



OPEN ACCESS

EDITED BY

Fang Huang,
University of Electronic Science and
Technology of China, China

REVIEWED BY

Yi Zeng,
Beijing Forestry University, China
Mohd Anul Haq,
Majmaah University, Saudi Arabia
Wei Li,
The University of Sydney, Australia
Xiaohui Ji,
China University of Geosciences, China

*CORRESPONDENCE

Yan Ma,
✉ mayan@aircas.ac.cn

RECEIVED 21 February 2025

ACCEPTED 27 March 2025

PUBLISHED 15 April 2025

CITATION

Ren X, Ma Y and Zhou Y (2025) Label semantics and image features aware remote sensing sample retrieval from multi-source datasets for AI-enabled remote sensing monitoring. *Front. Environ. Sci.* 13:1580797. doi: 10.3389/fenvs.2025.1580797

COPYRIGHT

© 2025 Ren, Ma and Zhou. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Label semantics and image features aware remote sensing sample retrieval from multi-source datasets for AI-enabled remote sensing monitoring

XingTao Ren^{1,2}, Yan Ma^{1*} and YiXin Zhou^{1,2}

¹Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, ²School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China

Introduction: With the advent of remote sensing (RS) big data, the evolving deep learning (DL) enabled analysis has shown remarkable potential in uncovering intricate features and detecting environmental changes from the vast influx of remote sensing imagery. This is particularly promising for data-driven, intelligent environmental change monitoring, including coastal land cover classification, urbanization impact analysis and so on. Accordingly, the remote sensing imagery sample datasets have become crucial in ensuring robust training models with satisfactory performance across various AI (Artificial Intelligence)-enabled remote sensing applications. However, as the significant demand for more abundant and diverse remote sensing imagery samples continues to surge, sample data scarcity has emerged as a critical challenge for large-scale AI-enabled remote sensing applications. Moreover, the inconsistencies in labeling semantic categories among sample datasets, coupled with the limits of the commonly used single-label representation of datasets, have posed a huge barrier to fully leveraging the training samples across diverse remote sensing sample datasets. Besides, the existing remote sensing training sample datasets dispersed across various hosting platforms are typically organized in supplier-specific and arbitrary data structures, making it rather trivial and difficult for applications to gather demand sample datasets to fit into training models.

Methods: To tackle the above challenges, we propose an intelligent remote sensing sample dataset retrieval approach with awareness of label semantics and visual features for fully integrating and leveraging sample datasets for AI-enabled remote sensing monitoring applications. Notably, it takes both label semantics and visual features into consideration during cross-dataset querying by measuring the similarity distance of visual features and label semantics between samples and the cluster center of label categories. Following this way, it could dynamically build application-tailored remote sensing sample datasets to better harness the multi-source sample datasets with single-label limits and label disparities. Moreover, it also establishes a dynamic RS label category system that is capable of dynamically expanding distinct categories from new sample datasets through label semantic similarity mapping to resolve label inconsistencies across sample datasets. In addition, it conducts in-memory

sample data discovery and integration across clouds supported by a virtual distributed storage system to sufficiently leverage the multi-source remote sensing sample datasets from platforms with limited interoperability.

Results: The comparative performance experiments have confirmed the effectiveness and efficiency of this approach. The proposed method is capable of dynamically integrating and leveraging multi-source sample datasets, effectively addressing the challenges of sample data scarcity, label inconsistencies, and data structure disparities. The dynamic RS label category system and the cross-dataset retrieval approach have shown significant improvements in sample dataset integration and retrieval accuracy.

Discussion: The proposed intelligent remote sensing sample dataset retrieval approach provides a comprehensive solution to the challenges faced by large-scale AI-enabled remote sensing applications. By integrating label semantics and visual features, the method enhances the accuracy and efficiency of sample dataset retrieval and integration. The dynamic label category system and in-memory data discovery mechanisms further improve the usability and accessibility of multi-source remote sensing sample datasets. Future work may focus on further optimizing the retrieval algorithms and expanding the range of supported data platforms to enhance the robustness and scalability of the system.

KEYWORDS

remote sensing sample datasets, remote sensing sample retrieval, remote sensing sample database, AI-enabled remote sensing application, remote sensing imagery, deep learning, label category system, remote sensing big data

1 Introduction

With the exponential growth of remote sensing big data and the blooming of Artificial Intelligence (AI) technologies, deep learning has emerged as a transformative approach, particularly in the context of remote sensing monitoring for coastal cities and environments. This includes applications such as coastal land cover classification (Lv et al., 2024), urbanization impact analysis (Qi, 2024), and shoreline detection and change monitoring (Dang et al., 2022). These applications are crucial for understanding and managing the complex interactions between urban development and environmental changes in coastal regions. By leveraging large-scale RS imagery and evolving training model structures, the deep-learning-enabled approach has shown increasing potential in improving the automatic interpretation of complex features and the detection of environmental changes that traditional methods may overlook (Yuan et al., 2020). Gradually, remote sensing training sample datasets have emerged as vital determinants in delivering optimal model performance for AI-powered RS applications. The availability and quality of these datasets directly impact the accuracy and reliability of AI models used for environmental change detection and urban impact assessment in coastal regions.

Up till now, significant efforts have been devoted to labeling training datasets derived from vast, multi-source satellite imagery for diverse AI-enabled remote sensing tasks. Some of the most widely used remote sensing datasets include the UC Merced dataset for urban land use classification (Yang and Newsam, 2010), and BigEarthNet (Sumbul et al., 2019) for large-scale multi-class classification (Sumbul et al., 2019). Additionally, the SpaceNet dataset has become prominent for tasks such as building detection and urban mapping (Van Etten et al., 2018). These datasets provide a diverse set of samples that cover different

geographic regions, seasons, and time periods. Meanwhile, there has been a surging demand for more abundant and comprehensively labeled training samples to improve the generalization performance of models on unseen data. However, despite the growing amount of RS sample datasets built from the vast deluge of remote sensing imagery, sample data scarcity has practically emerged as a critical challenge (Cheng et al., 2020).

Unfortunately, several challenging issues have further severe the scarcity of remote sensing sample datasets. *Firstly*, the existing remote sensing sample datasets are typically dispersed across different data hosting platforms, such as AWS Clouds and Kaggle, and each differs in access protocols. To complicate the case, the sample datasets from these distinct hosting platforms are mainly organized in diverse arbitrary and supplier-specific manners. These discrepancies have significantly impeded the interoperable harnessing of extensive datasets scattered across different repositories, which may greatly exacerbate data scarcity issues. Hence, an integrated sample management system is urgently demanded to facilitate comprehensive and interoperable sharing and access across clouds. *Secondly*, remote sensing sample datasets from different hosting platforms inevitably differ in label category systems and label representations. Diverse platforms or organizations may adopt varying categorizations to label the same land cover objects. However inconsistent semantic or granularity of label categories can cause models to struggle during training (Li et al., 2018), ultimately posing significant challenges in comprehensively leveraging multi-source and multi-sensor samples across datasets. *In addition*, current labeling practices, with coarse category labels and single-label representations, often fail to sufficiently depict the content of remote sensing sample imagery (Chaudhuri et al., 2017). Consequently, the traditional label-based sample retrieval often

struggles to properly locate the semantically matched samples. Although cross-dataset queries incorporate label semantics, such as LuoJiaset (Cao et al., 2023), they often result in compromised retrieval performance across datasets, as they rely solely on inadequately represented label semantics. Therefore, fostering innovative retrieval strategies that better present samples' context is crucial for enhanced retrieval outcomes in various remote sensing applications.

To properly settle the challenging issues above, we propose an intelligent remote sensing imagery samples retrieval approach with awareness of label semantics and image features. This approach aims to fully leverage multi-source sample datasets from various platforms to generate more imagery samples that accurately meet the model training needs of different remote sensing monitoring applications. It is worth noting that this approach takes both the imagery visual features and the label semantic features of sample datasets into consideration during sample data integration and cross-dataset retrieval procedure. By utilizing a BERT-based word embedding model to capture label semantics and a multi-label scene classification model to extract visual features, this approach enables comprehensive semantic data integration. For performance efficiency, the extracted feature vectors are then cached and indexed in feature database using Inverted File Product Quantization (IVF-PQ) encoding. The kernel retrieval method conducts a multi-level cross-dataset sample query, starting with spatial retrieval for an initial query, followed by label-semantic-based retrieval for more precise filtering, and ending with image-feature-based imagery retrieval for an extended sample retrieval to gather more supplementary samples. Following this retrieval method, it can provide composite sample datasets with more finely filtered samples from multi-source datasets, tailored semantically to meet various model training needs of remote sensing applications.

Moreover, to eliminate the inconsistency of supplier-specific label categories among multi-source samples, it adopts a knowledge-graph-based dynamic label category system on top of Million-AID (Long et al., 2021) label category network, which serves as a standard label category semantic network. Enabled by label semantic similarity and structural similarity mapping, the proposed label category system could build semantic relations between the existing categories and new labels to dynamically expand new label categories from newly integrated sample datasets. Withal, by virtue of a multi-task data crawler engine CrawlLab and Alluxio for in-memory virtual data accessing across clouds, it could also offer highly efficient sample data discovery and integration across platforms.

The rest of this paper is organized as follows. Section II reviews related work. In Section III, we present the framework and implementation of the cross-cloud dynamic remote sensing sample data management platform and describe the design principles of the system. Section IV discusses the experimental validation and analysis of the proposed platform. Finally, Section V concludes the paper.

2 Related work

2.1 Sample data management

2.1.1 Dataset-oriented sample data management

Current mainstream sample data integration management techniques employ a dataset-oriented management approach,

where datasets are treated as the smallest unit of management as shown in Table 1. Platforms such as AWS, Kaggle (Igloukov et al., 2017), and AIEarth (He et al., 2024) independently manage sample data from various datasets and integrates multiple datasets through user-uploaded data, providing a unified access interface for the datasets. Within each dataset, sample retrieval is enabled based on attributes such as labels and spatiotemporal ranges. While this management approach can offer a vast amount of remote sensing data, it falls short of meeting the needs of more advanced applications in remote sensing deep learning. As these applications evolve, a single dataset often fails to provide sufficient samples. Complex applications like coastal city and environmental monitoring applications seek to combine sample data from multiple datasets to train more robust models. However, dataset-oriented management lacks the ability to manage samples across datasets, thus limiting its capability to meet these emerging application requirements.

2.1.2 Sample-oriented sample data management

A single RS dataset often fails to meet the requirements of deep-learning-enabled remote sensing applications. There is an urgent need to organize sample data across multiple datasets to obtain the necessary sample data from various sources (Gominski et al., 2022).

Gong et al. (Cao et al., 2023) have proposed a label-based shared management platform that organizes samples by sample units across multiple datasets. Unlike dataset-oriented management, these platforms enable cross-dataset sample retrieval, allowing the free combination of multiple dataset samples.

Combining different dataset samples significantly enhances the intra-class diversity of the dataset. Variations in lighting, background, geographical location, and scale among similar samples from different datasets contribute to diversity, suggesting that a richer intra-class diversity of a sample set is closer to representing the real-world distribution, thus better reflecting the general characteristics of similar entities.

