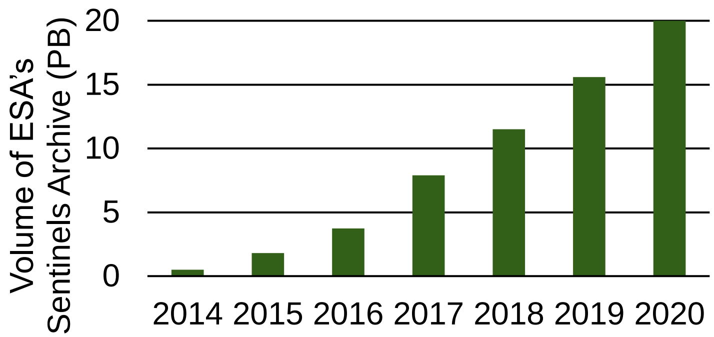
Deep Hashing for Scalable Remote Sensing Image Retrieval in Large Archives

Prof. Dr. Begüm Demir
Remote Sensing Image Analysis (RSiM) Group, TU Berlin
Email: demir@tu-berlin.de

Introduction: Space Renaissance

- ✓ Recent Earth Observation (EO) satellite missions have led to a significant growth of EO image archives.



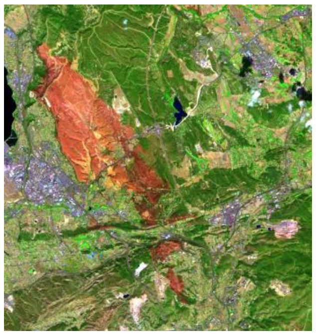
- ✓ Thus, development of efficient and accurate systems for information discovery within massive EO image archives is a growing research interest in remote sensing.

Sumbul et al, "Deep Learning for Image Search and Retrieval in Large Remote Sensing Archives", to appear as a book chapter in "Deep Learning for the Earth Sciences", John Wiley & Sons, 2020.

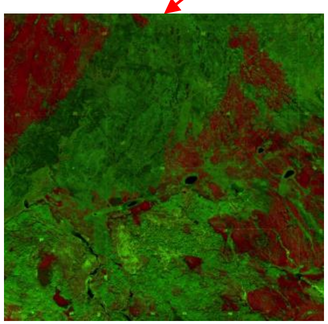
Information Discovery: Query by Example

EO image search/retrieval systems aim to explore crucial information from huge EO data archives.

