Satellite Image Analysis: A Review

Jatin Babbar
United States Technologies International Private Limited
Trivandrum, Kerala, India
jatinbabbar1696@gmail.com

Neeru Rathee
Maharaja Surajmal Institute of Technology
New Delhi, India
neeru1rathee@gmail.com

***Abstract*—Satellite image processing is an important area of research now a days due to its wide range of applications. Researchers and scientists have paid attention to satellite image processing so as to capture information from them. Satellite image analysis poses a great challenge to the researchers due to high variability, low resolution and big data of the satellite images. A lot of work has been done for satellite image analysis that covers the research from classification of hand crafted features to applying high performance computing on satellite images. The researchers have achieved great success in satellite image analysis. But a systematic review, which will lead researchers to identify the problem and to contribute in this field, is missing. In the presented chapter, an attempt has been made to present a detailed review of the various steps of satellite image processing, classification and available databases. This chapter will give an impetus towards further research in this field and will provide a baseline to research in the field of satellite image processing.**

Keywords—Satellite image processing · Satellite image Analysis · Satellite image Classification · Review.

I. Introduction

Remote sensing has contributed to the creation of huge datasets of satellite and airborne imagery. However, these datasets had very high resolution images which required fast processors and powerful GPUs for computation. Moreover, the image processing techniques present at the time were not enhanced enough to deal with or be used for high resolution satellite and airborne imagery.

The past few decades have contributed a lot to the advancements in the field of image processing. These advancements and new innovative technologies have made it possible to work on huge datasets of high resolution images like satellite image dataset.

This paper talks about some of the most popular image processing techniques and some image classification methods that are used for working on satellite images. The first section, satellite image processing, talks about one of the most popular image processing techniques, image segmentation, which is the process of segmenting the image into a number of 'objects' which helps to extract the region of interest from the entire image and helps to remove the background noise.

The next section of this paper talks about the image classification techniques that are used for high resolution satellite images. These techniques are used to label the images based on their classes or the objects present between them. These techniques are really helpful in Land Use Land Cover (LULC) study.

II. Satellite Image Processing

Image processing refers to the methods and operation that we use in order to extract useful information and various features from an image. These operations tend to enhance or modify the properties of the image so as to achieve better results after classification. Various mathematical algorithms are used to perform operations on images to process them for extraction of information. One of the most popular techniques in image processing is image segmentation.

In general aspects, Image segmentation is defined as the process of splitting an image into homogeneous regions in such a way that every region, when considered individually, is homogeneous [1]. Any two adjacent regions cannot be homogeneous. Image segmentation is one of the most necessitous steps in image processing.

One of the most important uses of image segmentation is to divide the image into a number of homogeneous entities that can later be classified into different categories and the region of interest or the entity under observation can be extracted from the rest of the image [2]. This way, the image gets easier to process further. Segmentation is also useful in Image Compression and Image Analysis [3].

Image segmentation can be accomplished in a number of ways. Each approach is affected differently by various factors [4]. The following section reviews about some of the techniques and approaches used for image segmentation.

Based on the approach, image segmentation can be classified into the following categories:

- Region Based
- Edge Based
- Threshold Based

A. Region Based Image Segmentation

In region based segmentation approach, the image is divided into different segments based on the homogeneity of the regions around a pixel. In this technique, adjacent pixels are compared to each other to find the similarities in the intensity and color values so as to group them together accordingly. If the pixel values and properties are similar, the two pixels are grouped into a single class, else, the pixel is compared with another pixel.

One of the most common algorithms for carrying out region based image segmentation technique is the watershed algorithm. A number of techniques have been put forth to compute watersheds [5,6]. The basic idea for building this model is using a geographical analogy. First, begin by entering the local minima of the surface [7]. Then gradually submerge the image into a lake. The water increasingly floods the basins based on all the minima. To stop the two different waters from different minima from merging, we put up a dam between the lines. When the surface is totally submerged, the various dams thus built give the watershed of our original image.

