International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Study of Benchmark Datasets for Performance Evaluation of Content Based Image Retrieval System for Remote Sensing Images

Nisha Gupta[1], Ajay Mittal[2], and Satvir Singh[3]

Submitted: 05/01/2024 **Revised:** 12/02/2024 **Accepted:** 20/02/2024

Abstract: The development of new feature extraction techniques and the testing of fresh datasets have greatly improved image retrieval systems for remote sensing. When using Remote Sensing Image Retrieval (RSIR) techniques and assessing their effectiveness, benchmark datasets are essential. Notable advancements in creating benchmark datasets for RSIR systems are highlighted in the literature. Unisource retrieval are the two categories into which these datasets fall. While the query and the images that are retrieved come from the same source in Unisource retrieval, they come from distinct sources in cross-source retrieval. In order to determine which datasets are best suited for the use of deep learning Convolutional Neural Networks and contemporary transfer learning techniques, this research offers a thorough examination of the salient features of both types of remote sensing datasets.

Keywords: *examination, determine, literature,*

1 Introduction

Using the characteristics of a query image, image retrieval looks through large digital image databases to identify the most relevant and related images. A Content-Based Image Retrieval (CBIR) system uses visual attributes like color, shape, and texture to retrieve relevant photos from enormous archives [1]. As shown in figure 1, the CBIR system uses a similarity index to process the visual query image and effectively retrieve relevant visual documents or images from a large repository.

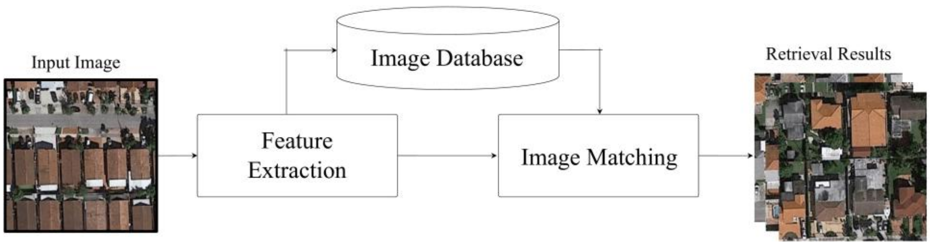


Figure 1: Query Image and Dataset Image matching and retrieval results

Remote sensing uses aerial vehicles like aeroplanes, drones, and balloons to gather data about the Earth's surface without being physically there. This method collects information remotely to investigate many facets of the planet

1Department of Computational Sciences, MRSPTU, Bathinda, India.
* *nishasbs2019@gmail.com*
2Department of Applied Sciences, Aryabhatta Group of Institutes, Barnala, India.
* *mittalajay@gmail.com*
3Department of Electronics and Communication, IKGPTU, Jalandhar, India.
* *drsatvir.in@gmail.com*
**Address correspondence to: nishasbs2019@gmail.com*

and is very useful in forestry, agriculture, and geographic surveys [2]. In order to help with crop optimization, flood detection, land vegetation analysis, water management, and ecosystem conservation, sensors aboard satellites and aeroplanes detect and analyze the Earth's surface and subsurface. Environmental monitoring and sustainable resource management depend on remote sensing [3]. It involves taking pictures from satellites or aeroplanes in order to keep an eye on physical traits from a distance. Large, varied, and complicated, the expanding databases of remote sensing photos have high-dimensional characteristics. The datasets selected for testing have a significant impact on how well a remote sensing image retrieval system performs. Figure 2 shows a general framework for such systems.