2.2 Label category system

2.2.1 File-directory structured label category system

Current remote sensing deep learning sample datasets often use a tree directory structure, with labels serving as directory names for datasets related to scene classification, object detection, and other applications. This label management approach provides a straightforward visualization of sample label information without the need for additional storage.

However, because label management is highly intertwined with sample management, directory-based label management is inefficient in querying and challenging to update. File system-based queries are generally less efficient than database management systems, and extensive directory restructuring may be necessary as datasets grow over time or when label categories need to be updated or modified. Furthermore, in modern deep learning applications where a single sample may have multiple labels, relying on a directory structure to define labels can lead to redundant storage of multi-labeled samples.

Some datasets, like MLRSNet (Qi et al., 2020), utilize additional mapping files to record the mapping from sample instances to their

TABLE 1 Summary of platform functionalities.

Platform	Data upload	Label storage format	Sample filtering conditions
GEE	Supported	Sample metadata attributes	Time, Space, Cloud Coverageetc.
AWS	Supported	Directory names and label mapping files	Time, Space
Kaggle	Supported	Label mapping files	Keywords, Upload Date
AIEarth	Supported	Sample metadata attributes	Time, Space, Label

corresponding labels. This label management approach separates the management of sample instances and labels, facilitating updates and integration with databases to support complex queries and data handling. Unlike the natural tree structure of directory names, this method does not reflect the dependency relationships between different labels.

2.2.2 Knowledge graph-based label category system

Early sample datasets treated different semantic labels as independent entities at the same granularity level, ignoring semantic relationships between categories. To consolidate existing sample datasets and eliminate semantic differences between labels, several sample datasets and platforms have proposed their own label category systems, referring to land cover/use standards and real-world application needs (Cao et al., 2023; Li et al., 2020; Long et al., 2021; Jin et al., 2018). These systems merge existing sample datasets by semantic mapping of labels and discarding ambiguous categories (Peng et al., 2020).

Recently, the dependency relationships between labels have gained significant attention in deep learning applications, particularly in multi-label learning. Based on structured knowledge graphs like WordNet to measure label similarity, these relationships are used to construct label knowledge graphs and are applied in multi-label zero-shot learning (Lee et al., 2018). Chen, Li (Chen et al., 2019; Li et al., 2022), et al. have modeled label dependency relationships as directed graphs and used graph neural networks to explore complex relationships between labels in these graphs.

Introducing graph-based label category systems enables detailed modeling of complex relationships between label categories, facilitating the learning of dependencies between different labels. Moreover, the graph structure offers expandability, supporting the introduction of new label categories.

2.3 Sample data retrieval methods

2.3.1 Label Keyword-Based Sample Retrieval

Traditional sample retrieval methods include Label Keyword-Based Sample Retrieval, where sample data within a single dataset is retrieved based on sample labels and keywords from sample metadata. The metadata of remote sensing samples primarily consists of additional information such as acquisition time, spatial extent, sensor details, and spectral band information, which can be used to characterize the sample. By providing sample label keywords or defining specific metadata, a retrieval process can return sample data that accurately matches the given criteria.

Due to its simple retrieval logic and fast response time, current mainstream sample data management platforms, such as Kaggle, Sense Earth, and GEE (Gorelick et al., 2017), support Label Keyword-Based Sample Retrieval. Sense Earth provides dataset search services based on various criteria, including dataset name, data type, data format, and creation time. Kaggle allows dataset retrieval based on keywords, data type, dataset creator, and upload date. GEE offers filtering capabilities by dataset name and, within datasets, supports filtering data based on time, spectral bands, cloud cover, and various geometric types (such as points, polygons, etc.). LuojiaSet offers filtering based on task type, dataset name, image type, label category, and spatiotemporal range.

However, due to the immaturity of current sample labeling techniques, sample labels often fail to fully represent the complex content of sample images. This leads to Label Keyword-Based Sample Retrieval missing some usable samples. Furthermore, due to differences in labeling systems and granularity across datasets, Label Keyword-Based Sample Retrieval is limited to retrieving samples within a single dataset and cannot obtain required samples across datasets, making it difficult to meet the needs of current intelligent remote sensing applications.

2.3.2 Cross-dataset sample retrieval

In modern intelligent remote sensing applications, diverse label category schemes are required to meet varying application-specific demands. This necessitates the selection of appropriate label categories from different sample datasets based on specific application needs. However, the sample size of a single dataset is often insufficient to meet the demands of complex applications, and significant differences in sample organization and labeling systems across datasets make traditional sample retrieval methods limited to a single dataset. To address this issue, Gong et al. proposed a cross-dataset compatible label category system. By mapping the label categories of each dataset to a unified labeling system, they integrated samples from multiple datasets into a unified sample repository and achieved cross-dataset sample retrieval. Based on label mapping, cross-dataset sample retrieval enables the retrieval of usable samples from multiple datasets. However, its essence remains label-based, which, limited by the representational capacity of labels, cannot fully explore and utilize potential usable samples.

2.3.3 Image feature-based sample retrieval

There is a certain degree of mismatch between existing sample labels and sample images, which manifests in two main aspects: 1) Coarse category labels often fail to represent detailed label categories in sample data. For example, the

“cultivated land” label cannot accurately describe sample images that contain arid land categories, leading to missed samples when retrieving more specific label categories. 2) The complexity of remote sensing images often results in sample images containing multiple object-level semantic categories. Single-label classification is usually insufficient to fully describe sample image content with complex categories (Shao et al., 2020). Consequently, single-label samples may contain additional semantic categories besides the labeled one. These issues prevent traditional label-based sample retrieval from comprehensively capturing all qualifying samples in the sample repository.

With the development of deep learning technologies, image feature-based retrieval techniques have gradually matured. Tang et al. utilized pre-trained sample representation models to extract visual features from sample images and applied the HNSW (Hierarchical Navigable Small World) approximate nearest neighbor matching algorithm to match samples with similar image content (Tang et al., 2024). Existing representation models can uncover high-level feature representations of images, such as color, shape, and texture, which provide more semantic information compared to vague label text. As such, performing search and matching based on the visual features of sample images can overcome the subjectivity and incompleteness of labels and metadata. This approach has the potential to discover visual similarities that users might not have anticipated. Tang et al. (2024). Have already made attempts to retrieve samples based on image features, but a mature and usable system has not yet been developed. Unlike other remote sensing data, remote sensing sample data typically comes with expert-annotated labels, so relying solely on image features for retrieval does not fully leverage the information provided by expert annotations.

3 Methodology

3.1 Main Solution

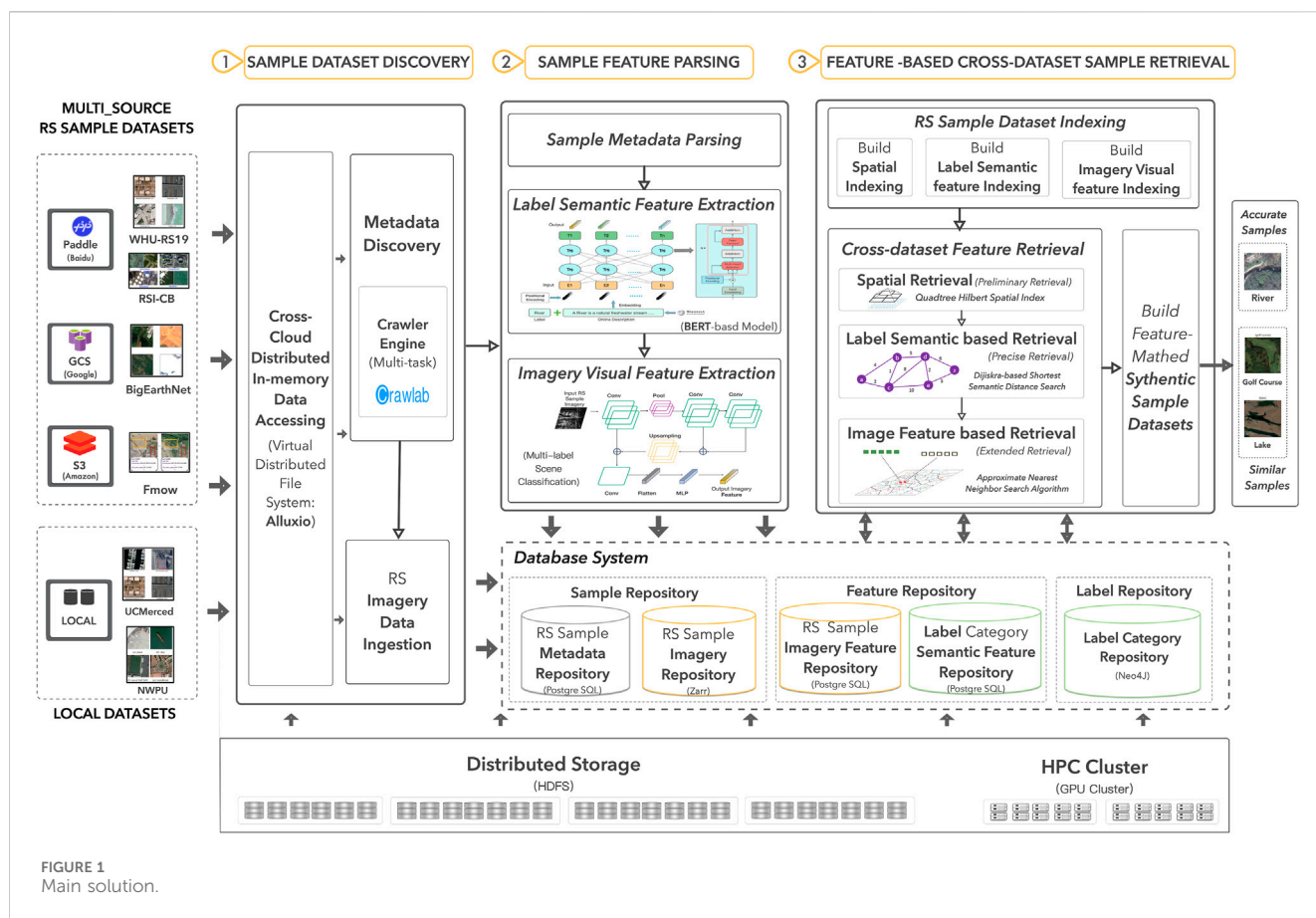
The inconsistency of label categories among multi-source samples, the limits of single-label representation of datasets, as well as the diverse supplier-specific data organization and arbitrary data structures, make it quite trivial and difficult to make sufficient use of the surging volume of diverse remote sensing imagery samples available from diverse platforms. As a result, obtaining abundant high-quality remote sensing sample datasets that properly fit the demand of various remote sensing monitoring applications remains quite challenging. To properly tackle these challenging obstacles, an intelligent remote sensing sample retrieval approach with awareness of both label semantics and image features is proposed in this paper. This approach is designed to dynamically integrate and leverage multi-source sample datasets, providing a rich and diverse set of samples that can effectively support the training of AI models for coastal city and environmental monitoring applications. The main idea of the retrieval approach is demonstrated in Figure 1.