Watershed algorithms falls under the category of morphological transformation techniques. This model uses the concepts of mathematical morphology [2] to segment images into homogeneous blocks [8]. Using this algorithm we obtain the homogeneous regions and eventually our region of interest in the image. However, this algorithms sometimes causes over-segmentation as noise and some irregularities in the image will still be there. The result of this is that regions with different labels get mixed together to form a single region.

There is an extension to the watershed algorithm, which is known as the marker-based watershed algorithm. In this algorithm, the points where the 'flooding' will start are explicitly defined. This is used to avoid the case of over-segmentation.

B. Edge detection based techniques

Edge detection segmentation techniques are somewhat more complex to compute in comparison to region based segmentation technique. To reduce the complexity, the input is first converted to a gray colored image before feeding it to the segmentation model. Also, when dealing with high resolution images, there is a high probability of presence of noise which needs to be removed before detecting the edges of the objects in the image.

There are various techniques used for detecting edges. The basic underlying approach is detecting the local changes in the pixel intensity, which depicts the boundary between two regions. The result of this technique gives the discontinuities in the image and helps to extract the area of interest from the background. In contrast to the region based segmentation techniques, where weak boundaries are 'submerged' and two different regions combine to form a single one, edge detection is able to extract weak boundaries as well.

The technique deals with maneuvering an image to obtain another image, removing redundant data and transforming a 2-D pixel array into an uncorrelated data set. Edges act as the local attribute of a pixel and its contiguous neighborhood and represents the boundaries. Hence, detecting Edges helps to extract useful characteristics of the image where there are unexpected changes [9].

1. Sobel detection

In Sobel detection, we use filters that give separate results for horizontal edges and vertical edges, which are then put together to form the resulting detected edges. The Sobel technique implements a 2 dimensional gradient quantity on an image to highlight regions that resemble edges [10].

2. Roberts edge detection

The Roberts edge detection is similar to Sobel edge detection. It comprises of 2 different kernels that give separate results for edges at different orientations, which are then put together to form the resulting detected edges. The main difference between the design of Sobel and Robert is that the kernels that each uses to obtain the image is different. The sobel kernel is more capable in detecting edges along the horizontal and vertical axis. The Roberts's kernel, on the other hand, detects them along the angles of 45○ and 135○ to the vertical axis [11].

C. Threshold based technique

Thresholding is one of the simplest and yet most important parts of image processing and analysis. The purpose of thresholding is to actually divide the image into two categories, which are the foreground, that is the region of interest, and the background. This process is carried out on a grayscale image. An appropriate value of threshold is found out, which is then used to turn all the pixel values into one of the two categories, which are zero and one. If the pixel value is smaller than the threshold value, it will be changed to zero, and if the pixel value is more than the threshold value, it will be changed to one.

One of the most popular thresholding techniques is Otsu's Binarization Thresholding. In image processing, Otsu's method is used as an automatic clustering-based image thresholding technique [12] or to reduce a grayscale image to a binary image. The algorithm theorizes that the image consists of two classes of pixels, the foreground and the background. Then it calculates the optimum value of threshold by separating the two classes such that the combined variance is minimum, and inter-class variance is maximum [13].

III. Satellite Image Classification

There are several methods and techniques used for the classification of satellite images. Different approaches have been adopted to extract and learn the features of the image during the processing/analysis part. Based on the approach that takes into account the spatial resolution of the desired region of interest, image classification can be divided into two broad categories:

- Pixel based classification

- Object based classification

Pixel based classification

During the early 70s, when there was low spatial resolution of the images, the various techniques and methods that were carried out were based on the per-pixel analysis [14]. This basically involved analyzing every pixel in the desired region of interest without considering other factors like orientation or other contextual information about the region of interest.