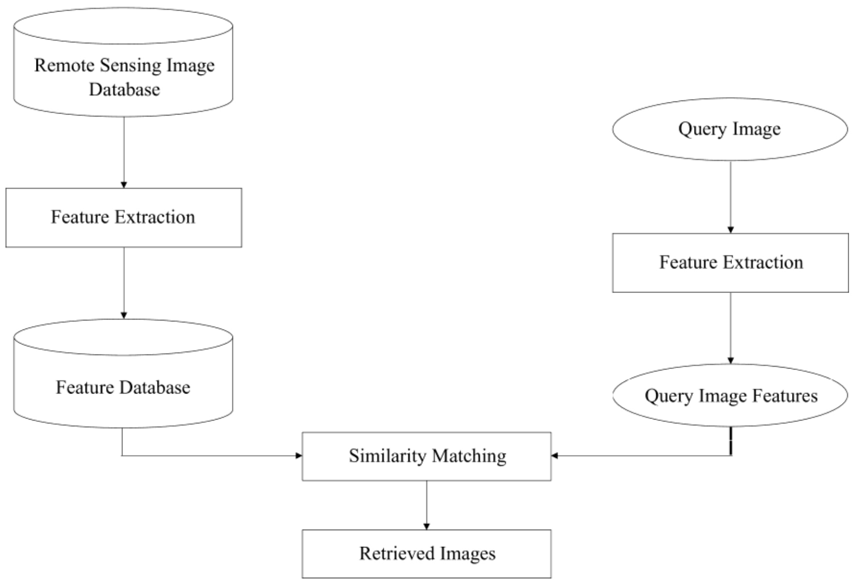


Figure 2: General Framework for Remote Sensing Image Retrieval System

2 Remote Sensing Datasets

To apply multiple Remote Sensing Image Retrieval (RSIR) techniques, datasets are collected according to distinct retrieval modes. Additional classifications for these datasets include SLRSIR, MLRSIR, RASRSIR, CMRSIR, and CVRSIR datasets. With an emphasis on the picture source, labels, and modalities, this study provides a thorough evaluation of the datasets used in remote sensing classification and retrieval research throughout the last ten years.

2.1 SLRSIR Datasets: Single Label Remote Sensing Image Retrieval

Each image in Single Label Image Retrieval (SLRSIR) is given a single label that encapsulates the most important details about it. The retrieved images and the query image are both in the same category. Trees, roads, and buildings are just a few examples of the many classes that are frequently seen in remote sensing photos. In the early days of Remote Sensing Image Retrieval (RSIR), researchers made extensive use of SLRSIR datasets. The UC Merced dataset [4], WHU-RS19 [5], RSSCN7 [6], AID [7], PatternNet [8], RSI-CB [9], SIRI-WHU [10], and NWPU-RESISC45[11] are notable SLRSIR datasets that are often cited in the literature. While the other datasets were originally collected for scene categorization, PatternNet was particularly collected for Remote Sensing Image Retrieval (RSIR).

2.1.1 UC-Merced Land Use Dataset

Another name for the dataset is UC-Merced or UCM. It focuses on classifying land use and land cover and includes 100 photos divided into 21 classifications. With a minimal number of classes and a consistent 256x256 pixel spatial resolution for every image, UCM is comparatively small. With a spatial resolution of 0.3 meters, the dataset consists of a variety of huge aerial photos from the US Geological Survey's National Map Urban Area. Aircraft, farms, forests, motorways, beaches, buildings, intersections, baseball diamonds, harbours, mobile home parks, chaparral, rivers, runways, sparse and dense residential areas, overpasses, parking lots, tennis courts, and storage tanks are just a few of the scene types covered by the dataset's classes. The medium, dense, and sparse classifications may overlap in some classes, such as residential areas [4].

2.1.2 WHU-RS19

The WHU-RS19 collection is made up of satellite photos from Google Earth that range in resolution from 0.5 meters to more. Airport, bridge, beach, commercial area, farming, desert, football field, forest, meadow, industrial area, park, mountain, parking lot, port, pond, railway station, residential area, river, and viaduct are among the 19 different kinds of high-resolution scenes that are included in it. The 50 samples in each class provide images with different illumination, scale, and orientation [5].

2.1.3 RSSCN7

There are 2,800 remote sensing images in seven categories make up the RSSCN7 dataset. Grassland, farming, forests, parking lots, industrial zones, residential zones, and bodies of water like lakes and rivers are some examples of these types of land. There are 400 photos in each category, all taken from Google Earth. Each image is 400 x 400 pixels and is taken at four distinct scales, with 100 photographs each scale. With photos captured at various scales and throughout various seasons and weather fluctuations, the dataset represents a variety of conditions [6].

2.1.4 AID

A vast collection of aerial images from Google Earth is called the AID dataset. The spatial resolutions of the pictures range from 0.5 to 8 meters. Airport, baseball pitch, bare land, bridge, beach, city center, church, commercial area, dense residential, forest, desert, meadow, farmland, industrial area, mountain, medium residential, park, parking lot, pond, port, resort, playground, railway station, river, school, sparse residential, square, stadium, storage tanks and viaduct are among the 30 scene categories included in the dataset. It has 10,000 images, with 220–420 images in each category, all in 600 by 600pixel resolution [7].

2.1.5 Patter Net

With an emphasis on different American locations, PatternNet is a massive dataset of high-resolution photos gathered from Google Earth imagery using the Google Maps API. There are 800, 256x256 pixel images in each of the dataset's 38 classifications. Aircraft, beach, baseball field, bridge, basketball court, chaparral, cemetery, Christmas tree farm, crosswalk, closed road, coastal mansion, dense residential, ferry terminal, football field, forest, harbour, golf course, freeway, nursing home, intersection, mobile home park, oil and gas field, oil well, overpass, parking lot, parking space, railway, river, runway, runway markings, shipping yard, solar panel, sparse residential, storage tank, tennis court, swimming pool, transformer station and Deep learning applications that need a lot of labelled data will benefit greatly from PatternNet, a high-quality dataset with labelled data created for remote sensing image categorization and retrieval [8].