Firstly, in-memory dataset discovery. It conducts in-memory remote sensing sample datasets discovery across diverse platforms

or repositories that host remote sensing sample datasets. By employing Alluxio, a virtual distributed file system across clouds, it can virtually mount a remote sensing sample data repository as a local data directory and offer a dynamic in-memory data cache. Thus, it can offer a unified, easy but efficient way to access the multi-source sample datasets from different platforms transparently. Then, based on a data crawler engine with multi-task management, it implements a data discovery process in parallel on a multi-threaded, Spark-enabled cluster to search remote sensing sample datasets on these virtually mounted and in-memory cached sample data repositories. Each data discovery thread conducts preliminary metadata retrieval and then ingests the target data samples into the local sample repository.

Secondly, remote sensing sample data integration of multi-source sample datasets through interpreting both visual features of imagery and semantic features of sample labels. At first, it abstracts the label and spatial metadata (such as spatial and temporal regions of imagery) of the sample datasets and ingests them into a metadata repository facilitated by PostgreSQL. Then, it incorporates a Bert (Devlin, 2018)-based word embedding model to extract the semantic features from the sample labels and cache them into the label category feature repository. Meanwhile, it also employs a multi-label scene classification model to extract the visual features from the remote sensing sample imagery and store them in the sample feature repository as feature vectors. Following this approach, the sample datasets can be fully understood and integrated into the sample database by parsing and interpreting not only the visual features of imagery but also the semantic features of sample labels.

Thirdly, label semantics and image features aware sample retrieval. First of all, to eliminate the inconsistency of label categories among multi-source samples, it adopts a graph-based dynamic label category system, which uses OSM (OpenStreetMap) semantic network as a fundamental category system. It is stored in a Neo4j-enabled label category repository. In case of category inconsistency, this dynamic label category system can expand new label categories from newly integrated sample datasets through building semantic relations between the existing categories and new labels using label semantic similarity. After that, it builds data indexing of the sample datasets, including the spatial indexing, label semantic feature indexing, and visual feature indexing. Here, the spatial data indexing incorporates a quadtree and Hilbert curve indexing approach. Meanwhile, the label semantic features and image features are indexed by employing the Inverted File Product Quantization (IVF-PQ) encoding algorithm. Then, it proceeds with an intelligent three-step cross-dataset sample query among various integrated datasets in the repository to fully search for the samples that are properly matched both in label semantics and visual features. The first retrieval is a spatial retrieval based on the spatial indexing using spatial conditions. The secondary retrieval is a label-semantic-based retrieval using label semantic similarities to match the samples within the label semantic repository. Then, the third step involves a feature-based imagery retrieval, which identifies similar samples from the sample feature repository by matching image feature vectors using an approximate nearest neighbor



(ANN) matching algorithm. Eventually, it reorganizes the matched remote sensing imagery samples returned from cross-dataset retrieval and packages them as a new application-tailored remote sensing sample dataset. The new supplementary samples obtained through expanded feature retrieval can be further filtered through user interaction for a refined dataset.

3.2 In-memory sample data discovery across clouds

To address the issue of dispersed sample datasets, the proposed method virtualizes remote sensing sample data repositories, like GEE, Kaggle, Paddle and AWS, as local directories using Alluxio, providing dynamic in-memory caching to enable unified, transparent, and efficient access to multi-source datasets. Utilizing the multi-task management data crawler engine Crawlhub, the method employs a distributed computing framework, Spark, to concurrently mine available sample datasets from the mounted data repositories. Subsequently, the directory structure and metadata of these datasets are recorded in Alluxio. Through Alluxio, the sample data is retrieved for subsequent integration.

The integration of sample datasets begins with a breadth-first traversal of the dataset's hierarchical directory structure to acquire the label category system. Once the dataset's label

category system is successfully obtained, each sample instance corresponding to a label category is located within its respective subdirectory. A parallel depth-first traversal is then performed on each label category to retrieve the corresponding sample images. For each accessed sample instance, the visual features of the sample image and the semantic features of the sample label are parsed. The label and spatial metadata (such as the spatial and temporal extent of the imagery) of the sample dataset are abstracted and imported into a metadata repository supported by PostgreSQL. Simultaneously, a BERT-based word embedding model is employed to extract the semantic features of the sample labels, which are then cached in the label category feature repository. Furthermore, a multi-label scene classification model is utilized to extract the visual features from the remote sensing imagery, storing these as feature vectors in the sample feature repository. This approach enables comprehensive understanding and integration of the sample datasets, not only by analyzing the visual features of the imagery but also by interpreting the semantic features of the sample labels.

3.3 Building dynamic label category system based on semantic similarity

Traditional label category systems consist of tree structures containing labels of different granularities. However, the coupling of label information with the dataset directory

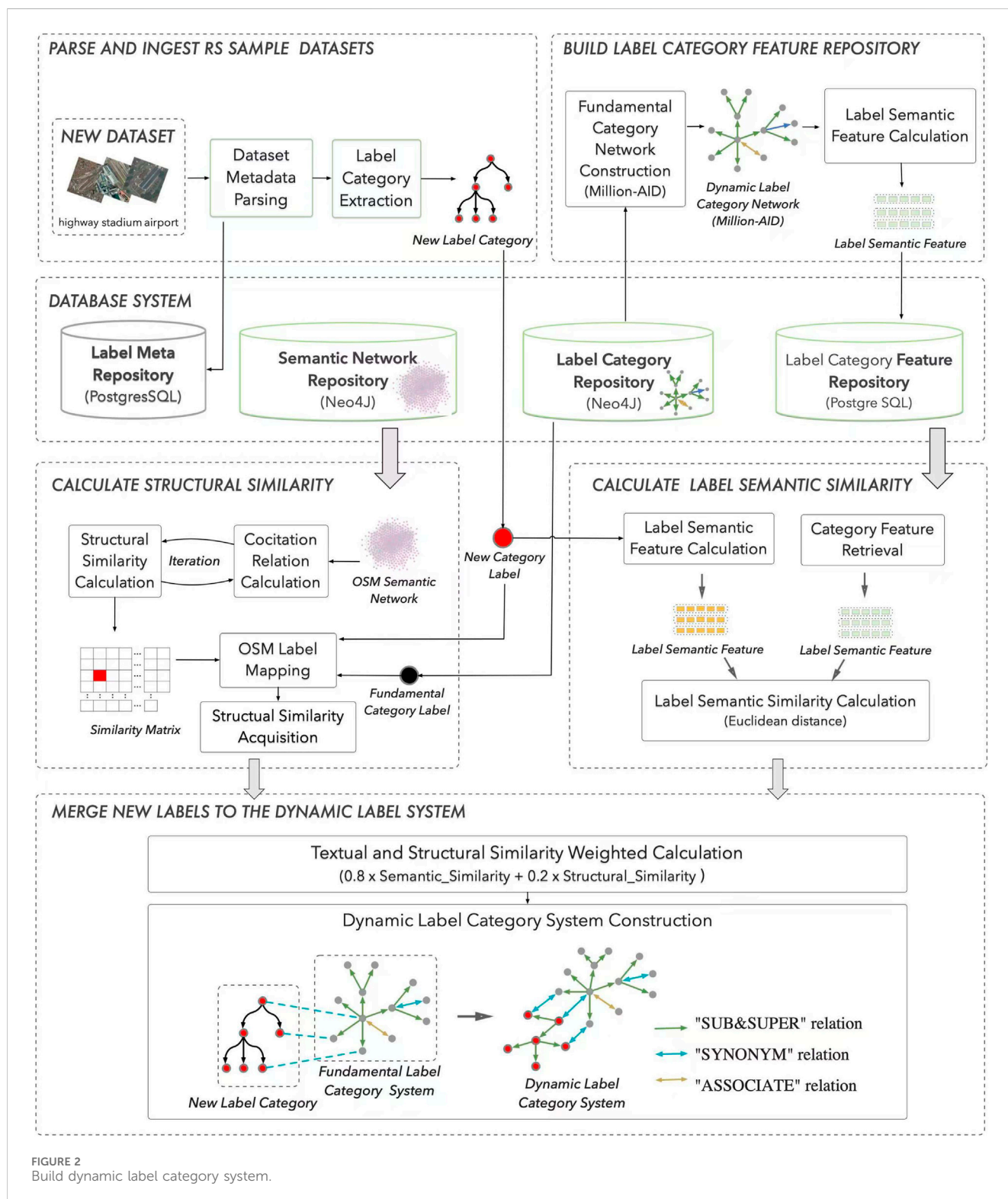


FIGURE 2
Build dynamic label category system.

structure leads to inefficient label management. Moreover, traditional classification systems only describe parent-child relationships between categories, providing a single, insufficient semantic relationship description to handle differences in label category systems across datasets.

To address these differences in label category systems and label representations among current remote sensing sample

datasets, this paper proposes a label category system compatible with multiple sample datasets based on land cover standards such as LCCS and existing open-source sample dataset classification systems (Herold et al., 2006). Labels and their relationships are stored and managed in a graph structure, mapping semantic labels from existing sample datasets to this classification system. This standardizes the label formats of

multi-source sample data, achieving unified organization and management of multi-source sample data labels.

To address synonym labels between different datasets, this paper introduces similarity relationships to establish connections among synonymous label categories. For differences in label granularity across datasets, the graph-based classification system provides more flexible management of label categories through parent-child relationships compared to traditional tree directory structures. For issues where most sample instances contain multiple label categories, the label system additionally introduces spatial adjacency relationships to represent the co-occurrence frequency of different categories within the same sample instance. Li et al. measured the spatial adjacency degree of categories based on content similarity between sample instances of different categories.

Based on the above, this method's label category system can encompass the label categories of most sample datasets and better describe relationships among categories, in addition to being extensible. The construction of the label category system for this method is illustrated in Figure 2.

3.3.1 Calculate label semantic similarity and structural similarity

Label similarity in this study is primarily calculated by weighting the textual similarity of the labels and the structural similarity based on an open-source geographic knowledge graph. As shown in Figure 3, the textual similarity is first enhanced by augmenting the label text with additional descriptive information retrieved from external sources such as Wikipedia. The enhanced label text is then transformed into feature vectors using a Bert-based word embedding model, which excels in

understanding contextual nuances through its bidirectional mechanism, providing an efficient approach to capturing complex language structures. Finally, the cosine similarity between the feature vectors is used to determine the textual similarity of the labels. The structural similarity describes the degree of association between two label entities within the geographic knowledge graph (Yu et al., 2018). In this study, the structural similarity is calculated based on the OSM Semantic Network. The OSM Semantic Network, derived from the OSM Wiki website, is a semantic network that includes a rich set of geographic entities and various semantic relationships between them. The co-occurrence algorithm iteratively computes the degree of association between two label entities based on the relationships among entities in the semantic network (Ballatore et al., 2013). A geographic knowledge graph organizes geographic data through ontologies and semantic links, where an ontology defines categories, attributes, and relationships of geographic entities (e.g., a "city" belongs to the category of "residential areas," and a "river" flows through a "city"). This structured representation enables the geographic semantic network to inherently encapsulate domain-specific knowledge and terminology in geography. In contrast, natural language processing (NLP) models are typically trained on general-purpose language corpora and may not be specifically optimized for geographic terminology and concepts. As a result, text similarity based on NLP models may fail to fully capture geographic semantic information due to a lack of specialized processing for geographic terms and relationships. To address this limitation, this study introduces structural similarity based on geographic knowledge graphs to complement text-based similarity, ensuring the inclusion of geospatial information that might be missing in purely textual representations.

