

Hardness-aware adaptive triplet learning approach for content-based image retrieval from remote sensing archives

Aswathy K R , Akshara Preethy Byju & Jorge Alberto Suárez Garea

To cite this article: Aswathy K R , Akshara Preethy Byju & Jorge Alberto Suárez Garea (2026) Hardness-aware adaptive triplet learning approach for content-based image retrieval from remote sensing archives, Geocarto International, 41:1, 2656085, DOI: [10.1080/10106049.2026.2656085](https://doi.org/10.1080/10106049.2026.2656085)

To link to this article: <https://doi.org/10.1080/10106049.2026.2656085>



© 2026 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 09 Apr 2026.



Submit your article to this journal [↗](#)



Article views: 90



View related articles [↗](#)



View Crossmark data [↗](#)

REPORT



Hardness-aware adaptive triplet learning approach for content-based image retrieval from remote sensing archives

Aswathy K R^a, Akshara Preethy Byju^a and Jorge Alberto Suárez Garea^b

^aDepartment of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Kollam, Kerala, India; ^bCentro Singular de Investigación en Tecnologías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain

ABSTRACT

Deep-metric learning using triplet loss has shown promise in remote sensing (RS) image retrieval by mapping similar images closer and dissimilar ones farther apart in embedding space. However, fixed-margin triplet loss struggles with intra-class and inter-class variability and often suffers from zero loss, where trivial triplets dominate training, causing ineffective convergence. To address this, we propose a novel hardness-aware adaptive margin triplet loss (HA²TL-CBIR) that dynamically adjusts the margin based on sample difficulty. By measuring the distance between positive and negative pairs, our method prioritizes hard triplets and reduces the impact of trivial ones. This adaptive margin strategy overcomes zero-loss issues and enhances the learning of robust, discriminative features. Experiments on benchmark RS datasets PatternNet and NWPU-RESISC45 demonstrate the effectiveness of our approach in handling the growing complexity and size of RS image archives, improving retrieval performance significantly.

ARTICLE HISTORY

Received 20 September 2025
Accepted 1 April 2026

KEYWORDS

Content-based image retrieval; remote sensing; deep metric learning; triplet training; adaptive margin

1. Introduction

Deep metric learning (DML) have recently emerged as a powerful technique to accurately characterize remote sensing (RS) images for several applications of content-based image retrieval (CBIR) (Cheng et al. 2018). DML learns an embedding space where similar images are brought closer and dissimilar images are pushed farther apart during training. The distance metric considers a novel representation of the images in the embedding space where they appear to be distinguishable. One of the most widely used methods in DML is triplet training, which involves the triplet loss function that uses triplets formed by (i) an anchor image; (ii) a positive image (belongs to the same class as that of the anchor image); and (iii) a negative image (belongs to a different class) (Wang et al. 2014; Schroff et al. 2015; Song et al. 2016). The triplet loss aims to minimize the distance between the anchor and the positive image while maximizing the distance between the anchor and the negative image (Feng et al. 2020).

This ensures that the learned embedding space separates similar and dissimilar images effectively. A predefined margin is typically introduced in traditional triplet loss to maintain a minimum separation between positive and negative pairs (Cao et al. 2020). In this fixed margin approach, the margin remains constant for all triplets during training. However, this approach ignores images with high intra-class variability and low inter-class dissimilarity, thereby treating all triplets with equal importance, regardless of their difficulty. As a result, the learning process is often dominated by *trivial triplets*, where a negative image is already well separated from the anchor image, thereby contributing a smaller extent of information required for training. On the other hand, in *hard triplets*, the negative image is closer to the anchor compared to the positive image, thereby providing substantial information required for training (Kertész 2022). However, in fixed-margin approaches, both the *trivial* and *hard triplets* are handled in the same way, which limits the discriminative power of the learned features.

In contrast, our proposed HA²TL-CBIR approach adopts a novel hardness-aware adaptive margin learning strategy. Instead of using a constant margin, the model dynamically adjusts the margin based on