2.1.6 RSI-CB 128

This dataset, which is intended to be large and scalable, is a varied collection of ground objects. Using information from Open Street Maps (OSM), the ground objects are categorized in accordance with China's official land-use categorization guidelines. It has two sub-datasets: one with 256x256 pixel images that has over 24,000 photos in 6 categories and 35 subclasses, and another with 128x128 pixel images that has over 36,000 images in 6 categories and 45 subclasses. Each of the six primary categories—which have multiple subclasses—consists of agricultural land, transportation and facilities, construction land and facilities, woodland, water and water conservancy facilities, and barren areas [9].

2.1.7 SIRI-WHU

SIRI-WHU (Remote Sensing Image Retrieval Group at Wuhan University) produced the dataset, which has 200 images in 12 classifications. The 200x200 pixel photos, which were taken from Google Earth, show Chinese cities with a spatial resolution of 2 meters. Agriculture, Commercial, Industrial, Harbour, Idle Land, Meadow, Overpass, Pond, Park, Residential, River, and Water are among the classes that are included in the dataset citezhao2015dirichlet.

2.1.8 NWPU-RESISC45

With 45 distinct scene classes, the NWPU-RESISC45 benchmark dataset is intended for remote sensing scene categorization. There are 700 photographs in each class, for a total of 31,500 256 x 256pixel images. The collection, which was produced by Northwestern Polytechnic University (NWPU), contains pictures with 0.2–30 m spatial resolutions. Airport, aeroplane, basketball court, baseball diamond, beach, bridge, church, chaparral, circular farmland, cloud, commercial area, dense residential, desert, forest, freeway, ground track field, golf course, harbour, intersection, industrial area, island, lake, medium residential, meadow, mobile home park, mountain, flyover, parking lot, palace, railway station, rectangular farmland, roundabout, river, runway, snowberg, sea ice, ship, sparse residential, storage tank, stadium, thermal power station, terrace, tennis court, and wetland [11].

2.2 MLRSIR Datasets: Multi Label Remote Sensing Image Retrieval

Multilabel Remote Sensing Image Retrieval, or MLRSIR, is the process of retrieving photographs that are linked to several labels. It takes a lot of effort and time to create a multi-label dataset for RSIR. DLRSD [12], WHDLD [3], and BigEarthNet [13] are a few of the datasets that are currently available for MLRSIR.

2.2.1 DLRSD

The Dense Labelling Remote Sensing Dataset, or DLRSD, is a multi-label RSIR (Remote Sensing Image Retrieval) dataset that was created as an expansion of the UC-Merced dataset. This multi-label dataset, which has 21 classes with 100 photos each, was created for the purpose of image retrieval and classification of remote sensing photographs. 17 of these classes—aircraft, bare soil, automobiles, buildings, courts, chaparral, grass, dock, field, mobile homes, pavement, sand, sea, ship, tanks, trees, and water—are comparable to those in the UC-Merced dataset. These 17 classes now have multi-labels instead of just labels. Some multi-labels were upgraded during the DLRSD labelling process and are known as revised multi-labels. For increased accuracy, each image’s pixels were manually labelled [14].

2.2.2 BigEarthNet

The Database Systems and Information Management Group (DIMA) and the Remote Sensing Image Analysis Group (RSIM) at Technische Universität Berlin (TU Berlin) created the extensive collection known as the BigEarthNet dataset. This dataset’s research is funded by the Big Data Centre (BBDB) and the European Research Council (ERC) through the Berlin Institute for the Foundations of Learning and Data’s (BIFOLD) Big Earth grant.

Covering regions in ten European nations (Austria, Belgium, Finland, Ireland, Kosovo, Lithuania, Luxembourg, Portugal, Serbia, and Switzerland), BigEarthNet comprises 590,326 pairings of Sentinel-1 and Sentinel-2 photos. The Sentinel-2 Level 2A product creation tool (sen2cor) was used to produce atmospheric correction for all pictures. 590,326 non-overlapping photographic scenes with multi-label land cover classes are included in the dataset. Only Sentinel-2 photos are included in the Beta-1 version (V1.0), which was subsequently enhanced with Sentinel-1 data to produce BigEarthNet-MM, a multi-modal version [13].