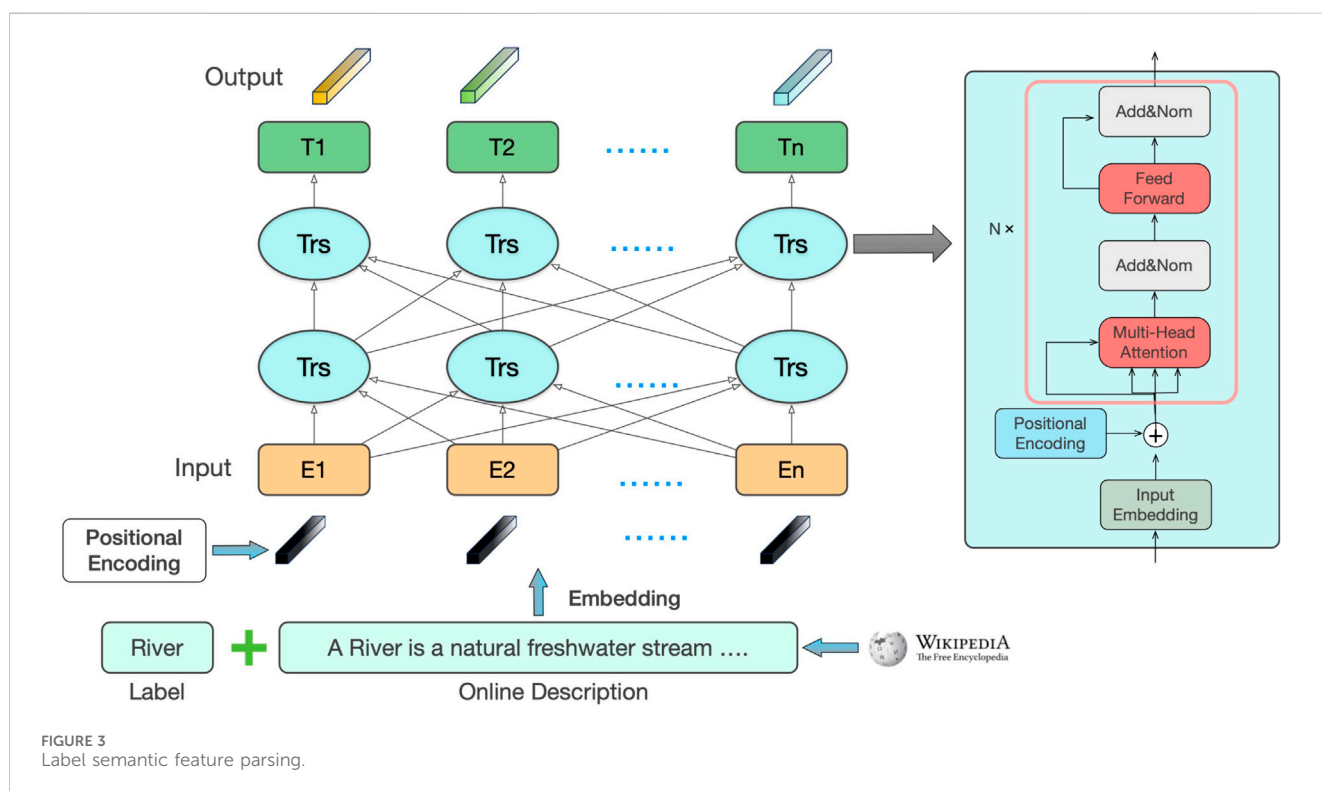


FIGURE 3
Label semantic feature parsing.


```

1: function MERGE_NEW_LABELS_TO_LABEL_CATEGORY_SYSTEM (new_
   classify)
2:   new_labels ← addToClassify (new_classify)
3:   for new_label in new_labels do
4:     sim_labels ← GET_SIMILAR_LABEL (new_label)
5:     for (label,score) in sim_labels do
6:       createSynonymLink(label,new_label,score)
7:     end for
8:   end for
9: end function
10: function GET_SIMILAR_LABEL (label_name)
11:   tag ← getOSMTag (label_name)
12:   descr ← tag.getDescr()
13:   feature ← generateFeature (label_name + descr)
14:   [(sim_label,sim_score)] ← knnSearch(feature)
15:   similarLabelList ← [ ]
16:   for (label,txt_sim) in [(sim_label,sim_score)] do
17:     struct_sim ←
       COCITATION(osm_network, tag.getOSMTag(label))
18:     score = a × txt_sim + b × struct_sim
19:     if score > Threshold then
20:       similarLabelList.append ((label,score))
21:     end if
22:   end for
23:   return similar_label_list
24: end function

```

Algorithm 1. Merge New Labels To Label Category System.

The label semantic matching in this study begins by parsing the label tree of the new dataset and adding it to the dynamic label category system. The new labels are then mapped to the OSM Semantic Network labels, which enhances the label text. The augmented label text is transformed into text features using a word embedding model, and these features are compared for similarity with the text feature library of the labels in the classification system. Subsequently, the co-occurrence algorithm is applied to calculate the structural similarity between the nodes in the OSM network. Finally, the textual and structural similarities are combined using a weighted approach to obtain a similarity score between the new labels and those in the label system. Similar relationships are established between highly similar labels, thus completing the integration of the new dataset's label tree into the label category system.

3.3.2 Merge new label to build dynamic label system through label mapping

In traditional label category systems for datasets, hierarchical relationships between labels of different granularities are typically defined in terms of parent and child categories. Building upon this foundation, this paper introduces additional relationships, such as similarity and spatial adjacency, to further construct the label category system. Yang et al. developed a hierarchical label category system, as illustrated in Figure 4, consisting of 87 categories, where the relationships between categories are restricted to parent-child connections. This paper adopts this system as the baseline label category system, and further extends it by merging label systems from

multiple existing datasets to construct a dynamic label system. When mapping the labels from a new dataset to the classification system, we first measure the similarity between the labels in the new dataset and those in the existing label system. Based on the results of this measurement, we establish similarity relationships between similar labels and record the number of matching samples with content similarities between different labels. If no direct relationship exists between categories, spatial adjacency is assumed once a certain threshold is reached, establishing spatial adjacency relationships between corresponding label pairs.

As shown in Figure 4, the extended dynamic label system is depicted with newly added nodes and relationships highlighted. This system comprises 171 label categories, covering most of the categories found in existing datasets, and supports the inclusion of new label categories. The relationships between these categories include parent-child, semantic similarity, and spatial adjacency, enabling a comprehensive representation of the complex interrelationships between categories. Given the richness of label categories and the complexity of the relationships among them, this paper models the labels in the expanded system as nodes, with the relationships between labels represented as weighted edges. The system is stored and managed using the Neo4j graph database to handle the graph-structured dynamic label category system. Compared to traditional label management methods, which rely on directory names and mapping files, the graph-based label category system allows for detailed modeling of complex relationships between labels, facilitating the learning of dependencies among different labels. Additionally, the graph structure offers scalability, supporting the introduction of new label categories.

3.4 Cross-dataset sample retrieval based on visual and semantic features

To tackle the inconsistencies in labeling semantic categories among datasets, and the limits of the commonly used single-label representation, we proposed a cross-dataset sample retrieval approach based on visual and semantic features. As shown in Figure 5, it begins by retrieving sample metadata that satisfies user-defined label semantics and spatiotemporal constraints. For spatial constraints, efficient spatial retrieval is achieved using a pre-constructed quadtree and Hilbert indexing. The retrieved sample data is then returned as the preliminary result. In the second step, label semantic based retrieval is performed based on the input label constraints. We uses the shortest path algorithm to identify associated labels within the system's label category hierarchy and conducts sample retrieval using the prototype features of these associated labels. The prototype features of a label category are derived from several cluster centers of image features obtained from all sample instances of that label category. Finally, based on the constructed feature index, the image-feature-based sample retrieval method performs feature matching between the prototype features of the associated labels and the image features of all integrated samples. It then conducts parallel approximate matching of samples whose feature

similarity exceeds a given threshold and returns these samples as reference samples.

The proposed semantic feature-based sample retrieval initially retrieves accurately matched sample data based on user-defined sample attribute constraints. Then, related labels are retrieved from the system's label category system, and prototype features corresponding to these related labels are obtained. A prototype feature is a category-level feature representation rather than an individual sample feature. It aims to capture the general characteristics of a given label category by aggregating feature representations of all sample instances belonging to the same category. Typically, multiple cluster centers or a global mean vector are computed to derive a generalized representation of

the category. In this study, extracted sample image features are classified according to their associated label categories to construct a category-level image feature library. Then, the K-means clustering algorithm is applied to compute multiple cluster centers, which are used as the prototype features for label categories. These prototype features are approximately matched with pre-extracted sample image features from the feature library to retrieve several content-similar supplementary samples.

3.4.1 Label semantic feature matching

Traditional sample retrieval methods typically search for samples belonging to specific label categories based on sample labels. However, since most current remote sensing sample