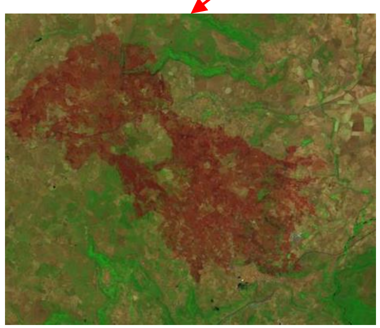
Burned Forest



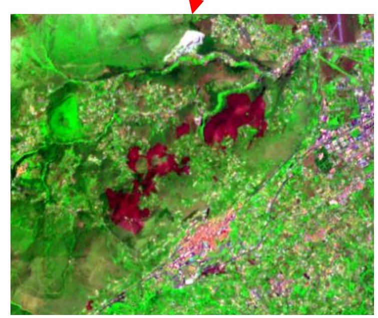
Marseille, France-22.09.2016



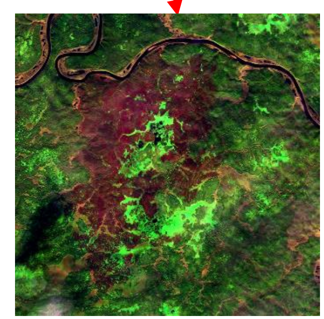
McMurray, Canada-30.08.2016



Sardinia, Italy-28.07.2016

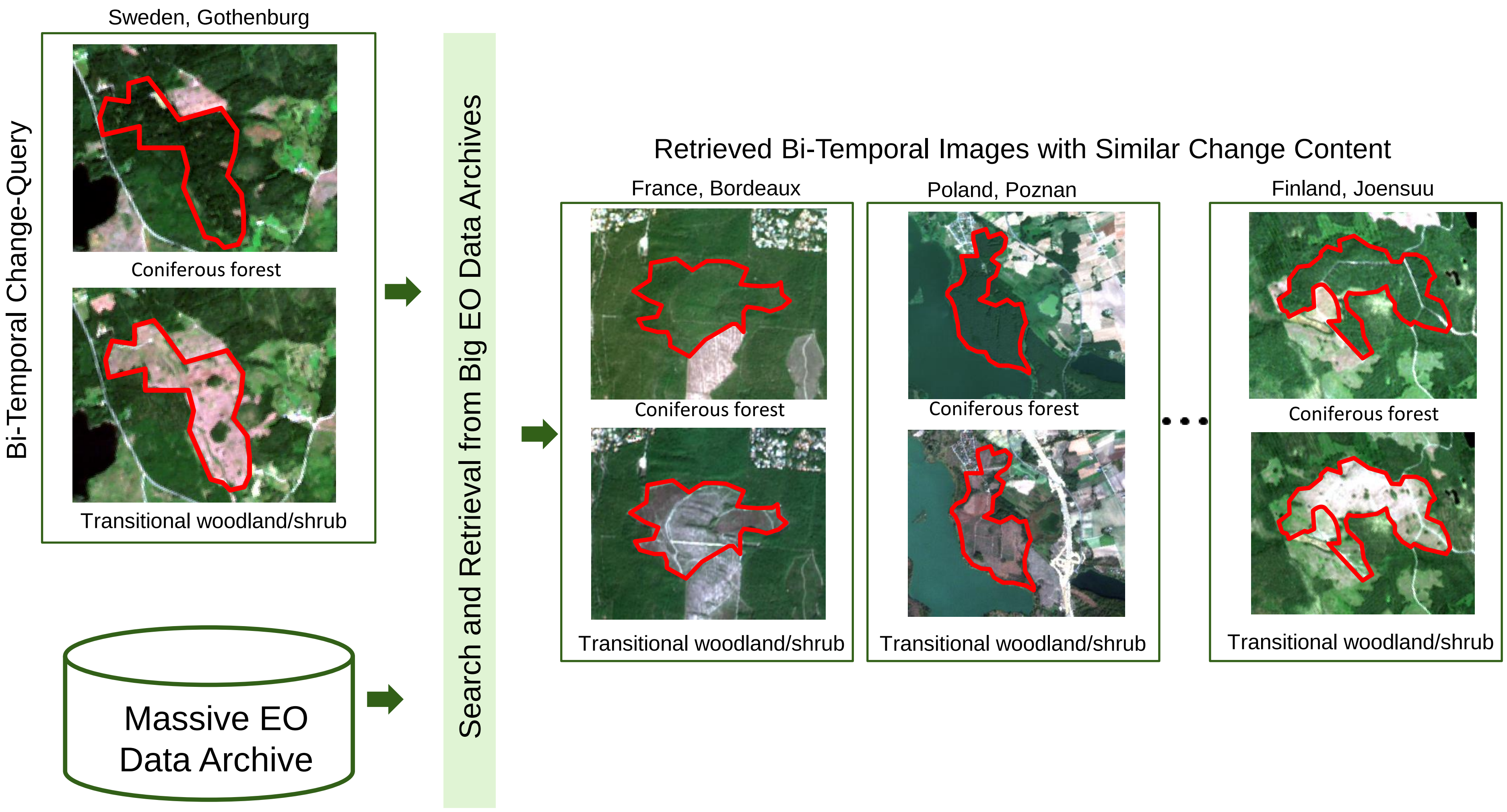


Sicily, Italy-25.05.2016

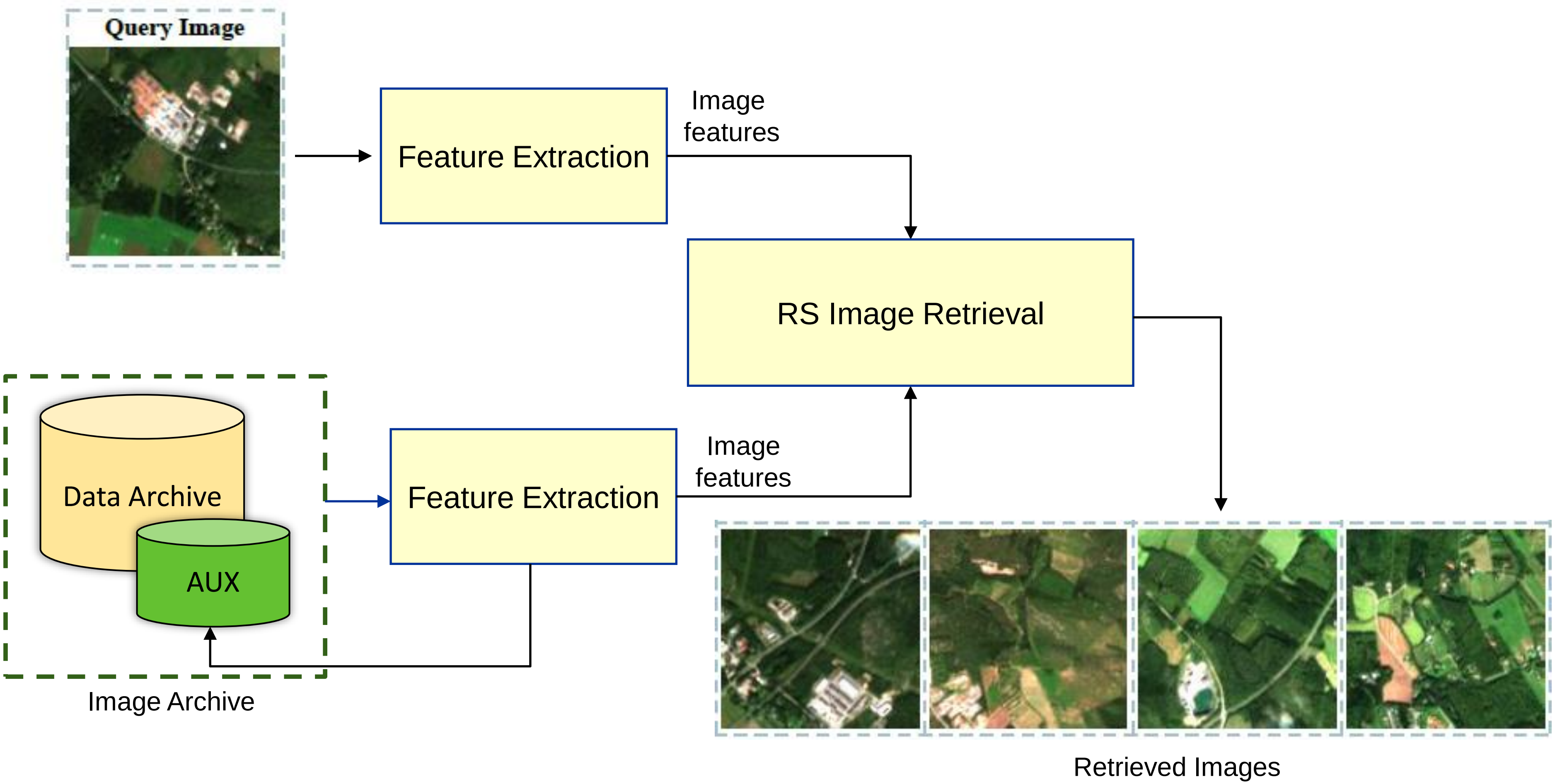


Siberia, Russia-08.10.2016

Information Discovery: Query by Example



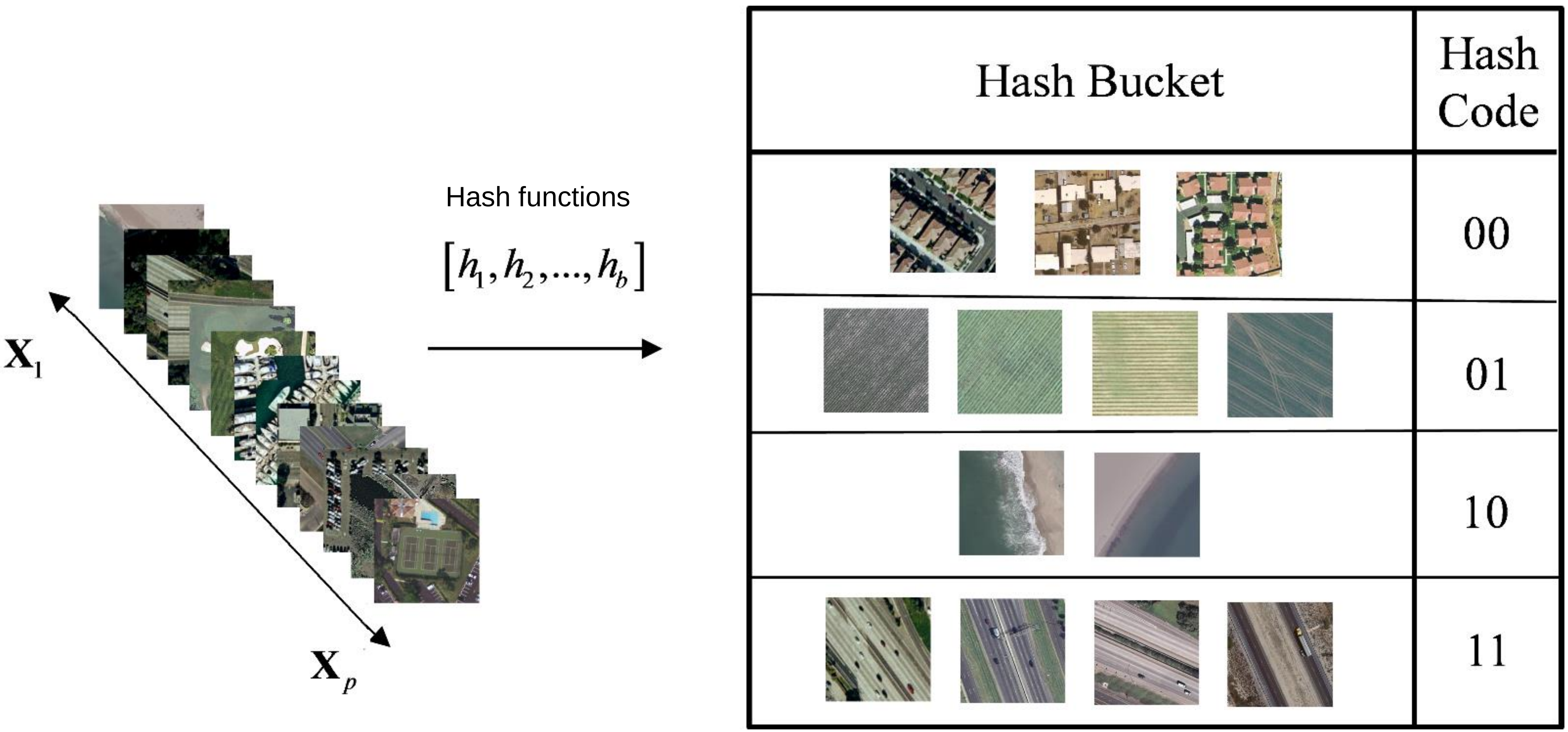
Query by Example: Traditional Systems



- ✓ The traditional systems separately optimize feature learning and image retrieval.
- ✓ Search complexity is $O(n)$ and storage complexity is $O(nd)$, where n is number of images in the archive and d is the image feature number.

  Problem: exhaustive search in huge remote sensing archives is time-demanding.

Hashing Methods in Image Retrieval



Demir et al, "Hashing based scalable remote sensing image search and retrieval in large archives", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no.2, pp. 892-904, 2016.

Kernel-based Hashing Methods

- ✓ Two popular hashing methods that define hash functions in the kernel space are:
  - Kernel-based unsupervised locality sensitive hashing.
  - Kernel-based supervised hashing LS locality sensitive hashing.
- ✓ Kernel-based methods express the Gaussian random vector as the weighted sum of *m* images selected from the archive as:

$$v_r = \sum_{j=1}^{m} \omega_r(j)\phi(\mathbf{X}_j)$$

nonlinear mapping function

Then the hash function becomes:

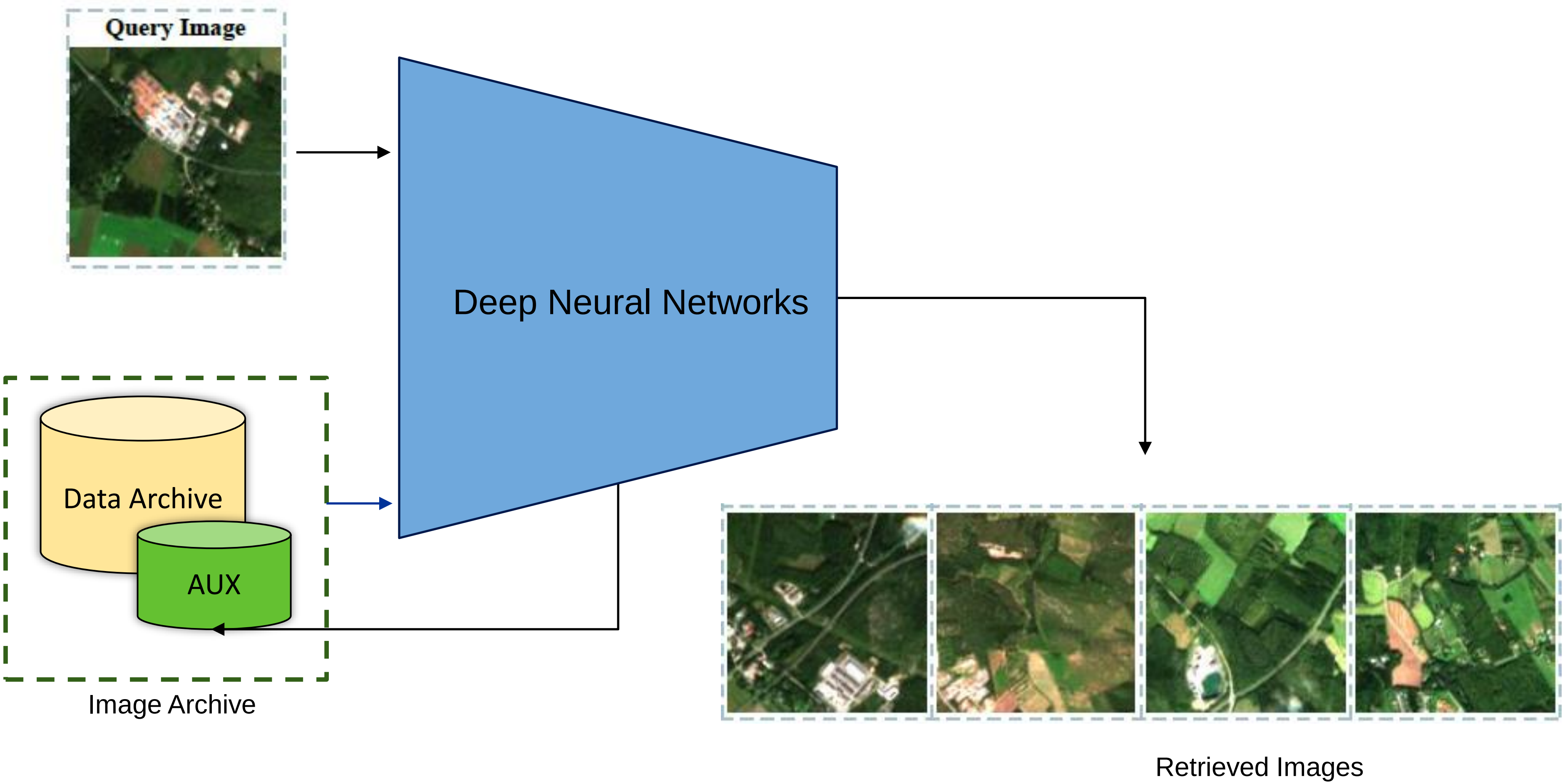
$$h_r(\mathbf{X}_i) = sign\left(\sum_{j=1}^{m} \omega_r(j)\phi(\mathbf{X}_j)\phi(\mathbf{X}_i)\right) = sign\left(\sum_{j=1}^{m} \omega_r(j)K(\mathbf{X}_j, \mathbf{X}_i)\right),\ r = 1, 2, \ldots, b$$

r-th hash function

kernel function

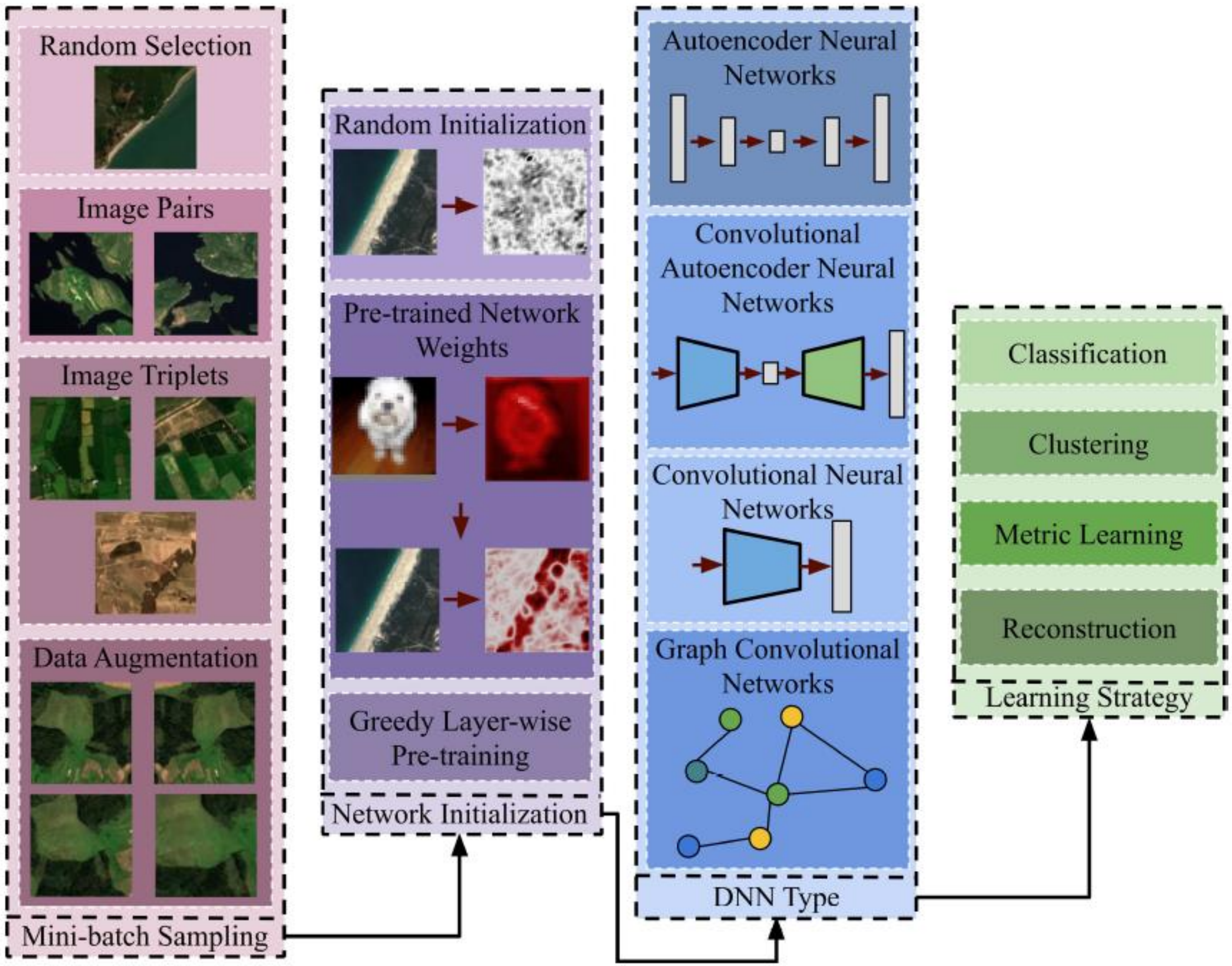
Demir et al, "Hashing Based Scalable Remote Sensing Image Search and Retrieval in Large Archives", IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no.2, pp. 892-904, 2016.