In pixel-based classification, different classification algorithms have been used throughout the past. The main focus of these classification techniques is to learn the features of the images and make predictions based on those features. In some cases, we have the samples which we can use to train the classifier models and then make predictions based on the features learned using the samples. In other cases, the machine learning model is supposed to learn the features of the input data (in this case, the image) on its own using analysis and various clustering and grouping techniques. Based on the learning approach of the features of the image, these techniques can be classified into the following categories:

- Supervised or handcrafted feature learning-based techniques

- Unsupervised feature learning-based techniques.

D. Supervised or Handcrafted feature learning-based techniques

In supervised or handcrafted based techniques, there is a requirement of a dataset which is pre-labeled. This dataset is known as the training dataset. This dataset is used to train the model which will be used as the classifier. The performance of this classifier depends majorly on the training dataset. This means that the samples collected to train determine how to classifier will work on the new dataset. The samples included in the dataset should have high diversity and variance in terms of the features that each data sample (in this case, an image) might have.

There are a number of methods used for supervised or handcrafted feature-based learning techniques. We review some of these methods, including Histogram of Oriented Gradients (HOG), Scale Invariant Feature Transform (SIFT), and Color Histograms.

1. Histogram of Oriented Gradient (HOG)

The Histogram of Oriented Gradient computes the distribution of the intensities of gradients and the edge directions of objects in the regions of the image. Basically, the image is divided into different regions called cells, and the pixels in the cells are used to obtain the histograms of local gradient intensities and edge orientations. These orientations are then used to calculate the image gradient vectors which are converted to angles. This technique is considered to be one of the most effective ways to learn features about the shape of objects in an image. Multiple cells are grouped together to form blocks which are used to calculate the normalized histograms for all the cells in the block.

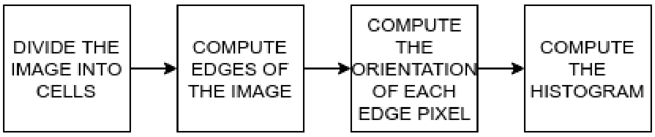


Fig 1: HOG

To compute the gradient values, derivative mask is applied to the horizontal and/or vertical directions. To do this, two filter kernels are required to filter the intensity data of the pixels in the image. The two filter colors are represented as:

$$[-1\ 0\ 1]\ ,\ [-1\ 0\ 1]^T$$

Based on these gradient computations, an orientation-based histogram channel is obtained.

2. Scale Invariant Feature Transform

The SIFT [15] technique is somewhat similar to HOG. In SIFT, the gradient information in the regions/blocks is computed around identified key points. These key points are identified from a large set of key points. These key pointscan be extracted in large amounts from an image, which makes the technique robust to extract features like small objects from clutter and noise [15].

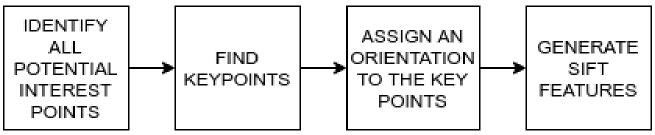


Fig 2: SIFT

3. Color Histograms

Color histogram feature learning is one the most basic techniques in image feature extraction. In this technique, a histogram is created for the different colors in the image in different bins, where the bars represent the number of pixels corresponding to each color. Even small illumination changes can be captured using color histograms. However, objects of same or similar colors are hard to distinguish [11]. The disadvantage of color histograms is that the shape and texture of the object are also ignored. In most images, color histogram is able to distinguish the foreground (region of interest) from the background.

E. Unsupervised feature learning-based techniques

These techniques use various clustering and grouping techniques to group images with similar properties and features and learn from them. In unsupervised feature learning techniques, there is no need of a training dataset, as the features are learned using the images and not the labeled sample data inputs. One of the most commonly used unsupervised feature learning-based techniques is k-means clustering.