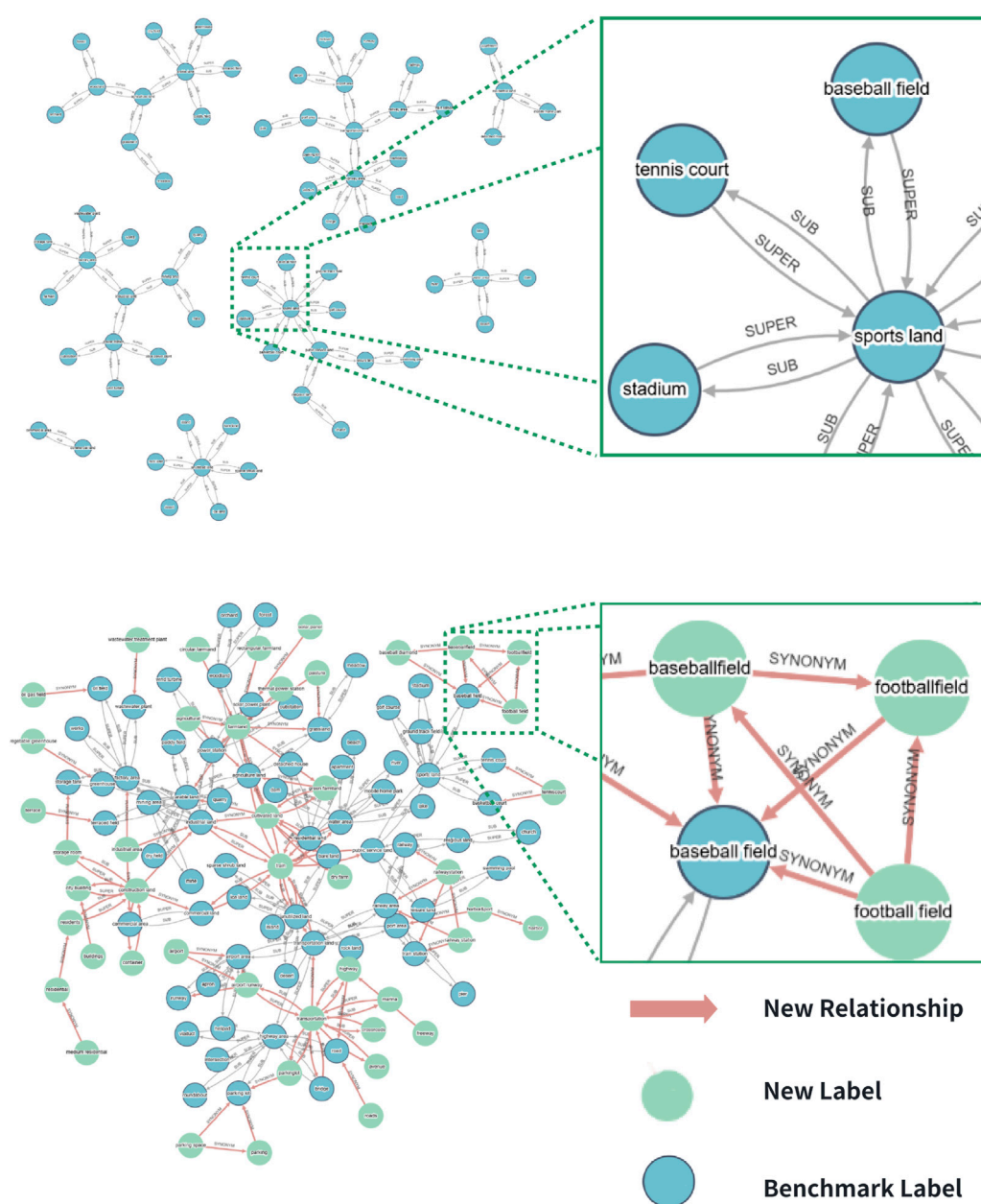


FIGURE 4
Fundamental and dynamic label category system after merging new labels.

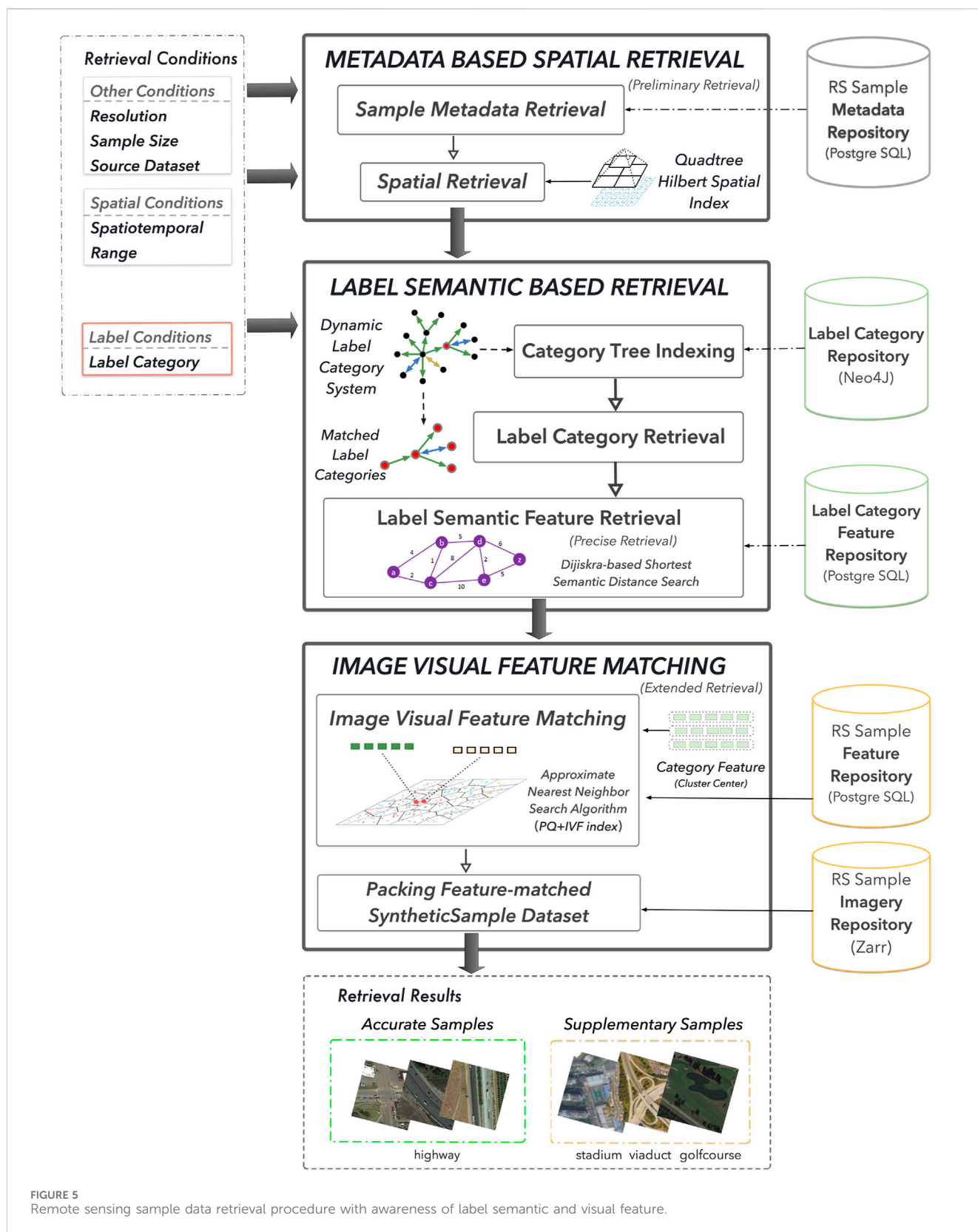


FIGURE 5 Remote sensing sample data retrieval procedure with awareness of label semantic and visual feature.

datasets have only a single label per sample, traditional retrieval methods often overlook other categories present in the sample images. In many cases, samples from certain label categories tend

to co-occur, meaning that images labeled as category A often also contain category B. This co-occurrence is commonly observed when there is a strong association between categories A and B,

such as a parent-child relationship, semantic similarity, or spatial adjacency.

```

1: function CROSS_DATASET_SAMPLE_RETRIEVAL (label_name)
2:   label_graph ← LOAD_LABEL_GRAPH
3:   dis ← DIJKSTRA (label_graph, label_name)
4:
   related_labels ← {label |
                     label ∈ label_graph,
                     dis[label] ≤ MAX_DISTANCE}
5:   supplementary_samples ← {}
6:   for label in related_labels do
7:
       proto_feature ← GET_prototype_
       feature_label
8:
       samples ← KNNSEARCH(feature_db,
       proto_features)
       supplementary_samples [label] ← samples
9:   end for
10:  return supplementary_samples
11: end function
12: function GET_PROTOTYPE_FEATURE (label)
13:  if label in
    cluster_db then
14:    prototype_features ← cluster_db[label]
15:    return prototype_features
16:  else
17:    features ← {feature_db[id] | (id, label)
                  ∈ feature_db.items
                  (), sample_label == label}
18:    if features is not empty then
19:      kmeans_result ← KMEANS (features, 10)
20:      prototype_features ← kmeans_result.
      centroid
21:      cluster_db[label] ← prototype_
      features
22:      return prototype_features
23:    else
24:      return None
25:    end if
26:  end if
27: end function
28: end function

```

Algorithm 2. Cross-Dataset Sample Retrieval.

To better exploit sample data containing instances of a specific label category, this paper extends traditional retrieval methods by using the dynamic label category system developed in Section 3.2. This system identifies additional label categories that are strongly associated with the search label category. Samples in these associated categories are more likely to contain the queried label category.

For example, with label A, the label-semantic-based sample retrieval aims to discover sample data from label categories associated with A that might contain samples of category A. The association weights between categories in the dynamic label category system vary according to the type of relationship. As shown in Equation 1, subclass relationships are assigned a weight of 0,

assuming that subclass samples always include instances of their parent class. The weights for parent-child and semantic similarity relationships are inversely proportional to the semantic similarity between categories: the more semantically similar the labels are, the lower the distance weight. Spatial proximity is computed by counting the co-occurrence of sample categories; if two categories co-occur in the same image above a certain threshold, they are considered to have a spatial adjacency relationship. The more frequent the co-occurrence, the lower the associated distance weight. Specifically, the distance weighting formula between labels is defined as (1):

$$W(A, B) = \begin{cases} 0 & \text{if } B \text{ is a subclass of } A \\ \frac{\alpha}{\text{sim}(A, B)} & \text{if } A \text{ and } B \text{ are semantically similar} \\ \frac{\beta}{\text{co-occurrence}(A, B) + \tau} & \text{if } A \text{ and } B \text{ are spatially adjacent} \\ \infty & \text{otherwise} \end{cases} \quad (1)$$

where α and β are tuning parameters, and τ is a threshold parameter.

Using the label category system with multiple types of association weights, the weighted shortest path algorithm is applied to compute a set of label categories whose distance to category A is below a specified threshold. These labels are considered associated labels for A.

3.4.2 Image visual feature matching

Given that image features better represent complex image content than labels, this paper further combines sample image features with label semantics for sample retrieval. To accommodate the multi-scale characteristics of sample images, this paper employs the multi-label scene classification model, as shown in Figure 6, to extract image features from the sample images. The feature extraction module is responsible for extracting suitable feature representations from the sample images, which typically possess varying scales. To address this, the feature extraction module fuses the outputs from multiple convolutional layers at different levels of the classification model, capturing multi-scale information present in the sample images, thereby obtaining image features that encapsulate multi-scale information. In order to enhance the extracted image features' ability to represent all potential label categories contained within the sample images, this paper further integrates the Binary Cross-Entropy (BCE) loss function and pre-trains the model using the multi-label dataset MLRSNet. The labels of multi-label samples are represented as one-hot vectors, and a weight matrix is constructed based on the similarity of the vectors. Image pairs with more shared labels are assigned higher weights. Through multiple iterations, the model minimizes the distance between similar label samples in the feature space to achieve parameter optimization.

We train an image retrieval model based on existing datasets to simultaneously learn suitable low-dimensional feature encodings, high-dimensional feature dimensionality reduction, and feature matching.