2.3 RASRSIR datasets:

Cross-sensor retrieval Image Recovery from Remote Sensing RARISR (Remote Sensing

Image Retrieval with multiple Sensors) is the term used to describe images taken by numerous sensors. Due to the usage of various sensor types, such as multispectral, panchromatic, and hyperspectral pictures, these images have variable resolutions. There aren't many RARISR databases available to the public; DSRSID is one of them.

2.4 DSRSID Dataset: Dual Source Remote Sensing Image Dataset

The Gaofen-1 optical satellite was used to capture images from panchromatic and multispectral sources for the DSRSID collection. The photos are arranged in pairs, with one panchromatic and one multispectral image in each pair. The multispectral photos have a spatial resolution of 8 meters and are 64 x 64 pixels, whilst the panchromatic images have 256 x 256 pixels and a spatial resolution of 2 meters. Aquafarm, cloud, forest, high building, low building, farmland, river, and water are the eight classes that make up DSRSID. Ten thousand pairs of panchromatic and multispectral photos are included in each class [15].

2.5 CMRSIR Datasets: Cross Modality Remote Sensing Image Retrieval

The images collected are captured from different modality is known as CMRSIR (Cross Modality Remote Sensing Image Retrieval). CMRSIR datasets perform RSIR between different data modalities such as optical and SAR images. Benchmark dataset developed for CMRSIR include RSketch [16], UCM/Syndey-Captions [17], RSICD [18], UCM/Syndey/RSICD-audio [19], TextRS [20] and CBRSIR VS [21] .

2.5.1 RSketch

The RSketch dataset is designed for CMRSIR (Cross-Modal Remote Sensing Image Retrieval) involving remote sensing images and sketches. It includes 20 categories, such as airplane, baseball diamond, basketball court, beach, bridge, closed road, crosswalk, football field, golf course, intersection, oil and gas field, overpass, railway, river, runway, runway marking, storage tank, swimming pool, tennis court, and wastewater treatment plant. Each category contains 200 remote sensing images and 45 sketches, with both image types fixed at a size of 256×256 pixels. The remote sensing images in this dataset are sourced from the UC Merced dataset [4] , WHU-RS19[5], AID [7], and PatternNet [8].

2.5.2 UCM-/Syndey-/RSICD-audio

The UCM-/Sydney-/RSICD-audio datasets are created based on the existing UCM-/Sydney-Captions and RSICD datasets. For each image in these datasets, five descriptive sentences are provided, with each sentence spoken by a different speaker. The spoken audio clips vary in length from 1 to 15 seconds [17].

2.5.3 TextRS

The TextRS dataset is compiled from four existing datasets: UC Merced [4], AID [7], PatternNet [8], and NWPU-45 [11]. It includes 2,144 images randomly selected from these sources. Each image is annotated with five sentences, each generated by a different person to ensure diversity [20].

2.6 CVRSIR: Cross View Remote Sensing Image Retrieval

Images captured from different viewpoints are referred to as cross-view remote sensing image retrieval (CVRSIR). In a CVRSIR dataset, images are typically gathered from various perspectives including ground aerial views, satellite views, and drone views. One example of a CVRSIR dataset is University-165 [22], CVACT [23], AiRound / CvBrCT [24], CVUSA subset [25], VIGOR [26], Vo and Hays [27].

2.6.1 University-1652

The University-1652 dataset is a multi-view, multi-source dataset designed for drone-based geo-localization. It includes ground-view, drone-view, and satellite-view images of 1,652 buildings from 72 universities. The training set comprises images of 701 buildings from 33 universities, while the testing set includes images of 701 buildings from the remaining 39 universities. University-1652 offers a rich collection of images per class, spanning multiple sources and views [22].

2.6.2 CVUSA subset

The CVUSA subset dataset is a smaller version of the original CVUSA. It consists of panoramas selected from the CVUSA dataset to represent ground-view images. For each ground-view panorama, aerial images at zoom level 19 are downloaded from Bing Maps for the same geographic area. Panoramas without corresponding aerial images are excluded, resulting in 35,532 training image pairs and 8,884 testing image pairs. The ground-view images have a resolution of 1232×224 pixels, while the satellite images are 750×750 pixels [25].

2.6.3 SAT-4 and SAT-6 airborne Datasets

SAT-4 and SAT-6 images are derived from the National Agriculture Imagery Program (NAIP) dataset. This dataset includes 330,000 images from the Continental United States (CONUS), comprising uncompressed Digital Ortho Quarter Quad (DOQQ) tiles and GeoTIFF images from the US Geological Survey (USGS). Each image tile is approximately 6000 pixels wide by 7000 pixels high, with an average size of around 200 megabytes. The entire NAIP dataset for CONUS totals about 65 terabytes and includes images in four bands: red, green, blue, and infrared.For SAT-4 and SAT-6, 1,500 images covering a variety of landscapes, such as rural and urban areas, mountainous terrain, small to large water bodies, densely forested regions, and agricultural areas in California, were selected. Each image is manually labeled for specific land cover classes. After labeling, 28x28 non-overlapping image patches are extracted and stored in the dataset with corresponding labels [28].