1. K-Means Clustering

K-Means Clustering is a vector quantization method, which aims to break up a set of data points/observations into a number of clusters based on their similarities. The number of clusters obtained is pre-defined and is arbitrary. The observations in one cluster are more comparable to one another than they are to observations in other clusters.

This technique is carried out when the data items do not have a label. K-Means clustering, broadly speaking, involves two steps:

- assigning a data item to the cluster with the centroid closest to it.

- recalculating the centroid of the cluster after adding the new data item.

When no centroids are getting updated, the convergence of the algorithm is achieved. The number of such centroids is predetermined, which signifies the count of clusters the user needs.

IV. Database for Satellite Image Analysis

The database play an important role in research by providing a standard to evaluate the algorithms proposed or developed by researchers. For satellite image analysis, many agencies are working on providing the database for different applications and contribute towards algorithm development in this field. The important databases are described in this section.

F. UC Merced Land Use Dataset

This database consist of images of diversified urban areas from the USGS National Map Urban Area Imagery collection. There are a total of 21 classes depicting different means of land use [16]. There are 100 images of each in the database resulting in 2100 images having pixel resolution of 1 foot and image size 256x256. The dataset is produced by manual cropping of the original images from USGS. This dataset is the most popular database used in scene classification [17,18,19] and information retrieval [20,21,22].

G. WHU-RS19 Dataset

This dataset is very small database and it was originally developed by extracting the images [23] from a bunch of satellite imagery from Google Earth and later on modified by combining various features [24]. There are 19 different types of scenes in this dataset. The size of the images is 600600 and every class contain 50 images and 1005 images in all. The high variation in orientation, different scaling, varying illumination and resolution extend a challenge to the researchers working in the field of satellite image processing and making it popular despite of its small size [25, 26, 27, 28].

H. SIRI-WHU Dataset

The SIRI-WHU dataset [29] contains images extracted from Google Earth images by Remote sensing group from Wuhan University by applying Intelligent Data Extraction and Analysis techniques. The size of the images was 200x200 pixels with spatial resolution of 2m. The dataset consists of 12 classes of different scenes mainly of China region. There are 200 images per class resulting in 2400 total images. The database has been used by researchers [30, 31, 32, 33] for evaluation of their algorithms but it is not so popular due to small size and lack of diversity.

I. RSSCN7

This database [19] is also collected from Google Earth database. These images were cropped on 4 different scales, which is the reason for challenging nature of this database. There are only seven classes of scenes in this database. Each class consists of 100 images on 4 different scales, thus resulting in 400 images for each scene and total images are 2800.

J. RSC11 Dataset

This dataset [19] is also extracted from Google Earth images. The size of the images is 512x512 and the spatial resolution is 0.2 m. The dataset is very challenging d because there are complicated scenes with high similarity between them. There are 11 different scenes from different cities such as Washington DC, L.A., San Francisco, New York, etc. The dataset is explored by the many researchers [34, 35, 36].