Query by Example: Advanced Systems

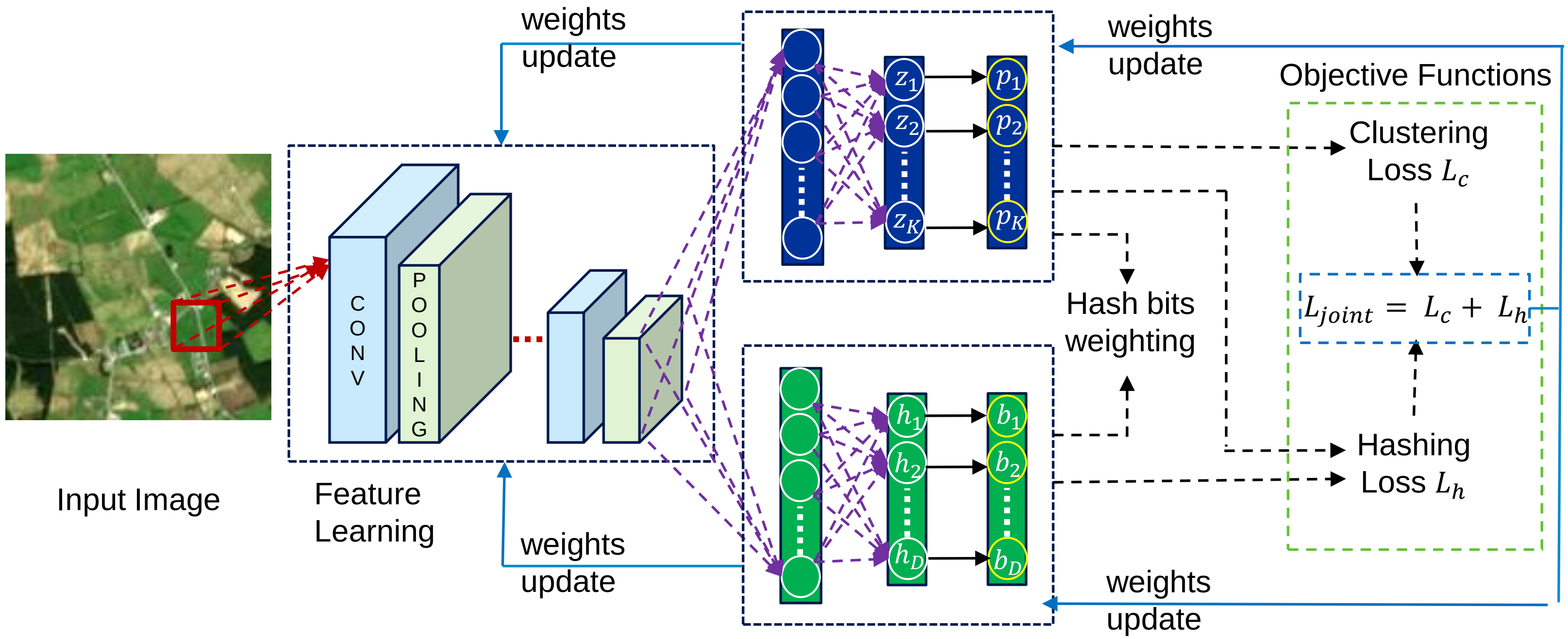


✓ Deep learning based systems jointly optimize feature learning and image retrieval.

Query by Example: Advanced Systems



Scalable Search and Retrieval



Deep class-wise hashing

Adversarial hashing

Semantic-preserving hashing

Graph-based hashing

Multi-modal hashing

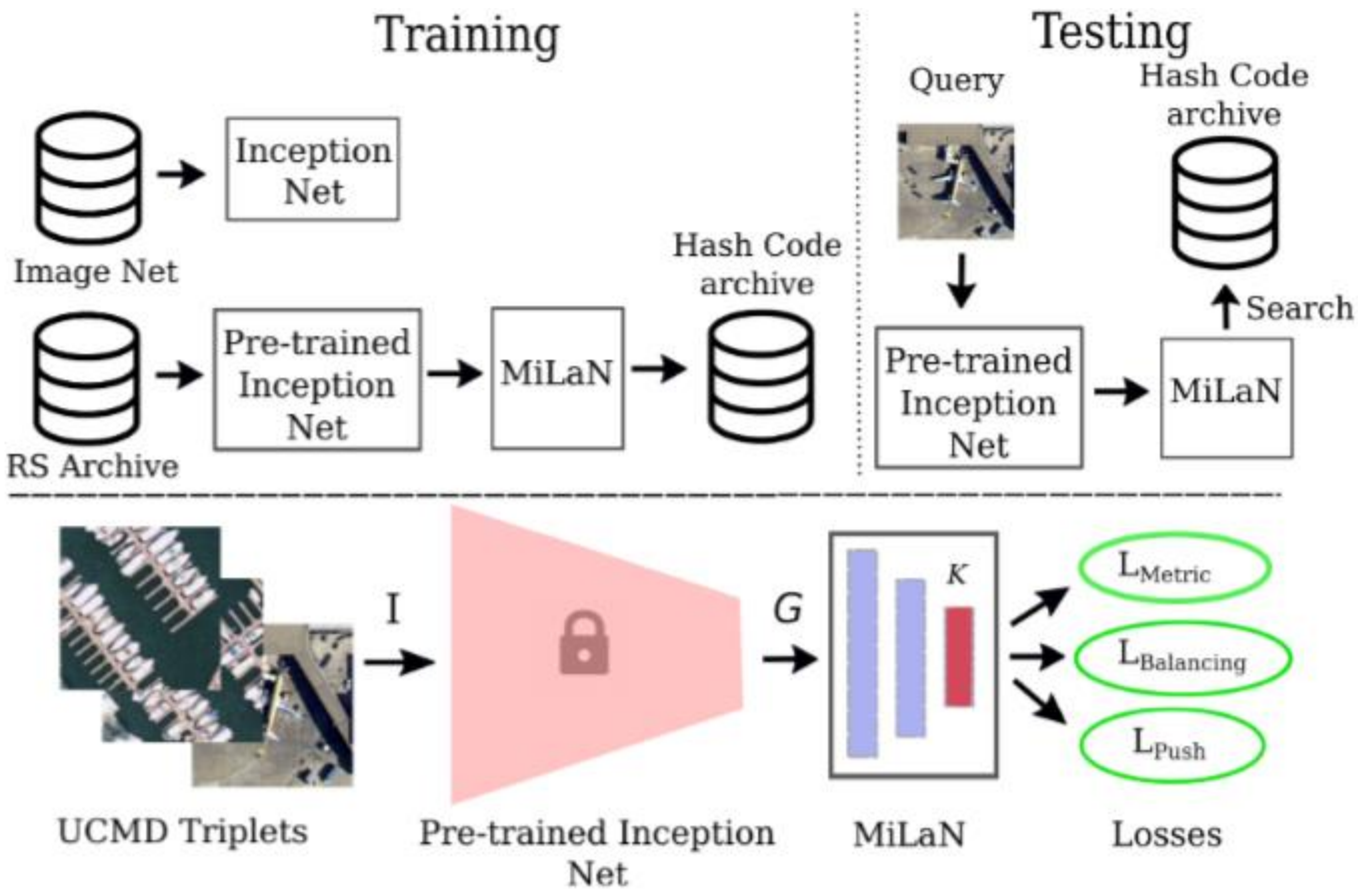
Attention guided hashing

Zero-shot hashing

Unsupervised hashing

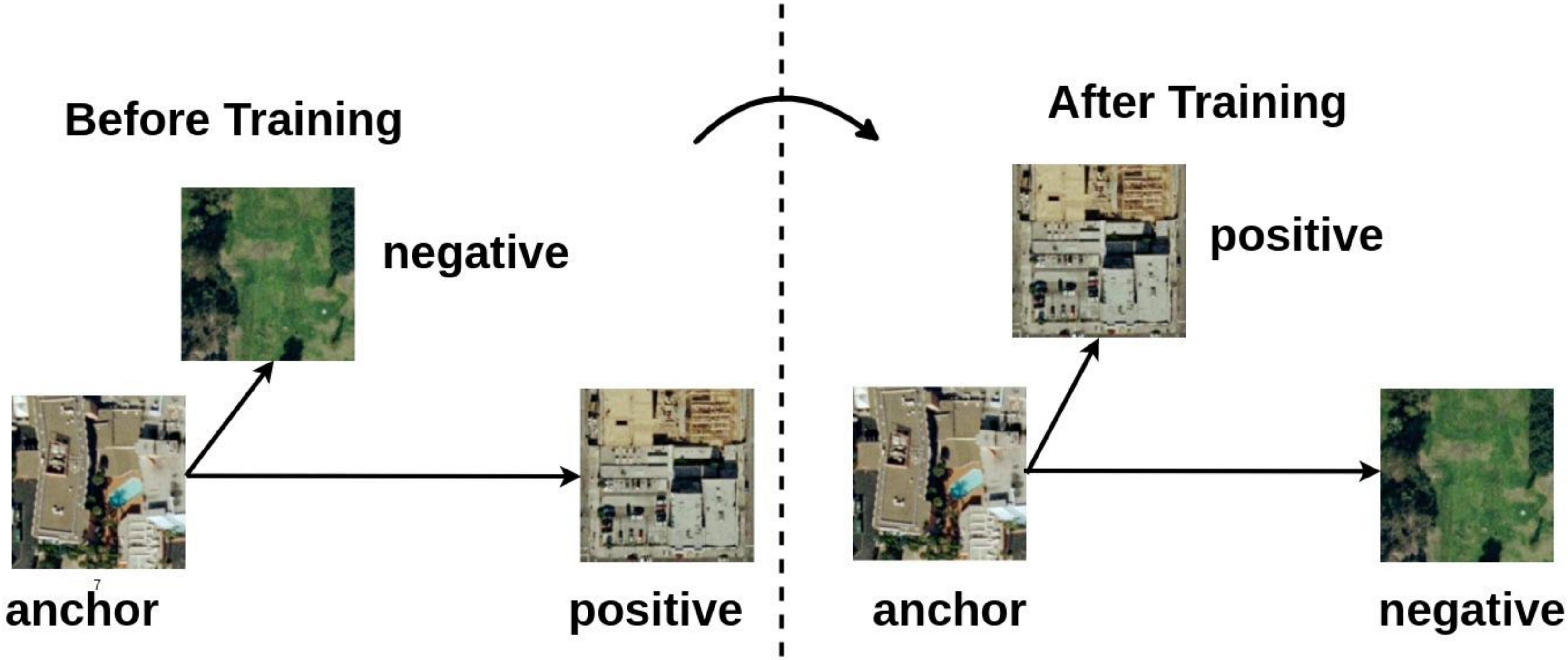
Weakly-supervised hashing

MiLAN: Metric Learning based Deep Hashing



Roy et al, "Metric-Learning based Deep Hashing Network for Content Based Retrieval of Remote Sensing Images", IEEE Geoscience and Remote Sensing Letters, 2020.