Multispectral remote sensing sample images significantly differ from optical RGB remote sensing images in terms of the number of

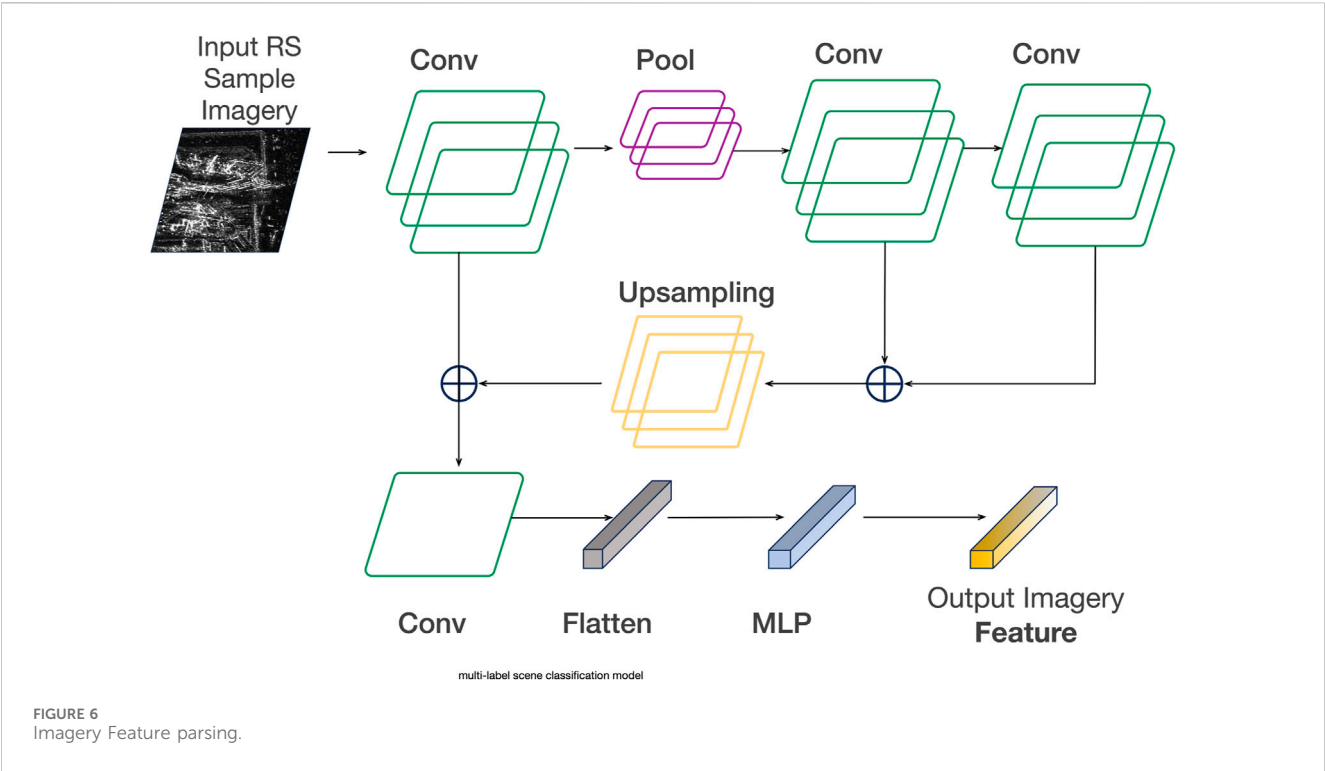


TABLE 2 Datasets overview.

Dataset	Number of classes	Number of instances	Resolution (m)	Sensor type
NWPU-RESISC45	45	31,500	0.2 ~ 30	Optical
AID	30	10,000	0.5 ~ 0.8	Optical
RSI-CB256	35	24,000	0.22 ~ 3	Optical
AID++	46	> 400,000		Optical
PatternNet	38	30,400	0.06 ~ 4.7	Optical
CLRS	25	15,000	0.26 ~ 8.85	Optical
BigEarthNet-MM	19	590,326	10, 20, 60	Multispectral
DFC-15	8	3,342		Optical
OPTIMAL-31	31	1,860		Optical
RSSCN7	7	2,800	0.5	Optical
VArcGIS	38	59,071	0.07 ~ 19.11	Optical
VGoogle	38	59,404	0.075 ~ 9.555	Optical
VBing	38	58,944	0.07 ~ 38.22	Optical
Million-AID	51	1,000,848	0.5 ~ 153	Optical
fmow	63	> 10,000,000	2 ~ 30	Multispectral
WHU-RS19	19	1,005	≤ 0.5	Optical
EuroSAT	10	27,000	10	Optical, Multispectral
MLRSNet	46	109,161	0.1 ~ 10	Optical
RS_C11	11	1,232	0.2	Optical
UCMerced	21	2,100	0.3	Optical

spectral bands and their spectral distribution. While RGB images contain only the visible red (R), green (G), and blue (B) bands, multispectral images typically include multiple spectral bands, such as visible, near-infrared (NIR), and shortwave infrared (SWIR). This disparity introduces certain mismatches in feature extraction and analysis processes between the two data types. Additionally, even within multispectral remote sensing imagery, data collected by different sensors may exhibit variations in band structure, central wavelength, and bandwidth, posing challenges in band alignment and feature unification when conducting cross-sensor analysis. To address these differences, we adopt a band selection and dimensionality reduction fusion approach. First, all multispectral image bands undergo normalization to ensure consistency in data distribution. Next, we configure band mapping files to align the RGB-related visible bands while preserving key non-visible bands (e.g., NIR) to enhance feature expressiveness. Finally, Principal Component Analysis (PCA) is applied to reduce the multispectral bands to three principal components, enabling unified feature extraction between multispectral and optical remote sensing data within the same feature space.

To support fast approximate matching during the retrieval phase, we build an IVF-PQ clustering index for the extracted image features and maintain several feature cluster centers for each label category. The IVF index divides the entire vector space into smaller regions using clustering algorithms such as K-means, where each region is represented by a cluster center. The PQ index quantizes the feature vectors by segmenting them and clustering each segment, replacing each segment with the nearest cluster center. During the approximate matching phase, the IVF index groups the vectors based on the nearest cluster centers, and within each group, the PQ index further compresses the vectors. The quantized vectors are then used for further matching.

Finally, based on the constructed feature index, the image-feature-based sample retrieval method first performs segment-wise quantization on the image features. Subsequently, it calculates the Euclidean distance between the quantized feature vectors and the cluster centroids in the index. The features from the nearest cluster centroids are then selected and returned as the matching results.

3.4.3 Generating application-tailed synthetic remote sensing sample dataset

To meet the increasingly complex requirements of remote sensing applications and provide sample data that accurately matches these needs, this paper employs semantic feature-based sample retrieval to obtain qualifying sample data and construct dynamic datasets based on the application-specific dataset label category sets and sample attributes. According to the user-selected label category sets and other sample attribute constraints, the process first splits the task into multiple parallel sample retrieval subtasks based on the label category sets. Each task retrieves metadata for samples that meet the criteria, and these metadata are then used to construct a logical dataset. The retrieval results include accurately matched samples from the initial search and supplementary samples obtained through feature matching. Users interactively decide whether to retain these supplementary samples in the dataset, leading to the dynamic construction of the sample dataset upon completion of the interaction. The system interface

presents a visual preview of all retrieved samples in a list format, enabling users to quickly assess whether the supplementary samples contain the desired labels. Users can then complete the interactive selection by checking the relevant samples within the list. Additionally, the filtered-out supplementary samples are recorded as user feedback, which serves as a reference for subsequent refinement of the image feature extraction model. The label categories included in the dynamic dataset are entirely specified by the user, thus meeting the needs of most deep learning applications. Furthermore, the sample data for the dynamic dataset come from various source datasets, and the heterogeneity of samples across different datasets naturally endows the dynamic dataset with high intra-class diversity and broad spatiotemporal distribution.

4 Experiment

In this paper, we design and implement a cross-cloud dynamic sample repository that provides a unified and efficient access method for distributed sample datasets. By constructing a dynamic label category system and implementing sample retrieval based on image features, we enable the customized generation of sample datasets tailored to specific application requirements. This system maximally exploits available sample data, helping to mitigate the issue of scarce sample data to some extent.

To evaluate the system's performance, we conducted a series of experiments, as outlined below:

- (1) External dataset integration performance evaluation.
- (2) Time performance and accuracy evaluation of supplementary sample retrieval based on feature mining.
- (3) Comparison of dynamic sample datasets generated by the cross-cloud dynamic sample repository for specific applications with general scene classification datasets.

The experiments were conducted across five computing nodes, each equipped with a processor featuring a four-core CPU (3.0 GHz), 8 GB of memory, running CentOS 7.0. The GPU model used was the RTX 3090, with a memory capacity of 24 GB.

4.1 Experimental datasets

The current sample repository contains approximately 20 remote sensing (RS) sample datasets, encompassing 178 label categories. The sample distribution spans the globe and includes data from optical and multispectral sensor types. The total number of sample instances is 1,397,210. Detailed information about the RS sample datasets included in the repository is provided in [Table 2](#).

4.2 Dataset integration performance experiment

Dataset integration aims to add new sample datasets from external data platforms to the sample repository, involving

TABLE 3 Summary of datasets with storage and integration details.

Dataset name	Number of instances	Storage size	Image format	Image size	Storage platform	Integration time
UCMerced	2,100	424MB	tif	256	Local	1 min
NWPU-RESISC45	31500	470MB	jpg	256	Kaggle	11 min
Million-AID-mini	10000	3.7GB	jpg	512	Local	5 min
EuroSAT	27000	134MB	jpg	64	Local	8 min
MLRSNet	109161	1.6GB	jpg	256	Local	32 min
fmow	132716		tif	74–16184	AWS	48 min
RSI-CB256	14052	2.7GB	tif	256	Paddle	7 min

several key steps: dataset mounting, label category system parsing, sample feature extraction, and feature index updating. In this section, to evaluate the system's performance in integrating external datasets, we assess the dataset integration process across various scales and data source platforms. The datasets are sourced primarily from remote cloud platforms and local disks, with examples including Amazon AWS, Kaggle, and PaddlePaddle. Sample data is retrieved from remote platforms via client applications and platform APIs. The size of the datasets is measured by the number of sample instances and the storage space occupied by the dataset. The experimental results are as follows:

From the experimental results shown in Table 3, it is observed that as the number of sample instances increases, the integration time grows linearly. However, the storage space of the sample dataset has minimal impact on the integration time. This is because the majority of the time consumption during dataset integration occurs during the feature extraction process of sample images. Since the model requires fixed image dimensions, the input images undergo resampling, which ensures that the performance of feature extraction is not significantly influenced by the size of the sample images. For a dataset with 100,000 instances and a total size of 3.7 GB, the integration time is approximately 30 min. The integration time for remote platforms is slightly longer than for local datasets due to network transmission overhead, but the system is still capable of quickly integrating the majority of datasets.

This efficiency is attributed to the use of the distributed memory layer Alluxio, which caches some sample images for pre-mounted datasets. Additionally, sample images stored in the Zarr array database format allow for chunked and parallel loading, resulting in faster image loading speeds. Furthermore, both dataset parsing and sample image feature extraction steps are performed using multi-node and multi-threaded parallelization, further reducing time consumption.