K. Brazilian Coffee Scene Dataset

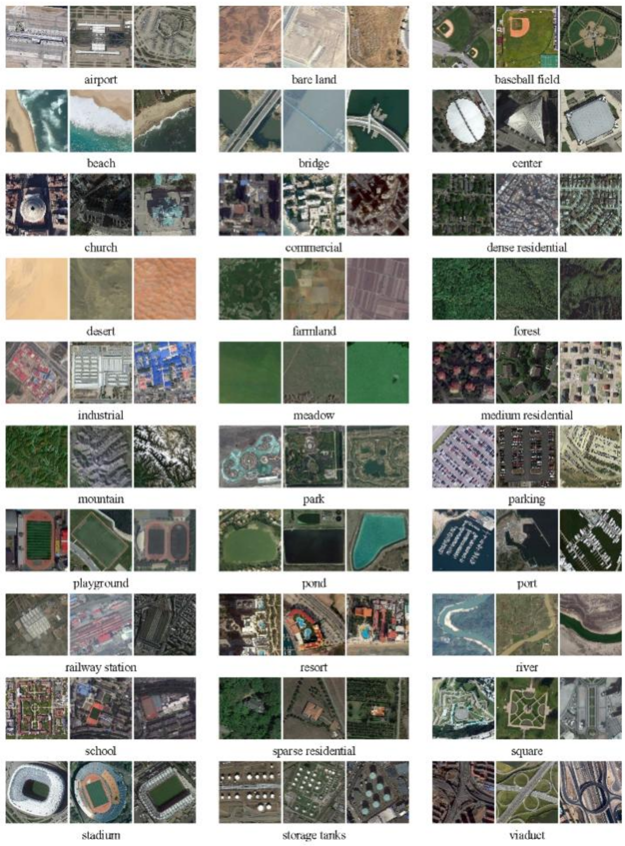
This dataset [37] is popularly used for identification of coffee crop in various countries. The size of images is very small i.e. 64x64 as they were extracted from SPOT satellite images. There are 2876 images in total (1438 images in one class). This dataset explored various bands such as green, red and near-infrared so as to have better clarity of vegetation area. the labelling of the images was done manually by agricultural researchers by considering the images with greater than 85% pixels as coffee images and rest as non-coffee images. Though researchers have evaluated their algorithms using this database [38, 39, 40] but still it is not so popular due to limited number of classes.

L. Aerial Image Dataset (AID)

Aerial image dataset [41] is created by using images from Google earth images. The idea behind using Google Earth images is the proof that Google Earth images are quite similar to the real opticalaerial images at the pixel level for detection of Land Cover/Use mapping. This dataset describes 30 aerial scenes such as railway station, airport, resort, river, farmland, forest, industrial, meadow, school, storage tanks and viaduct and consists of 10000 images in total. The labelling of the images is done manually by the experts. The advantage of this dataset is that it contains multi source images rather than single source. Another advantage of the dataset is that it contain images with high intra class diversities as the images are acquired at different times, in different imaging and atmospheric conditions and include various countries such as China, England, France, Italy, Japan, Germany, etc. A The sample of the images is shown in the figure 2.

M. NWPU-RESISC45

NWPU-RESISC45 dataset [15] has been recently proposed and is very large in size (31500 images) making it suitable for evaluation deep learning algorithms. There are 45 different scenes in the dataset. The images are extracted from Google Earth images and have size 256x256. The images in this dataset vary largely in spatial resolution



(varying from 0 m to 0.2 m per pixel), region (covers nearly 100 countries), image variation (covering 45 scenes), high interclass variation and low intraclass variation. The sample of the images is shown in figure 3. The deep learning algorithms have been well evaluated on this dataset [42, 43, 44].

Fig 3: Sample images from AID database [41]

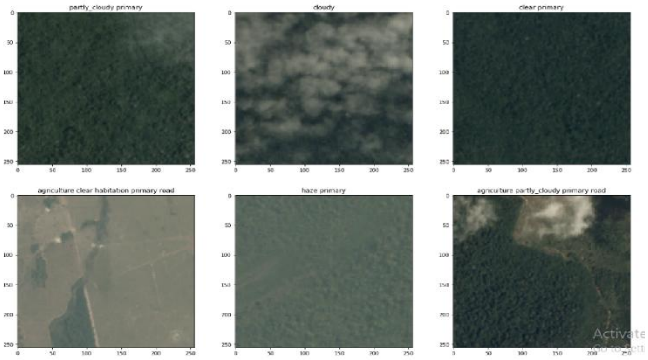


Fig 4: Sample of NWPU-RESISC45 dataset [15]

REFERENCES

[1] , Zhang, Y., Zhong, M.: A review on image segmentation techniques with remote sensing perspective (2010)

[2] Gonzalez, Rafael C., and Richard E. Woods. "Image processing." Digital image processing 2 (2007).