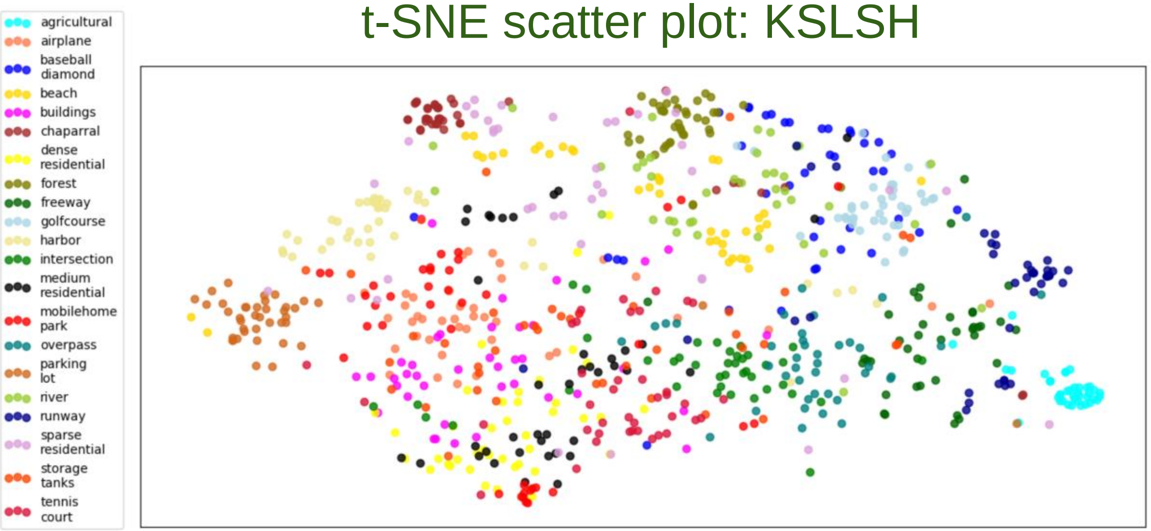
MiLAN: Metric Learning based Deep Hashing



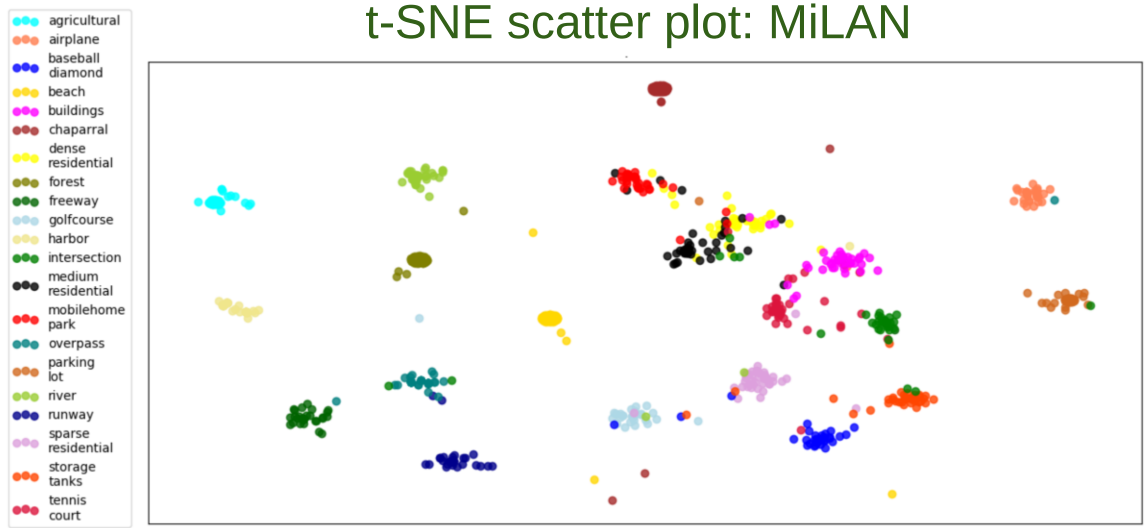
The intuition behind the **triplet loss**: after training, a positive sample is “moved” closer to the anchor sample than the negative samples of the other classes.

MiLAN: Results

t-SNE scatter plot: KSLSH



t-SNE scatter plot: MiLAN



T-SNE: t-distributed stochastic neighbor embedding

MiLAN: Results

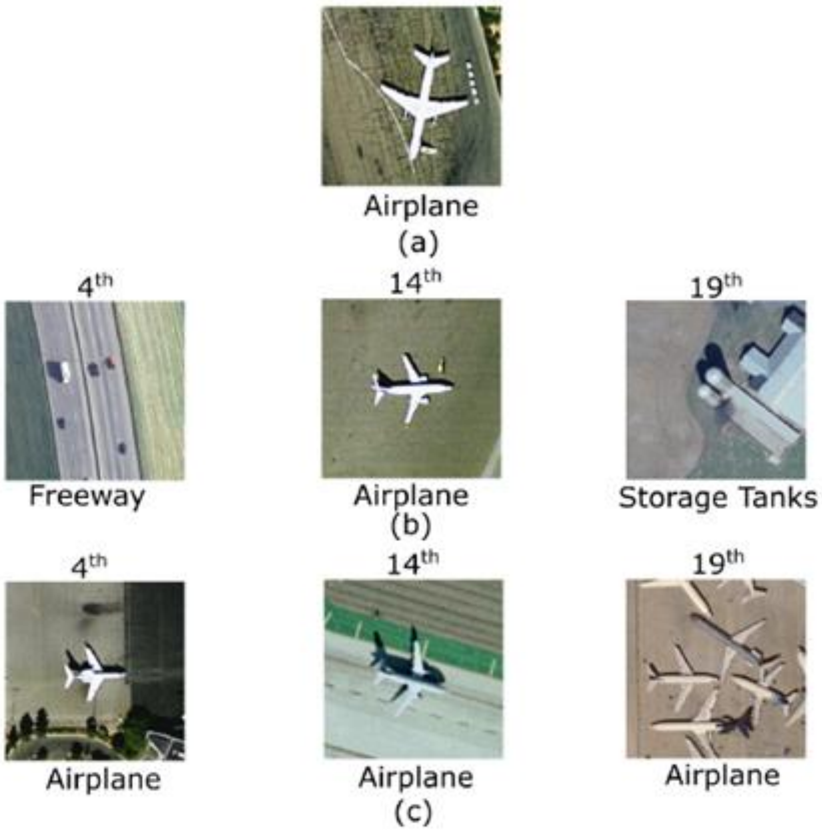


Figure: (a) The query image, (b) Images retrieved by KSLSH. and (c) Images retrieved by the MiLaN.