4.3 Supplementary sample retrieval experiment

Supplementary sample retrieval aims to extract usable samples from those with mismatched labels. The system retrieves associated label categories from the dynamic label category system based on

one or more input labels and then mines content-similar samples from other label categories using feature matching, based on prototype features of the associated label categories. A supplementary sample retrieval example is provided in Figure 7. This section evaluates the proposed retrieval method in terms of both time performance and the effectiveness of the retrieved supplementary samples.

4.3.1 Evaluation of retrieval time efficiency

To validate the time performance of supplementary sample retrieval, this experiment measures the time consumption of supplementary sample retrieval under different dataset sizes. Additionally, the experiment evaluates the system's retrieval performance in high-concurrency scenarios by incrementally increasing the number of requests per unit of time.

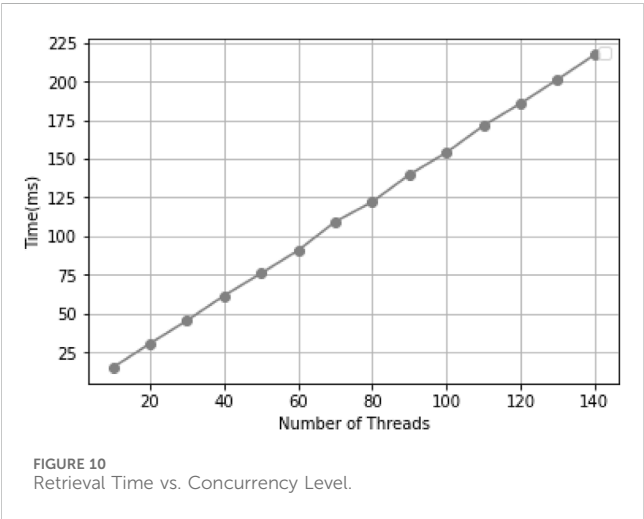
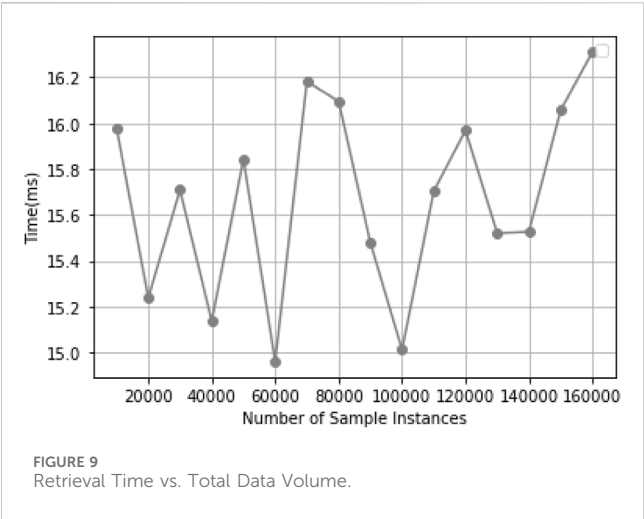
According to the experimental results shown in Figure 8, as the total data volume of the sample repository increases, the time consumption for supplementary sample retrieval does not exhibit significant changes. This is due to the efficient PQ-IVF feature index. The IVF algorithm partitions the feature set into multiple clusters. As the dataset size increases, the number of samples within each cluster increases, while the number of clusters remains relatively fixed. As a result, even as the data volume grows, the number of clusters involved in the retrieval remains essentially constant, leading to stable query time consumption. Moreover, the PQ algorithm further quantizes and reduces the dimensionality of feature vectors, reducing the storage space and the computational complexity of similarity measures. This prevents a significant increase in query time as the dataset size grows.

From Figure 9, it is observed that retrieval time increases linearly with parallelism. When the parallelism reaches 140, the retrieval time is 224 m, indicating that the supplementary sample retrieval system can efficiently discover available samples even under high parallelism conditions.

4.3.2 Evaluation of retrieval performance

As shown in Figure 10, the supplementary samples retrieved through the system are highly likely to contain samples belonging to the corresponding label categories. These supplementary samples cover a variety of complex scene categories, and their label categories exhibit spatial adjacency to the queried labels, such as "river" and "bridge," or "viaduct" and "highway."





Supplementary sample retrieval relies on the label category system and sample image features. To evaluate its performance, we construct a label category system using 20 scene classification datasets, each containing multiple label categories. Prototype features are extracted from sample image features, with cluster centers representing each category. Using this system, we assess supplementary sample retrieval on a mixed sample set from three multi-label datasets: UCMerced_multi_label, AID_multi_label, and DCF. Multi-label datasets provide a more comprehensive representation of complex scenes than single-label samples, making them more suitable for evaluating retrieval effectiveness.

Experiments were conducted on this mixed sample set to assess cross-dataset generalizability, i.e., the ability to handle samples from different datasets. Since dataset samples may vary, this setup provides a comprehensive evaluation in a multi-source heterogeneous data environment. The experiment focuses on 10 common label categories, present across datasets but differing in textual representation. For each category, associated labels are retrieved, prototype features are obtained, and feature matching is performed to retrieve supplementary samples. Accuracy and

TABLE 4 Accuracy and Precision comparison across different datasets.

	UCMerced_multi_label	AID_multi_label	DCF
Accuracy	64.61%	60.44%	75.17%
Precision	57.45%	73.15%	25.28%

precision are calculated based on whether retrieved samples contain the queried label within their associated label categories. Since retrieval spans multiple datasets, samples are grouped by dataset, and accuracy is computed separately for each group. A weighted overall accuracy is then derived, with dataset weight proportional to the number of positive instances of the target label. Table 4 presents the overall accuracy.

Results indicate that cross-dataset supplementary sample retrieval effectively identifies potential usable samples. However, supplementary sample accuracy is limited and requires manual verification. Precision varies across datasets due to differences in sample quantity, with the DCF dataset, having fewer instances, showing the lowest precision. Given that the goal is maximizing usable sample discovery, a lower precision is acceptable.

4.4 Dataset application experiment

The dynamic sample repository presented in this paper aims to generate dynamic sample datasets that match application requirements, particularly in scenarios where sample data is scarce. In this section, to comprehensively evaluate the effectiveness of the generated dynamic sample datasets, we compare the performance of the new dataset with that of other datasets of varying scales across multiple model architectures.

Three commonly used network architectures—VGGNet16, ResNet50, and DenseNet—are employed in this study. By comparing the performance of these convolutional neural network (CNN) architectures on the new dataset, we can fully assess the dataset’s effectiveness and the strengths of different models. These network architectures are pre-trained on three widely-used general scene classification datasets—AID, UCMerced, and NWPU-RESISC45—as well as on the dynamic dataset. Transfer learning experiments are conducted on the Million-AID dataset (Long et al., 2021), and the classification accuracy and precision of the models are validated on the validation set. By comparing the classification performance of the same network architectures after training on samples from different datasets, the performance of the datasets is evaluated. UCMerced, AID, and NWPU-RESISC45 represent small, medium, and large-scale datasets in the remote sensing scene classification domain. They differ in terms of label sets, image resolution, and data sources. These datasets have been widely used in various studies, and comparing them with the dynamic dataset allows us to more clearly demonstrate the practical effectiveness of the dynamic dataset, while also ensuring high generalizability and reproducibility of the experimental results.

The label set of the dynamic dataset is constructed by mapping the labels of the aforementioned datasets. The generated dataset

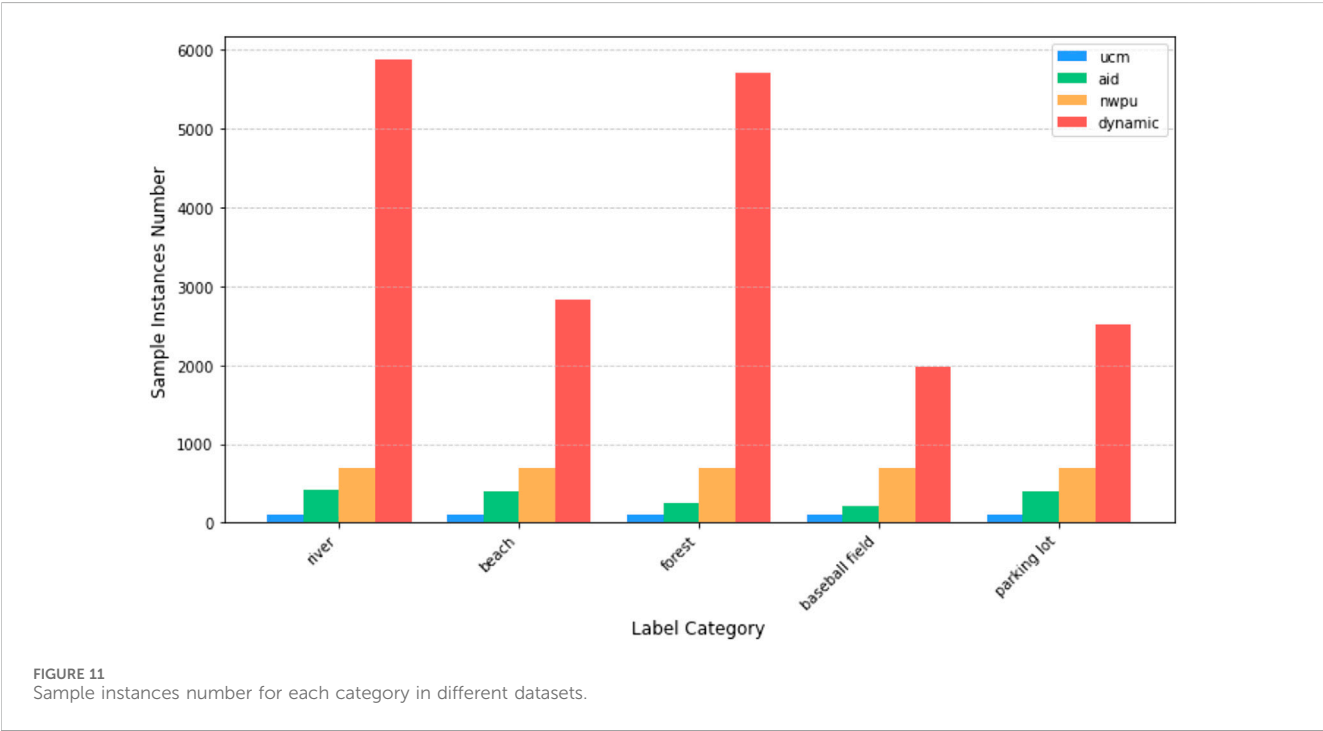


TABLE 5 Accuracy and Precision comparison across different models.

Dataset	Classification model	Accuracy	Precision
UCMerced LandUse	ResNet50	93.30%	89.15%
	VGG16	56.80%	36.54%
	DenseNet	91.9%	85.48%
NWPU-RESISC45	ResNet50	95.25%	90.06%
	VGG16	69.75%	52.78%
	DenseNet	94.4%	87.97%
AID	ResNet50	91.10%	84.84%
	VGG16	71.85%	55.96%
	DenseNet	93.2%	88.46%
Dynamic Dataset	ResNet50	96.30%	92.49%
	VGG16	72.90%	54.18%
	DenseNet	97.05%	94.23%

contains 70 label categories, with each category having between 100 and 5,900 sample instances. As shown in Figure 11, there are notable differences in the number of sample instances between some common categories across the four datasets. The dynamic dataset has more sample instances within each category compared to the other datasets.