[3] Yogamangalam, R., Karthikeyan, B.: Segmentation techniques comparison in image processing. International Journal of Engineering and Technology (IJET) 5(1), 307–313 (2013)

[4] Bins, L.S., Fonseca, L.G., Erthal, G.J., Ii, F.M.: Satellite imagery segmentation: a region growing approach. Simp´osio Brasileiro de Sensoriamento Remoto 8(1996), 677–680 (1996)

[5] Beucher, Serge. "Use of watersheds in contour detection." In Proceedings of the International Workshop on Image Processing. CCETT, (1979)

[6] Najman L, Schmitt M. Geodesic saliency of watershed contours and hierarchical segmentation. IEEE Transactions on pattern analysis and machine intelligence.1163-73. (1996)

[7] Serra, Jean. "Image segmentation or image understanding?." Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82.. Vol. 7. IEEE, 1982.

[8] Vincent, L., Soille, P.: Watersheds in digital spaces: an efficient algorithm based on immersion simulations. IEEE Transactions on Pattern Analysis & Machine Intelligence (6), 583–598 (1991)

[9] Vincent, O. Rebecca, and Olusegun Folorunso. "A descriptive algorithm for sobel image edge detection." In Proceedings of Informing Science & IT Education Conference (InSITE), vol. 40, pp. 97-107.

[10] Muthukrishnan, R., Radha, M.: Edge detection techniques for image segmentation. International Journal of Computer Science & Information Technology 3(6), 259 (2011)

[11] Lakshmi, S., Sankaranarayanan, D.V., et al.: A study of edge detection techniques for segmentation computing approaches. IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT pp. 35–40 (2010)

[12] Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. Journal of Electronic imaging.(1):146-66. (2004)

[13] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." IEEE transactions on systems, man, and cybernetics 9, no. 1 (1979): 62-66.

[14] Cheng, G., Han, J., Lu, X.: Remote sensing image scene classification: benchmark and state of the art. Proceedings of the IEEE 105(10), 1865–1883 (2017)

[15] Lowe, David G.: Distinctive image features from scale-invariant key points, International journal of computer vision 60.2 : 91-110. (2004)

[16] Yang Y, Newsam S. "Bag-of-visual-words and spatial extensions for land-use classification". InProceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems (pp. 270-279). ACM (2010)

[17] Wu, H., Liu, B., Su, W., Zhang, W., Sun, J.: Hierarchical coding vectors for scene level land-use classification. Remote Sensing 8(5), 436 (2016)

[18] Liu, Y., Zhang, Y.M., Zhang, X.Y., Liu, C.L.: Adaptive spatial pooling for image classification. Pattern Recognition 55, 58–67 (2016)

[19] Zou, Q., Ni, L., Zhang, T., Wang, Q.: Deep learning based feature selection for remote sensing scene classification. IEEE Geosci. Remote Sensing Lett. 12(11), 2321–2325 (2015)

[20] Xia, G.S., Wang, Z., Xiong, C., Zhang, L.: Accurate annotation of remote sensing images via active spectral clustering with little expert knowledge. Remote Sensing 7(11), 15014–15045 (2015)

[21] Hu, J., Xia, G.S., Hu, F., Zhang, L.: A comparative study of sampling analysis in the scene classification of optical high-spatial resolution remote sensing imagery. Remote Sensing 7(11), 14988–15013 (2015)

[22] Zhang, J., Li, T., Lu, X., Cheng, Z.: Semantic classification of high-resolution remote-sensing images based on mid-level features. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9(6), 2343–2353 (2016)

[23] Xia, G.S., Yang, W., Delon, J., Gousseau, Y., Sun, H., Maître, H.: Structural high-resolution satellite image indexing. In: ISPRS TC VII Symposium-100 Years ISPRS. vol. 38, pp. 298–303 (2010)

[24] Sheng, G., Yang, W., Xu, T., Sun, H.: High-resolution satellite scene classification using a sparse coding based multiple feature combination. International journal of remote sensing 33(8), 2395–2412 (2012)