Cross-Domain Hashing

Research in Progress:

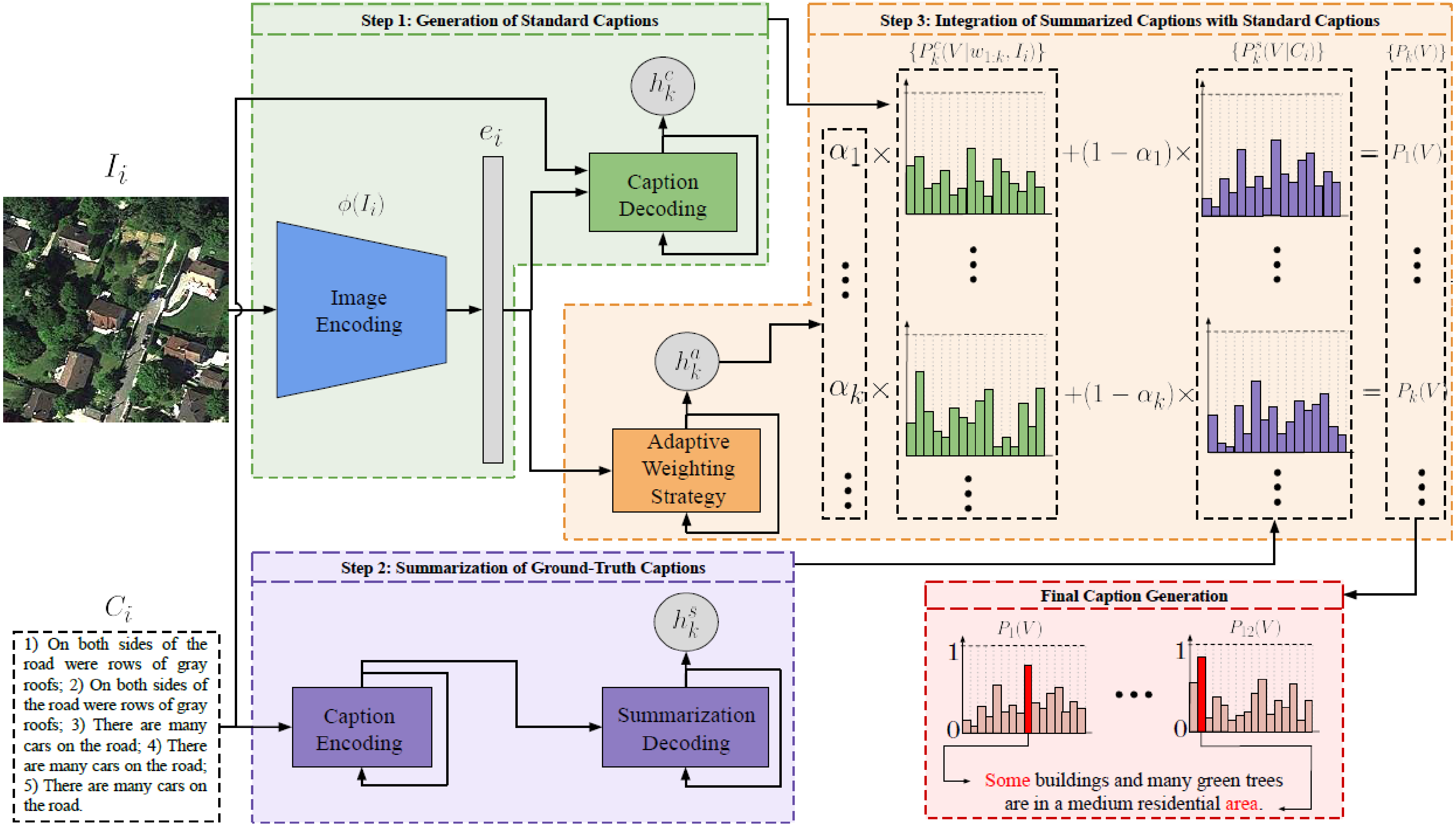
✓ Extension of the hashing methods to be operated as cross-modal retrieval problems.

✓ Cross-modal retrieval aims to search semantically similar images in one modality (e.g., Sentinel-2) by using a query from another modality (e.g., Sentinel-1).

✓ We develop hashing methods that:

- effectively learn the common representations for the heterogeneous EO data by preserving the semantic discrimination and modality invariance simultaneously in an end-to-end manner.
- consist of intermodality similarity-preserving learning and semantic label-preserving learning modules based on different types of loss functions simultaneously.
- include an inter-modal invariance triplet loss and inter-modal pairwise loss functions in the framework of the cross-modal retrieval problems.

RS Image Captioning with LSTMs



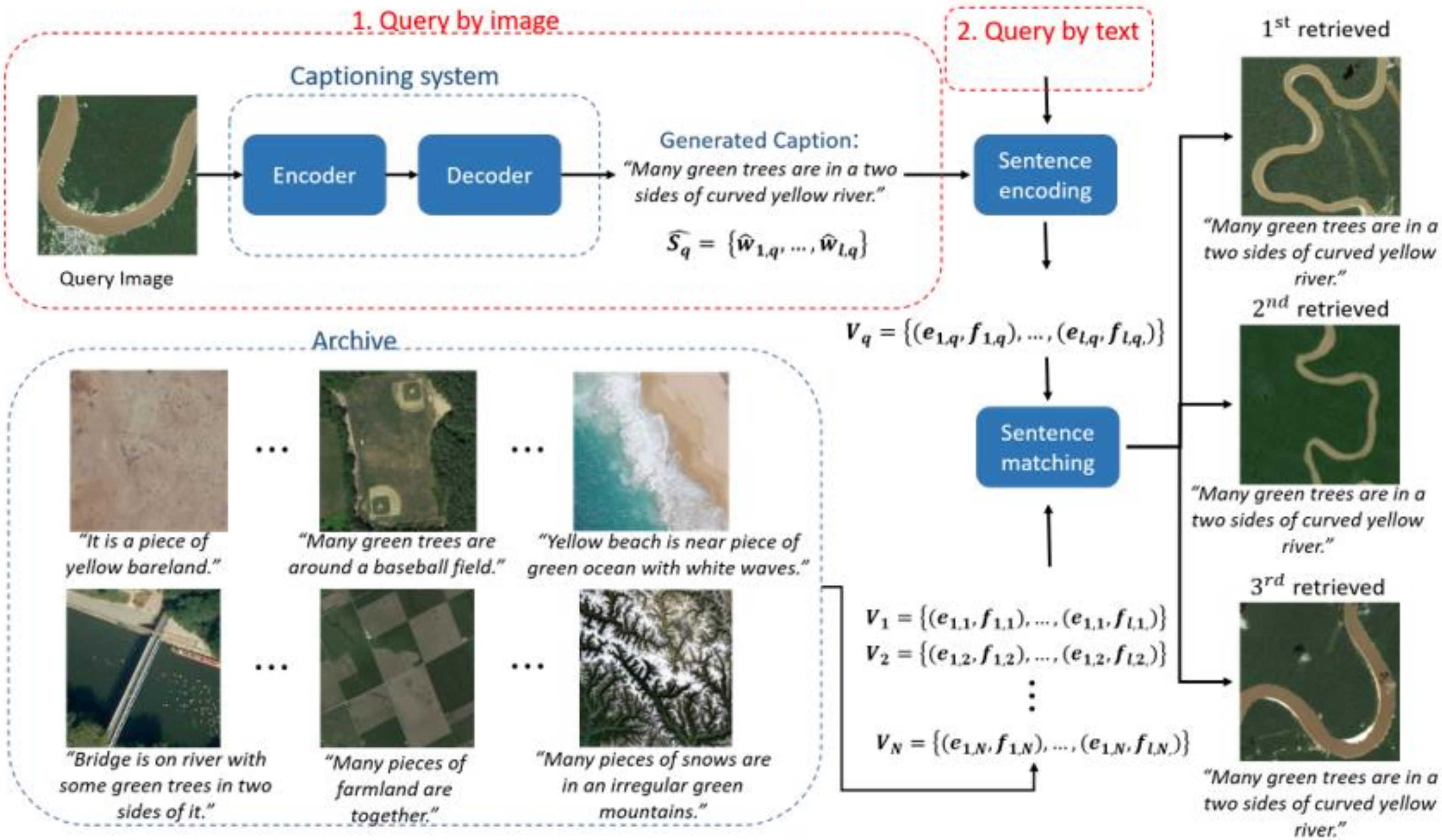
Sumbul et al, "SD-RSIC: Summarization Driven Deep Remote Sensing Image Captioning", IEEE TGRS, 2020.