SAT-4 is named for its training labels, which are formatted as 1x4, with a single index from 0 to 4 representing the class. SAT-6 is named for its test labels, formatted as 1x6, with a single index from 0 to 6 representing the class [29].

2.6.4 Eurosat

Eurosat is a benchmark dataset designed for deep learning applications in land use classification within geospatial images. It consists of Sentinel-2 satellite images across 13 spectral bands. The dataset includes 27,000 georeferenced images categorized into 10 classes [30]. A summary of publicly available benchmark datasets for various Remote Sensing Image Retrieval methods in recent years is presented in table 1.

3 Candidate Datasets for Remote Sensing Image Classification and Retrieval suitable for Deep Learning Implementations

Benchmark datasets are the basis for implementing Remote Sensing Image Retrieval (RSIR) methods and performance evaluation. Performance of experiments is solely dependent upon appropriate dataset selection [11]. After the indepth study of characteristics of remote sensing datasets four RS datasets UC-Merced [4], AID [7], NWPU-RESISC45 [11] and PatterNet [8][31] with unisource retrieval may be considered well suitable for remote sensing image classification and retrieval system. Summary of the characteristics of these datasets are presented in table 2.

Table 2: Populated Datasets used by researchers for implementing deep learning algorithms

DATASET SELECTED	TOTAL CLASSES CONTAINED IN DATASET	TOTAL IMAGES CONTAINED IN DATASET	IMAGE SIZE (Pixels)	SPATIAL RESOLUTIO N (m)
UC-Merced [4]	21	2100	256×256	0.3
AID [7]	30	10000	600×600	8-0.5
NWPU-RESISC45[11]	12	31500	256×256	30-0.2

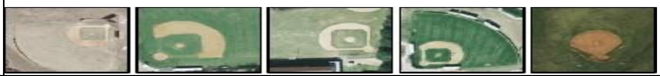
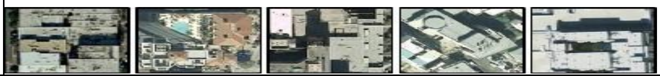
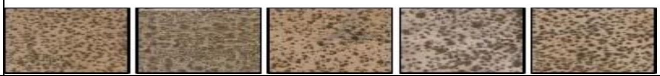
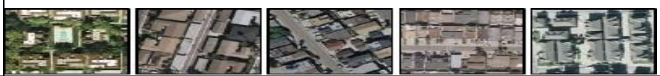
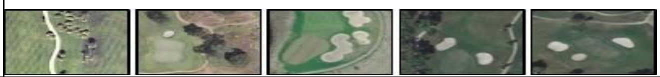
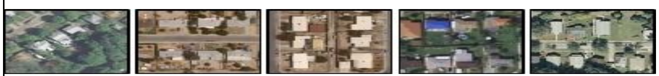
PatterNet [8]	38	30400	256×256	4.69-0.06

Sample images of UC-Merced, AID, NWPU-RESISC45 and PatterNet dataset are presented in table 3, table 4, table 5, table 6 respectively.

Table 1: Summary of Remote Sensing Datasets

DATASET	MAGE COUNT IN DATASET	SIZE OF IMAGES	ANNOTATION INFORMATION	SINGLE LABEL/ MULTILABEL
UCM [4]	2,100	256×256	21 Categories Single Label	Singlelabel Unisource Retrieval
AID in [7]	600*600	30 Categories Single Label	Singlelabel Unisource Retrieval	
NWPU-RESISC45 [11]	31,500	256×256	21 Categories Single Label	Singlelabel Unisource Retrieval
PatternNet [8]	30,400	256×256	38 Categories Single Label	Singlelabel Unisource Retrieval
WHU-RS19 [5]	1,005	600×600	19 Categories Single Label	Singlelabel Unisource Retrieval
RSSCN7 [6]	2,800	400×400	7 Categories Single Label	Singlelabel Unisource Retrieval
SIRI-WHU [10]	2,400	200×200	12 Categories Single Label	Singlelabel Unisource Retrieval
DSRSID [14]	60,000	64×64- 256×256	6 Categories Single Label	Crosssource Retrieval
RSI-CB128 [9]	36,707	128×128	45 Categories Single Label	Singlelabel Unisource Retrieval
SAT-4 [29]	500,000	64×64	4 Categories from Single Label	Singlelabel Unisource Retrieval
SAT-6 [28]	405,000	64×64	6 Categories from Single Label	Singlelabel Unisource Retrieval
EuroSat [30]	27,000	64×64	10 Categories from Single Label	Singlelabel Unisource Retrieval
BigEarthNet [13]	590,326	20×20 - 120×120	43 Categories from Multi Label	Multilabel Unisource Retrieval