The Million-AID dataset, a larger-scale scene classification dataset, includes label categories that cover the three datasets mentioned above. Validating the transfer-learned models on Million-AID better simulates the model’s performance in real-world scenarios where the model faces various unknown data. This allows for a more comprehensive assessment of the impact of the dataset on model performance.

As shown in Table 5, the classification accuracy and precision after pre-training with the dynamic dataset outperforms the results on UCMerced, AID, and NWPU-RESISC45 across all three network architectures. Moreover, the dynamic dataset shows superior performance in more complex classification models, indicating that models pre-trained on the dynamic dataset exhibit better generalization and classification precision. This is due to the higher number of sample instances in the dynamic dataset, as well as the increased intra-class diversity resulting from samples coming from multiple datasets with variations in lighting conditions, viewpoints, and backgrounds. As a result, the dynamic dataset enables the training of more robust classification models.

The results demonstrate that the dynamic dataset outperforms datasets such as UCMerced and AID, particularly when using complex classification models. This improvement can be attributed to the higher number of sample instances in the dynamic dataset and the significant inter-class variability among samples from different datasets, which contributes to training more robust classification models.

However, the current remote sensing scene classification field still suffers from a lack of diverse datasets. The number of remote sensing sample instances is insufficient to bridge the differences in sample styles and dataset scales between different datasets. Consequently, the generated dynamic dataset exhibits some performance differences compared to MillionAID and MLRSNet.

5 Conclusion

Remote sensing sample data, especially those used for coastal city and environmental monitoring, currently face challenges such as dispersion, inconsistency in dataset structures and labeling systems, and insufficient representation capability of sample labels. To address these issues, this paper presents an intelligent remote sensing sample dataset retrieval approach with awareness of label semantics and visual features. By leveraging distributed storage, a dynamic labeling system, and a hybrid retrieval mechanism that integrates textual and image-based label features, our approach effectively consolidates heterogeneous remote sensing sample data and enables efficient data management. Experimental results demonstrate that the proposed multi-source data unification strategy based on a dynamic labeling system significantly enhances the compatibility of diverse datasets while improving the accuracy of data retrieval and sample matching in intelligent remote sensing analysis tasks. However, the current system still has certain limitations. For instance, it lacks support for multimodal remote sensing data, as it does not yet fully incorporate SAR, speech, and text modalities. Additionally, the adaptability of the labeling system needs further improvement to better accommodate various application scenarios and remote sensing tasks. Moreover, discrepancies remain between the data distribution in our system and real-world environments. The existing sample datasets cannot fully simulate real-world conditions, particularly in extreme environments (e.g., polar regions, high-altitude observations, and nighttime remote sensing) or specific tasks (e.g., disaster monitoring and underground imaging). These limitations may affect the system's generalization ability in complex remote sensing applications. Future research will focus on the following three key directions: *First*, integration of multimodal remote sensing data: Incorporate SAR, speech, and textual data while exploring cross-modal feature alignment techniques to enhance the fusion of multi-source information, thereby supporting more complex intelligent remote sensing analysis tasks. *Second*, development of an adaptive label evolution system: Utilize knowledge graphs and large-scale vision-language pre-trained models to enable dynamic label expansion and automated normalization, improving the scalability and intelligence of the labeling system. *Third*, optimization of remote sensing sample data distribution: Continuously integrate new sample datasets to expand the

number of sample instances, ensuring that data distributions better approximate real-world scenarios, thereby enhancing the system's adaptability in complex remote sensing environments. With the incorporation of multimodal data, optimization of the intelligent labeling system, and improvements in data distribution, the proposed system is expected to further advance the organization and management of remote sensing sample data, providing a highly efficient, accurate, and scalable data foundation for future intelligent remote sensing analysis.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <http://weege.vision.ucmerced.edu/datasets/landuse.html><https://captain-whu.github.io/DiRS/><https://gcheng-nwpu.github.io/#Datasets>.

Author contributions

XR: Formal Analysis, Methodology, Software, Validation, Writing – original draft, Writing – review and editing. YM: Funding acquisition, Investigation, Resources, Writing – review and editing. YZ: Investigation, Project administration, Visualization, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the National Natural Science Foundation of China: 42071413, and the Science and disruptive technology project of Aerospace Information Research Institute of Chinese Academy of Sciences, grant number E2Z206010F.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Ballatore, A., Bertolotto, M., and Wilson, D. C. (2013). Geographic knowledge extraction and semantic similarity in openstreetmap. *Knowledge Inf. Syst.* 37, 61–81. doi:10.1007/s10115-012-0571-0
- Cao, Z., Jiang, L., Yue, P., Gong, J., Hu, X., Liu, S., et al. (2023). A large scale training sample database system for intelligent interpretation of remote sensing imagery. *Geo-Spatial Inf. Sci.* 27, 1489–1508. doi:10.1080/10095020.2023.2244005
- Chaudhuri, B., Demir, B., Chaudhuri, S., and Bruzzone, L. (2017). Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method. *IEEE Trans. Geoscience Remote Sens.* 56, 1144–1158. doi:10.1109/tgrs.2017.2760909
- Chen, Z.-M., Wei, X.-S., Wang, P., and Guo, Y. (2019). “Multi-label image recognition with graph convolutional networks,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5177–5186.
- Cheng, G., Xie, X., Han, J., Guo, L., and Xia, G.-S. (2020). Remote sensing image scene classification meets deep learning: challenges, methods, benchmarks, and opportunities. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 13, 3735–3756. doi:10.1109/jstars.2020.3005403
- Dang, K. B., Vu, K. C., Nguyen, H., Nguyen, D. A., Nguyen, T. D. L., Pham, T. P. N., et al. (2022). Application of deep learning models to detect coastlines and shorelines. *J. Environ. Manag.* 320, 115732. doi:10.1016/j.jenvman.2022.115732
- Devlin, J. (2018). Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv Prepr. arXiv:1810.04805*. doi:10.48550/arXiv.1810.04805
- Gominski, D., Gouet-Brunet, V., and Chen, L. (2022). “Cross-dataset learning for generalizable land use scene classification,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1382–1391.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R. (2017). Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. doi:10.1016/j.rse.2017.06.031
- He, S., Zhou, B., Wang, B., Ding, Z., and Zhang, J. (2024). Land classification research based on the improved deeplabv3+ model and the ai earth remote sensing cloud platform. *IEEE Access* 12, 182516–182531. doi:10.1109/ACCESS.2024.3468384
- Herold, M., Woodcock, C. E., Di Gregorio, A., Mayaux, P., Belward, A. S., Latham, J., et al. (2006). A joint initiative for harmonization and validation of land cover datasets. *IEEE Trans. Geoscience Remote Sens.* 44, 1719–1727. doi:10.1109/tgrs.2006.871219
- Iglovikov, V., Mushinskiy, S., and Osin, V. (2017). Satellite imagery feature detection using deep convolutional neural network: a kaggle competition
- Jin, P., Xia, G.-S., Hu, F., Lu, Q., and Zhang, L. (2018). “Aid++: an updated version of aid on scene classification,” in *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium (IEEE)*, 4721–4724.
- Lee, C.-W., Fang, W., Yeh, C.-K., and Wang, Y.-C. F. (2018). “Multi-label zero-shot learning with structured knowledge graphs,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1576–1585.
- Li, H., Dou, X., Tao, C., Wu, Z., Chen, J., Peng, J., et al. (2020). Rsi-cb: a large-scale remote sensing image classification benchmark using crowdsourced data. *Sensors* 20, 1594. doi:10.3390/s20061594
- Li, P., Chen, P., and Zhang, D. (2022). Cross-modal feature representation learning and label graph mining in a residual multi-attentional cnn-lstm network for multi-label aerial scene classification. *Remote Sens.* 14, 2424. doi:10.3390/rs14102424
- Li, Y., Zhang, H., Xue, X., Jiang, Y., and Shen, Q. (2018). Deep learning for remote sensing image classification: a survey. *Wiley Interdiscip. Rev. Data Min. Knowledge Discov.* 8, e1264. doi:10.1002/widm.1264
- Long, Y., Xia, G.-S., Li, S., Yang, W., Yang, M. Y., Zhu, X. X., et al. (2021). On creating benchmark dataset for aerial image interpretation: reviews, guidances, and million-aid. *IEEE J. Sel. Top. Appl. earth observations remote Sens.* 14, 4205–4230. doi:10.1109/jstars.2021.3070368
- Lv, Q., Wang, Q., Song, X., Ge, B., Guan, H., Lu, T., et al. (2024). Research on coastline extraction and dynamic change from remote sensing images based on deep learning. *Front. Environ. Sci.* 12, 1443512. doi:10.3389/fenvs.2024.1443512
- Peng, C., Li, Y., Jiao, L., and Shang, R. (2020). Efficient convolutional neural architecture search for remote sensing image scene classification. *IEEE Trans. Geoscience Remote Sens.* 59, 6092–6105. doi:10.1109/tgrs.2020.3020424
- Qi, X., Zhu, P., Wang, Y., Zhang, L., Peng, J., Wu, M., et al. (2020). Mlrnsnet: a multi-label high spatial resolution remote sensing dataset for semantic scene understanding. *ISPRS J. Photogrammetry Remote Sens.* 169, 337–350. doi:10.1016/j.isprsjrs.2020.09.020
- Qi, Y. (2024). Evaluation of urbanization quality based on deep learning and intelligent algorithms. *Int. J. High Speed Electron. Syst.*, 2540144. doi:10.1142/s0129156425401445
- Shao, Z., Zhou, W., Deng, X., Zhang, M., and Cheng, Q. (2020). Multilabel remote sensing image retrieval based on fully convolutional network. *IEEE J. Sel. Top. Appl. Earth Observations Remote Sens.* 13, 318–328. doi:10.1109/jstars.2019.2961634
- Sumbul, G., Charfuelan, M., Demir, B., and Markl, V. (2019). “Bigearthnet: a large-scale benchmark archive for remote sensing image understanding,” in *IGARSS 2019-2019 IEEE international geoscience and remote sensing symposium (IEEE)*, 5901–5904.
- Tang, X., Wu, S., Hou, J., and Chen, G. (2024). Efficient sample retrieval techniques for multimodal model training. *J. Softw.* 35, 1125–1139. In Chinese.
- Van Etten, A., Lindenbaum, D., and Bacastow, T. M. (2018). Spacenet: a remote sensing dataset and challenge series. *arXiv Prepr. arXiv:1807.01232*.
- Yang, Y., and Newsam, S. (2010). “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*, 270–279.
- Yu, L., Qiu, P., Liu, X., Lu, F., and Wan, B. (2018). A holistic approach to aligning geospatial data with multidimensional similarity measuring. *Int. J. digital earth* 11, 845–862. doi:10.1080/17538947.2017.1359688
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., et al. (2020). Deep learning in environmental remote sensing: achievements and challenges. *Remote Sens. Environ.* 241, 111716. doi:10.1016/j.rse.2020.111716