RS Image Captioning with LSTMs

Images

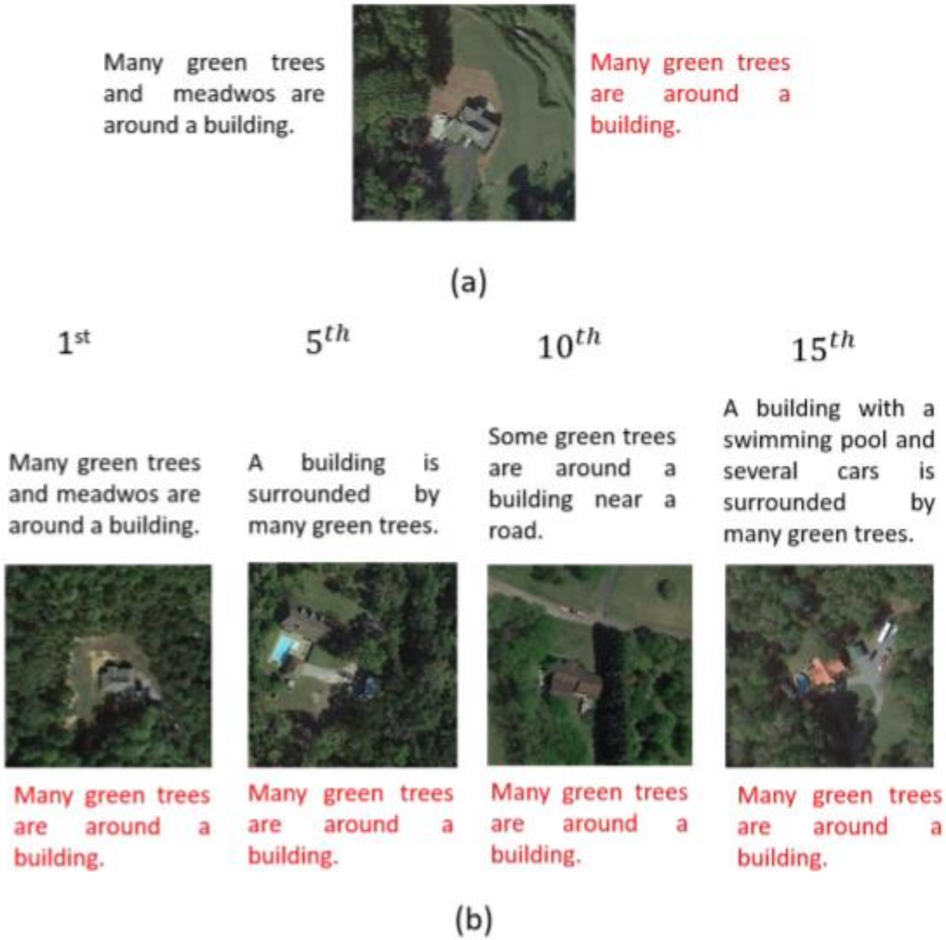
SD-RSIC	A bridge is on a river with some green trees in two sides.	Many cars are on a bridge over a river with many green trees in two sides of it.	Some buildings and many green trees in a medium residential area.	It is a piece of green meadow.	Many storage tanks are in a factory near a river.
NIC	A bridge is over a river in a bridge over it.	Many cars are on a bridge over a parking lots.	Many green trees and a swimming pool are in a resort.	It is a large piece of green mountain.	Many green trees and green and parking.
Ground-truth Captions	1. On either side of the river there are many grey roofed houses. 2. On either side of the river there are many grey roofed houses. 3. On either side of the river there are many grey roofed houses. 4. There is a magnificent bridge over the river. 5. There is a magnificent bridge over the river.	1. There are many cars running on the road. 2. There are many cars running on the road. 3. There are many cars running on the road. 4. There are many tall trees planted on both sides of the river. 5. There are many tall trees planted on both sides of the river.	1. The residential with black villages is in the center of the forest. 2. The residential with black villages is in the center of the forest. 3. The residential with black villages is in the center of the forest. 4. This lush woods is surrounding the peaceful neighborhood with roads passes by. 5. Several buildings and many green trees are in a residential area.	1. A furcate road separates the grass green farmland. 2. A furcate road separates the grass green farmland. 3. The green farmland is divided by a furcate road. 4. It is a green farmland with several curved roads through it . 5. Many pieces of green farmlands are together.	1. There is a factory beside the river. 2. There is a factory beside the river. 3. There is a factory beside the river. 4. There are many storage tanks in the factory. 5. There are many storage tanks in the factory.

Remote Sensing Image Retrieval under a Deep Captioning Perspective



Hoxha et al, "Toward Remote Sensing Image Retrieval under a Deep Image Captioning Perspective", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020.

Remote Sensing Image Retrieval under a Deep Captioning Perspective



Sparse residential image retrieval.

a. Query image.
b. Images retrieved using the proposed retrieval system.

Generated textual descriptions (highlighted by red) are reported below the related images of (b). One ground truth description is reported above the related images of (b).

The order of the retrieved images is reported above each image.

Hoxha et al, "Toward Remote Sensing Image Retrieval under a Deep Image Captioning Perspective", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2020.

Compressed Domain Analysis

- ✓ Deep Learning approaches have the capability to learn high-level semantic content of remote sensing (RS) images resulting in high classification accuracy.

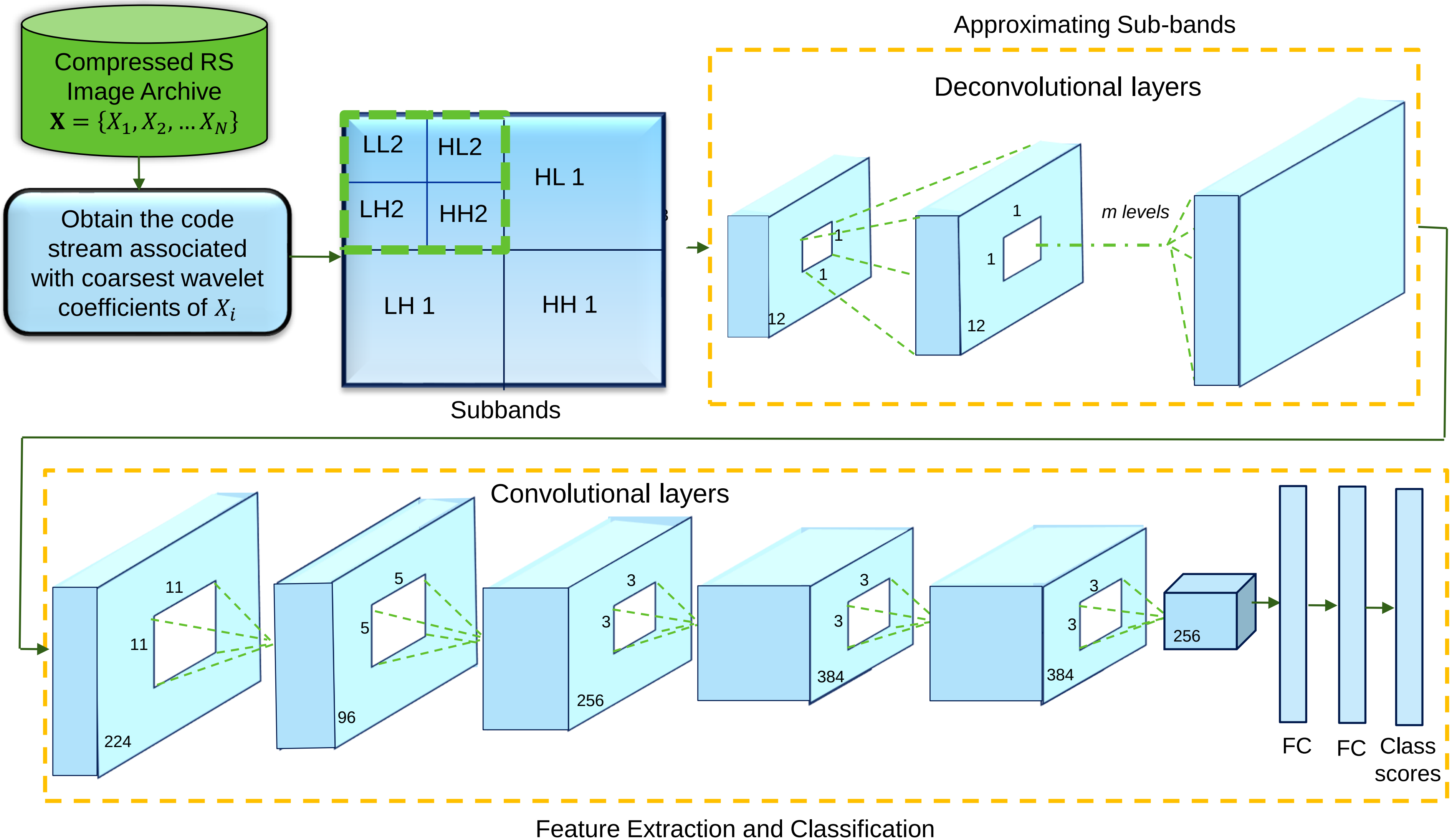
- ✓ RS images are usually stored in compressed format in large-scale archives.

- ✓ Conventional deep neural networks (DNNs) do not directly operate on the compressed streams of the RS images.

Problem: RS image scene analysis using DNNs requires decoding which is computationally-demanding and time-consuming.

Byju et al, "Remote Sensing Image Scene Classification with Deep Neural Networks in JPEG 2000 Compressed Domain", IEEE Transactions on Geoscience and Remote Sensing, 2020.