Table 3: UC-MERCED: Remote Sensing Dataset

UC-Merced: Remote Sensing Dataset		
Class 1.	AGRICULTURAL	
Class 2.	AIRPLANE	
Class 3.	BASE BALL DIAMOND	
Class 4.	BEACH	
Class 5.	BUILDINGS	
Class 6.	CHAPARRAL	
Class 7.	DENSE RESIDENTIAL	
Class 8.	FOREST	
Class 9.	FREEWAY	
Class 10.	GOLF COURSE	
Class 11.	HARBOR	
Class 12.	INTERSECTION	
Class 13.	MEDIUM RESIDENTIAL	
Class 14.	MOBILE HOME PARK	
Class 15.	OVERPASS	
Class 16.	PARKING LOT	

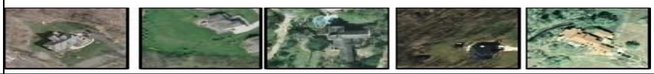
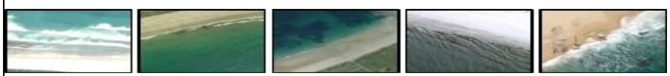
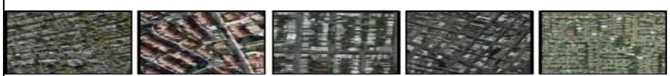
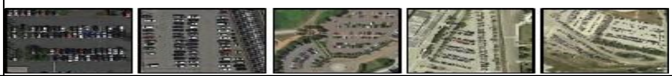
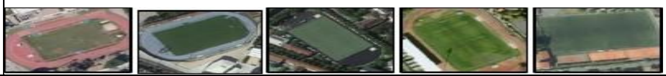
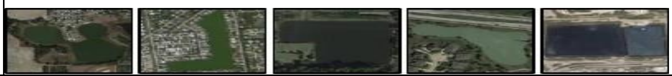
Class 17.	RIVER	
Class 18.	RUNWAY	
Class 19.	SPARSE RESIDENTIAL	
Class 20.	STORAGE TANKS	
Class 21.	TENNIS COURT	

Table 4: AID: Remote Sensing Dataset

REMOTE SENSING IMAGES DATASET : AID		
Class 1.	AIRPORT	
Class 2.	BARE LAND	
Class 3.	BASEBALL FIELD	
Class 4.	BEACH	
Class 5.	BRIDGE	
Class 6.	CENTER	
Class 7.	CHURCH	
Class 8.	COMMERCIAL	
Class 9.	DENSE RESIDENTIAL	
Class 10.	DESERT	

Class 11.	FARMLAND	
Class 12.	FOREST	
Class 13.	INDUSTRIAL	
Class 14.	MEADOW	
Class 15.	MEDIUM RESIDENTIAL	
Class 16.	MOUNTAIN	
Class 17.	PARK	
Class 18.	PARKING	
Class 19.	PLAYGROUND	
Class 20.	POND	
Class 21.	PORT	
Class 22.	RAILWAY STATION	
Class 23.	RESORT	
Class 24.	RIVER	
Class 25.	SCHOOL	
Class 26.	SPARSE RESIDENTIAL	
Class 27.	SQUARE	

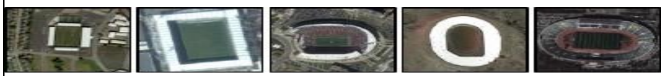
Class 28.	STADIUM	
Class 29.	STORAGE TANK	
Class 30.	VIADUCT	