[25] Huang, L., Chen, C., Li, W., Du, Q.: Remote sensing image scene classification using multi-scale completed local binary patterns and fisher vectors. Remote Sensing 8(6), 483 (2016)

[26] Zou, J., Li, W., Chen, C., Du, Q.: Scene classification using local and global features with collaborative representation fusion. Information Sciences 348, 209–226 (2016)

[27] Hu, F., Xia, G.S., Wang, Z., Huang, X., Zhang, L., Sun, H.: Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(5) (2015)

[28] Zhao, L., Tang, P., Huo, L.: Feature significance-based multibag-of-visual-words model for remote sensing image scene classification. Journal of Applied Remote Sensing 10(3), 035004 (2016)

[29] Zhao, B., Zhong, Y., Xia, G.S., Zhang, L.: Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing 54(4), 2108–2123 (2016)

[30] Zhu, Q., Zhong, Y., Zhao, B., Xia, G.S., Zhang, L.: Bag-of-visual-words scene classifier with local and global features for high spatial resolution remote sensing imagery. IEEE Geoscience and Remote Sensing Letters 13(6), 747–751 (2016)

[31] Liu, Y., Zhong, Y., Fei, F., Zhang, L.: Scene semantic classification based on random-scale stretched convolutional neural network for high-spatial resolution remote sensing imagery. In: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. pp. 763–766. IEEE (2016)

[32] Han, X., Zhong, Y., Cao, L., Zhang, L.: Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. Remote Sensing 9(8), 848 (2017)

[33] Yuan, L.X., Li, S.J., Jiang, Y.P.: Remote sensing scene classification using a pre-classification strategy and an improved structural feature. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10(9), 4094–4103 (2017)

[34] Ebied, H., Al-Berry, M., Tolba, M.: Deep learning for satellite image classification. In: Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018. p. 383. Springer (2019)

[35] Birdi, P.K., Kale, K.: Deep convolutional neural network models for land use and land cover identification using dataset created from liss-iv satellite images (2018)

[36] Liu, G., Li, S., Shao, Y.: Low-quality and multi-target detection in rsis (2016)

[37] Penatti, O.A., Nogueira, K., dos Santos, J.A.: Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 44–51 (2015)

[38] Nogueira, K., Penatti, O.A., dos Santos, J.A.: Towards better exploiting convolutional neural networks for remote sensing scene classification. Pattern Recognition 61, 539–556 (2017)

[39] Nogueira, K., Miranda, W.O., Dos Santos, J.A.: Improving spatial feature representation from aerial scenes by using convolutional networks. In: Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on. pp. 289–296. IEEE (2015)

[40] Wang, G., Fan, B., Xiang, S., Pan, C.: Aggregating rich hierarchical features for scene classification in remote sensing imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10(9), 4104–4115 (2017)

[41] Xia, G.S., Hu, J., Hu, F., Shi, B., Bai, X., Zhong, Y., Zhang, L., Lu, X.: Aid: A benchmark data set for performance evaluation of aerial scene classification. IEEE Transactions on Geoscience and Remote Sensing 55(7), 3965–3981 (2017)

[42] Zhu, X.X., Tuia, D., Mou, L., Xia, G.S., Zhang, L., Xu, F., Fraundorfer, F.: Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geoscience and Remote Sensing Magazine 5(4), 8–36 (2017)

[43] Cheng, G., Yang, C., Yao, X., Guo, L., Han, J.: When deep learning meets metric learning: remote sensing image scene classification via learning discriminative cnns. IEEE transactions on geoscience and remote sensing 56(5), 2811–2821 (2018)

[44] Li, Y., Zhang, Y., Huang, X., Zhu, H., Ma, J.: Large-scale remote sensing image retrieval by deep hashing neural networks. IEEE Transactions on Geoscience and Remote Sensing 56(2), 950–965 (2018)