Compressed Domain Image Analysis



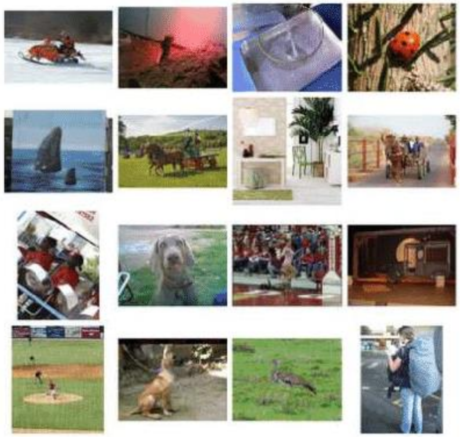
Byju et al, "Remote Sensing Image Scene Classification with Deep Neural Networks in JPEG 2000 Compressed Domain", IEEE Transactions on Geoscience and Remote Sensing, 2020.

Requirements on Annotated Images

- ✓ Most DL models require huge amounts of annotated images during training to optimize all parameters and reach a high-performance during evaluation.

- ✓ The availability and quality of such data determine the feasibility of many DL models.

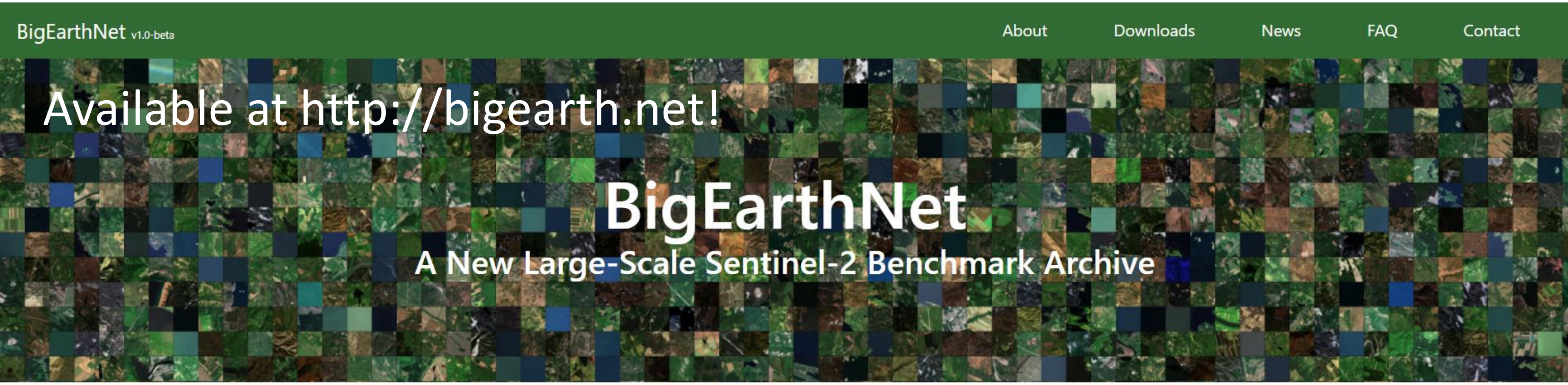
- ✓ One solution is to use of deep learning models pre-trained on large scale computer vision archives (e.g., ImageNet)



Problem: Differences on the characteristics of images between computer vision and remote sensing.

Requirements on huge amounts of annotated images

✓ Recent datasets have been constructed based on the publicly available thematic products (e.g., Corine Land-Cover database).



Sumbul et al, “BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding", IEEE International Conference on Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019.

BigEarthNet: Results

Methods	Recall	F_1 Score	F_2 Score
Pre-trained Inception-v2 on ImageNet*	43.5 %	0.45	0.44
RGB Bands in Standard CNN	56.8 %	0.57	0.56
All Bands in Standard CNN	62.0 %	0.61	0.62
All Bands in K-Branch CNN	**71.5 %**	**0.67**	**0.70**

* We apply fine-tuning to the pre-trained Inception-v2 architecture.

Sumbul et al, “BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding", IEEE International Conference on Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019.

Corine Limitations

✓ Some classes can be missing and noisy:

Image Scene		
Multi-Labels	Broad-leaved forest Discontinuous urban fabric Complex cultivation patterns	Land principally occupied by agriculture, with significant areas of natural vegetation Non-irrigated arable land Coniferous forest
Predictions	Discontinuous urban fabric Green urban areas Complex cultivation patterns Broad-leaved forest	Discontinuous urban fabric Non-irrigated arable land Complex cultivation patterns Land principally occupied by agriculture, with significant areas of natural vegetation

✓ According to the validation report of the CLC, the accuracy is around 85%.

Image Search and Retrieval with Noisy Labels

- ✓ Collecting large-scale data with complete-clean labels for supervised training of neural networks is practically challenging for EO.

- ✓ We research on robust training at label noise regimes by developing:
  - noise-tolerant loss functions;
  - early stopping approaches;
  - importance re-weighting approaches;
  - decoupling, co-teaching, student-teacher based DL models.

Li et al. “Learning to learn from noisy labeled data”, IEEE Conference on Computer Vision and Pattern Recognition. 2019, pp. 5051–5059.

Ghosh et al. “Robust loss functions under label noise for deep neural networks”, ArXiv preprint arXiv:1712.09482, 2017.

Nguyen et al. “Robust Learning Under Label Noise with Iterative Noise-Filtering”, ArXiv preprint arXiv:1906.00216, 2019.

Yi et al. “Probabilistic end-to-end noise correction for learning with noisy labels”, IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7017–7025.

Concluding Remarks

✓ Hashing-based methods are promising for RS retrieval problems due to their capability on fast and scalable image search and retrieval. We currently work on:

- Multi-modal/cross-modal hashing;
- Uncertainty sensitive hashing;
- Volunteered geographic information driven hashing.
- Physics-guided image retrieval and hashing.
- DNN based compressed domain image retrieval and hashing.
- Image retrieval operated on data-cubes.
- Data-privacy preserving image retrieval and hashing.

Accurate and Fast Discovery of Crucial Information for Observing Earth from Big EO Archives

http://bigearth.eu/

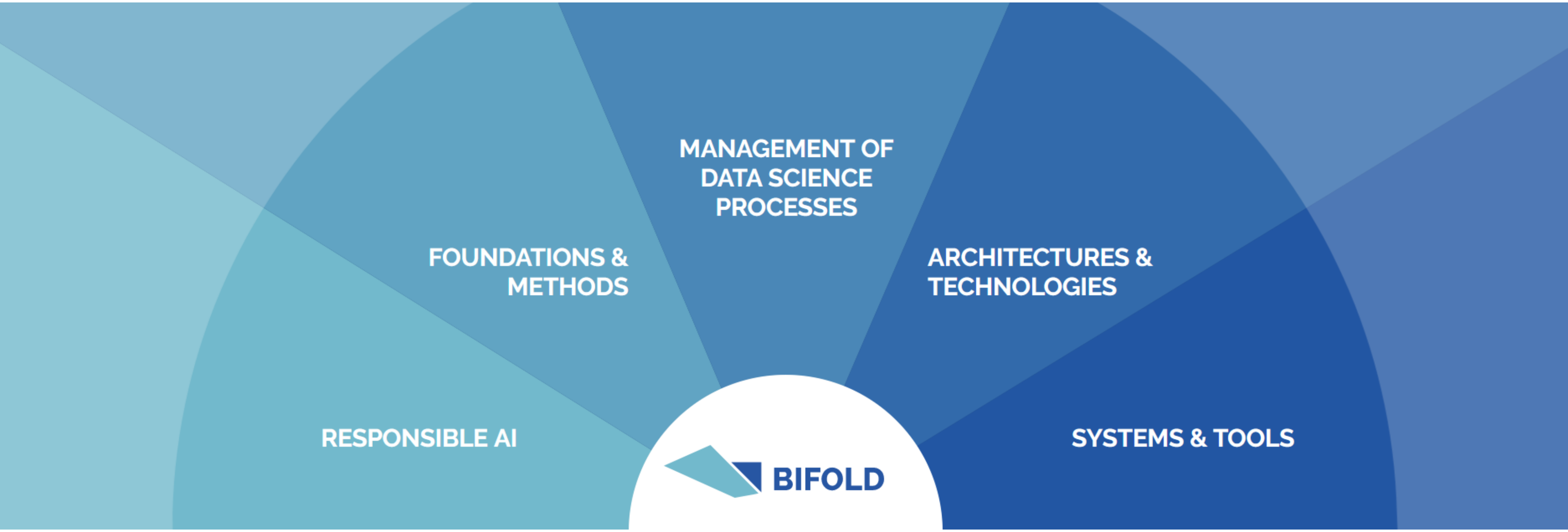
The codes developed and maintained at our group are publicly available:

https://www.rsim.tu-berlin.de/menue/software

Follow us in twitter: @BigEarthERC

BIFOLD: Berlin Institute for the Foundations of Learning and Data

https://bifold.berlin



✓ BIFOLD conduct research into the scientific foundations of Big Data and Machine Learning, to advance AI application development, and greatly increase the impact to society, the economy, and science.