Table 5: NWPU-RESISC45:Remote Sensing Dataset

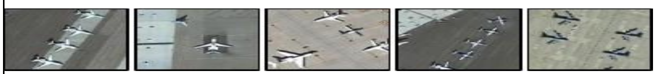
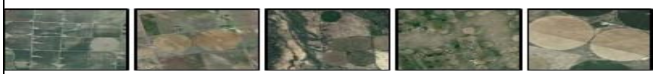
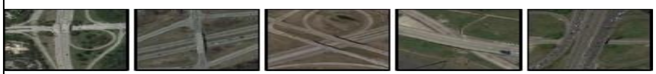
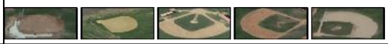
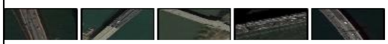
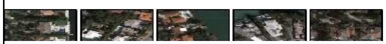
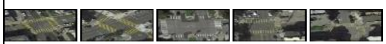
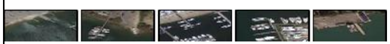
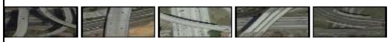
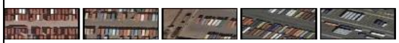
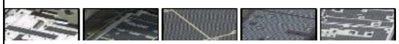
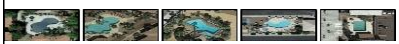
REMOTE SENSING IMAGES DATASET : NWPU-RESISC45		
Class 1.	AIRFEILD	
Class 2.	ANCHORAGE	
Class 3.	BEACH	
Class 4.	DENSE RESIDENTIAL	
Class 5.	FARM	
Class 6.	FLYOVER	
Class 7.	FOREST	
Class 8.	GAME SPACE	
Class 9.	PARKING SPACE	
Class 10.	RIVER	
Class 11.	SPARSE RESIDENTIAL	
Class 12.	STORAGE CISTERNS	

Table 6: PatterNet: Remote Sensing Dataset

REMOTE SENSING IMAGES DATASET : PATTERNET		
Class 1.	AIRPLANE	
Class 2.	BASEBALL FIELD	
Class 3.	BASKETBALL COURT	
Class 4.	BEACH	
Class 5.	BRIDGE	
Class 6.	CEMETERY	
Class 7.	CHAPARRAL	
Class 8.	CHRISTMAS TREE FARM	
Class 9.	CLOSED ROAD	
Class 10.	COASTAL MANSION	
Class 11.	CROSSWALK	
Class 12.	DENSE RESIDENTIAL	
Class 13.	FERRY TERMINAL	
Class 14.	FOOTBALL FIELD	
Class 15.	FOREST	
Class 16.	FREEWAY	
Class 17.	GOLF COURSE	
Class 18.	HARBOR	
Class 19.	INTERSECTION	
Class 20.	MOBILE HOME PARK	
Class 21.	NURSING HOME	
Class 22.	OIL GAS FIELD	
Class 23.	OIL WELL	

		
Class 24.	OVERPASS	
Class 25.	PARKING LOT	
Class 26.	PARKING SPACE	
Class 27.	RAILWAY	
Class 28.	RIVER	
Class 29.	RUNWAY	
Class 30.	RUNWAY MARKING	
Class 31.	SHIPPING YARD	
Class 32.	SOLAR PANEL	
Class 33.	SPARSE RESIDENTIAL	
Class 34.	STORAGE TANK	
Class 35.	SWIMMING POOL	
Class 36.	TENNIS COURT	
Class 37.	TRANSFORMER STATION	
Class 38.	WASTEWATER TREATMENT PLANT	

4 Conclusion

This paper has explored various categories of remote sensing datasets, including SLRSIR, MLRSIR, RASRSIR, CMRSIR, and CVRSIR datasets, used for image classification and retrieval. It provides a detailed examination of the number of classes, varying resolutions, and image sizes. Researchers in remote sensing have leveraged these benchmark datasets to test their proposed techniques. Based on the literature review, experiments conducted on different datasets have led to the selection of four benchmark datasets: UC-Merced, AID, NWPU-RESISC45, and PatterNet.

PatterNet is an extensive dataset featuring high-resolution images sourced from Google Earth imagery via the Google Maps API, covering various US cities. It includes 38 classes with 800 images per class, each sized at 256×256 pixels. PatterNet is a highly valuable dataset for remote sensing image classification and retrieval, offering labeled data that is particularly suited for deep learning techniques requiring large amounts of labeled data [8][31].

References

[1] A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N. I. Ratyal, B. Zafar, S. H. Dar, M. Sajid, T. Khalil, *et al.*, "Content-based image retrieval and feature extraction: a comprehensive review," *Mathematical problems in engineering*, vol. 2019, 2019.

[2] Y. Li, J. Ma, and Y. Zhang, "Image retrieval from remote sensing big data: A survey," *Information Fusion*, vol. 67, pp. 94–115, 2021.

[3] Z. Shao, W. Zhou, X. Deng, M. Zhang, and Q. Cheng, "Multilabel remote sensing image retrieval based on fully convolutional network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 318–328, 2020.

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

[5] G.-S. Xia, W. Yang, J. Delon, Y. Gousseau, H. Sun, and H. Maˆıtre, "Structural high-resolution satellite image indexing," in *ISPRS TC VII Symposium-100 Years ISPRS*, vol. 38, pp. 298–303, 2010.

