhanced interactive remote sensing image retrieval with scene classification convolutional neural networks model,” in IEEE Intl. Geosci. Remote Sens. Symp., July 2018, pp. 4748–4751.

[19] R. Imbriaco, C. Sebastian, E. Bondarev, and P. H. N. de With, “Aggregated deep local features for remote sensing image retrieval,” Remote Sens., vol. 11, p. 493, 2019.

[20] U. Chaudhuri, B. Banerjee, and A. Bhattacharya, “Siamese graph convolutional network for content based remote sensing image retrieval,” Comput. Vis. Image Understand., vol. 184, pp. 22–30, 2019.

[21] R. Cao, Q. Zhang, J. Zhu, Q. Li, Q. Li, B. Liu, and G. Qiu, “Enhancing remote sensing image retrieval using a triplet deep metric learning network,” Int. J. Remote Sens., pp. 1–12, 2019.

[22] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, “Large-scale remote sensing image retrieval by deep hashing neural networks,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 2, pp. 950–965, Feb 2018.

[23] Y. Li, Y. Zhang, X. Huang, and J. Ma, “Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 11, pp. 6521–6536, 2018.

[24] S. Roy, E. Sangineto, B. Demir, and N. Sebe, “Metric-learning based deep hashing network for content based retrieval of remote sensing images,” arXiv preprint arXiv:1904.01258, 2019.

[25] W. Song, S. Li, and J. Benediktsson, “Deep hashing learning for visual and semantic retrieval of remote sensing images,” arXiv preprint arXiv:1909.04614, 2019.

[26] X. Tang, C. Liu, X. Zhang, J. Ma, C. Jiao, and L. Jiao, “Remote sensing image retrieval based on semi-supervised deep hashing learning,” in IEEE Intl. Geosci. Remote Sens. Symp., July 2019, pp. 879–882.

[27] C. Liu, J. Ma, X. Tang, X. Zhang, and L. Jiao, “Adversarial hash-code learning for remote sensing image retrieval,” in IEEE Intl. Geosci. Remote Sens. Symp., July 2019, pp. 4324–4327.

[28] J. Kang and B. Demir, “Graph structure driven unsupervised deep hashing,” in IEEE Conf. Comput. Vis. Pattern Recog., in review, 2020.

[29] Y. Cao, B. Liu, M. Long, and J. Wang, “Hashgan: Deep learning to hash with pair conditional wasserstein gan,” in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 1287–1296.