Report

To develop a Content-Based Image Retrieval (CBIR) system suitable for multi-resolution remote sensing imagery, three primary datasets were analyzed and integrated to formulate a robust, generalised dataset. This report delineates the datasets utilized, the methodologies adopted, and the outcomes of the preliminary implementations.

Datasets Overview (data-analysis. ipynb)

1. **FAIR1M:** A benchmark dataset for fine-grained object recognition within high-resolution remote sensing imagery. For simplicity in developing a foundational CBIR system, the object bounding boxes were omitted. Instead, the folder names were used as labels for all images within respective folders, which included categories such as Airplane, Neighborhood, and Ship. This approach facilitates adaptability across various datasets designated for similar tasks **(DOI: 10.48550/arXiv.2103.05569)**.

2. **RESISC45:** Provided by Northwestern Polytechnical University, this publicly accessible benchmark dataset supports Remote Sensing Image Scene Classification (RESISC). It comprises 31,500 images distributed across 45 scene classes, each containing 700 images. Key scenes within this dataset include Airplanes, Bridges, and Ships.

3. **Sentinel-2:** Utilized for its distinct classification of maritime scenes labelled as 'Ship' and 'No Ship,' with varied backgrounds enhancing the dataset's diversity and utility in specialised scenarios.

4. **AID:** The Aerial Image Dataset (AID) supplements our data sources by providing 30 classes ready for aerial scene classification. This dataset was pivotal for model validation and accuracy assessment, particularly when the initial datasets proved insufficient.

Methodology and Implementation

The initial phase involved integrating the above datasets to create a unified dataset containing multi-resolution images. This amalgamation allowed for a comprehensive analysis of scene class distribution, aiding in the nuanced development of our CBIR system. This integration and initial analysis were conducted using a Jupyter Notebook titled data-analysis.ipynb.

For the CBIR system's inference, two primary models were employed:

Torchvision with ResNet50: Leveraged as a feature extraction mechanism, this model, pre-trained on IMAGENET1K_V2, contains 23,542,865 parameters. It aids in creating feature vectors for images, subsequently used to estimate similarities and identify the k-nearest neighbours for any query image.

Report

SWIN Models: With 87,930,848 parameters, the SWIN models were tested for their efficacy in handling large parameter spaces and complex image datasets.

Additionally, the OpenAI CLIP model, a vision-language model, was utilized in an explorative approach. Its performance was exemplary, particularly when tested against both initial and AID datasets. The queries in this model were text-based (e.g., 'airplane'), showcasing the potential of integrating natural language processing with image retrieval.

Validation and Results

The accuracy of the CBIR system was gauged using conventional metrics such as accuracy scores and confusion matrices. These were instrumental in validating the models' performance and were detailed within another Jupyter Notebook titled **image-search.ipynb.**

Conclusion

The preliminary results underscore the viability of using deep learning for CBIR systems in remote sensing applications. Both employed models demonstrated substantial potential in searching and analyzing images across multiple resolutions. However, to enhance the system's effectiveness, it is recommended that the pre-trained models undergo fine-tuning specific to multi-resolution remote sensing imagery. This two-page report encapsulates the project's current scope and foundational findings, setting the stage for further refinement and scaling.