[6] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classifi- cation," *IEEE Geoscience and remote sensing letters*, vol. 12, no. 11, pp. 2321–2325, 2015.

[7] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[8] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS journal of photogrammetry and remote sensing*, vol. 145, pp. 197–209, 2018.

[9] H. Li, X. Dou, C. Tao, Z. Hou, J. Chen, J. Peng, M. Deng, and L. Zhao, "Rsi-cb: A large scale remote sensing image classification benchmark via crowdsource data," *arXiv preprint arXiv:1705.10450*, 2017.

[10] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 2108–2123, 2015.

[11] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[12] Z. Shao, K. Yang, and W. Zhou, "Performance evaluation of single-label and multi-label remote sensing image retrieval using a dense labeling dataset," *Remote Sensing*, vol. 10, no. 6, p. 964, 2018.

[13] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for re- mote sensing image understanding," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 5901–5904, IEEE, 2019.

[14] Z. Shao, K. Yang, and W. Zhou, "A benchmark dataset for performance evaluation of multi-label remote sensing image retrieval.," *Remote Sensing*, vol. 10, no. 6, 2018.

[15] U. Chaudhuri, B. Banerjee, A. Bhattacharya, and M. Datcu, "Cmir-net: A deep learning based model for cross-modal retrieval in remote sensing," *Pattern recognition letters*, vol. 131, pp. 456–462, 2020.

[16] F. Xu, W. Yang, T. Jiang, S. Lin, H. Luo, and G.-S. Xia, "Mental retrieval of remote sensing images via adversarial sketch-image feature learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7801–7814, 2020.

[17] B. Qu, X. Li, D. Tao, and X. Lu, "Deep semantic understanding of high resolution remote sensing image," in *2016 International conference on computer, information and telecommunication systems (Cits)*, pp. 1–5, IEEE, 2016.

[18] X. Lu, B. Wang, X. Zheng, and X. Li, "Exploring models and data for remote sensing image caption generation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2183–2195, 2017.

[19] G. Mao, Y. Yuan, and L. Xiaoqiang, "Deep cross-modal retrieval for remote sensing image and audio," in *2018 10th IAPR workshop on pattern recognition in remote sensing (PRRS)*, pp. 1–7, IEEE, 2018.

[20] T. Abdullah, Y. Bazi, M. M. Al Rahhal, M. L. Mekhalfi, L. Rangarajan, and M. Zuair, "Textrs: Deep bidi- rectional triplet

network for matching text to remote sensing images," *Remote Sensing*, vol. 12, no. 3, p. 405, 2020.
[21] Y. Sun, S. Feng, Y. Ye, X. Li, J. Kang, Z. Huang, and C. Luo, "Multisensor fusion and explicit semantic preserving-based deep hashing for cross-modal remote sensing image retrieval," *IEEE Transactions on Geo- science and Remote Sensing*, vol. 60, pp. 1–14, 2021.
[22] Z. Zheng, Y. Wei, and Y. Yang, "University-1652: A multi-view multi-source benchmark for drone-based geo- localization," in *Proceedings of the 28th ACM international conference on Multimedia*, pp. 1395–1403, 2020.
[23] L. Liu and H. Li, "Lending orientation to neural networks for cross-view geo-localization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5624–5633, 2019.
[24] G. Machado, E. Ferreira, K. Nogueira, H. Oliveira, M. Brito, P. H. T. Gama, and J. A. dos Santos, "Airound and cv-brct: Novel multiview datasets for scene classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 488–503, 2020.
[25] M. Zhai, Z. Bessinger, S. Workman, and N. Jacobs, "Predicting ground-level scene layout from aerial imagery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 867–875, 2017.
[26] S. Zhu, T. Yang, and C. Chen, "Vigor: Cross-view image geo-localization beyond one-to-one retrieval," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3640–3649, 2021.
[27] N. N. Vo and J. Hays, "Localizing and orienting street views using overhead imagery," in *Computer Vision– ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 494–509, Springer, 2016.
[28] R. M. Khan and S. Ghuffar, "Benchmarking deepsat dataset with a simple convolutional neural network," in *2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pp. 1215–1220, IEEE, 2018.
[29] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. Nemani, "Deepsat: a learning framework for satellite imagery," in *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, pp. 1–10, 2015.
[30] P. Helber, B. Bischke, A. Dengel, and D. Borth, "Eurosat: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, pp. 2217–2226, 2019.
[31] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Learning low dimensional convolutional neural networks for high- resolution remote sensing image retrieval," *Remote Sensing*, vol. 9, no. 5, p. 489, 2017.