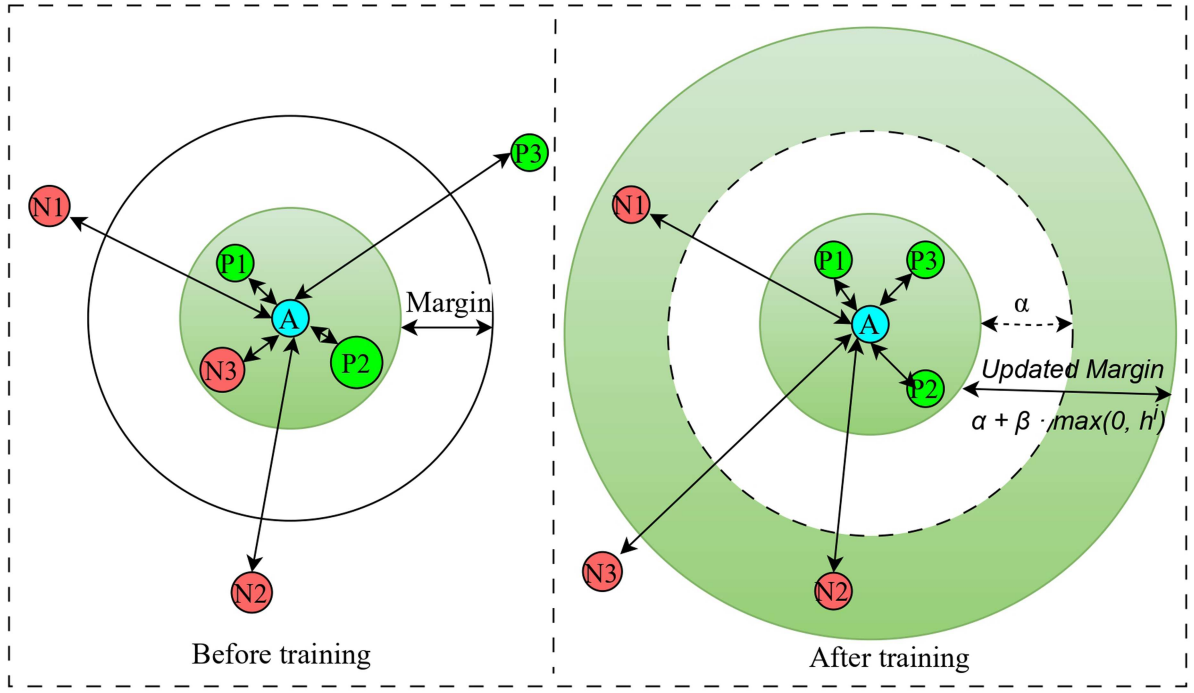


Figure 1. HA²TL-CBIR: adaptive triplet learning approach for content-based image retrieval.

the hardness (or difficulty) of each triplet (see Figure 1). The *hardness* of the triplet reflects how challenging it is for the model to correctly separate the positive and negative image samples. The *hardness* is computed as the difference between the anchor-positive and anchor-negative distances. If a triplet is *hard*, a larger margin is applied to enforce stronger separation and drive learning. For *trivial triplets*, a smaller margin is used to reduce their influence and avoid unnecessary updates. This adaptive mechanism focuses training on informative triplets and reduces the dominance of trivial samples that produce zero loss. Figure 1 shows an illustration of triplet training using adaptive margin triplet loss highlighting the margin change for *hard* triplets. (A, P_1, N_1) and (A, P_2, N_2) are *trivial* triplets whereas (A, P_3, N_3) is a *hard* triplet. After trained using our adaptive margin triplet training approach, negative image N_3 will be pushed away from the anchor A and positive image P_3 will be pulled closer to anchor image A . As a result, HA²TL-CBIR improves convergence, increases retrieval accuracy, and produces more robust and discriminative embeddings for CBIR tasks in complex RS image archives.

In summary, our contributions are *three-fold*:

- (1) We identified the need to consider the dynamic learning mechanism that considers adaptive margin;
- (2) We proposed a novel approach to handle the varying margin distances between positive and negative triplets;
- (3) With extensive experiments, we compare the performance of the proposed approach with recent DML-based methods.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 presents the proposed adaptive margin approach. Section 4 describes the experimental design. Section 5 discusses the results and analysis. Finally, Section 6 presents the conclusions.

2. Related work

CBIR system has evolved significantly with the advent of deep learning. Conventional image retrieval systems mainly include two modules: (i) feature characterization; and (ii) similarity assessment. Feature characterization describes an input image through mathematical representation and later performs similarity assessment between

these features to retrieve similar images from the archive. The main challenge is to accurately characterize images using these feature descriptions. Thus, early works mainly relied on hand-crafted features, such as colour histograms, texture descriptors, and shape features, which often struggled to generalize across diverse landscapes and imaging conditions (Sudha and Aji 2022). The advent of convolutional neural networks (CNNs) revolutionized RS image retrieval tasks by enabling automatic extraction of high-level semantically rich features (Cheng et al. 2018; Cao et al. 2020). Architectures such as ResNet (He and Zhang, 2016), DenseNet (Huang et al. 2017), and EfficientNet (Tan and Le 2019) have been widely employed to capture discriminative representations from RS imagery. Transfer learning from large-scale natural image datasets (e.g. ImageNet) has further improved performance, particularly in scenarios where annotated RS datasets are scarce (Cheng et al. 2017; Zhou et al. 2018). This generalization capability of deep features across domains has also been demonstrated in prior studies, where features learned from everyday object datasets were shown to transfer effectively to remote sensing imagery (Penatti et al. 2015).

To enhance the performance of image retrieval tasks, DML has emerged as a powerful paradigm in RS. Unlike conventional classification models, metric learning focuses on structuring the embedding space so that semantically similar images are placed closer together. Early methods such as contrastive loss and triplet loss introduced the concept of learning embeddings by enforcing distance constraints between pairs or triplets of samples (Wang et al. 2014; Schroff et al. 2015; Song et al. 2016; Hermans et al. 2017; Ha and Blanz 2021; Yuan et al. 2021). Within the RS domain, triplet-based deep metric learning has demonstrated substantial improvements in retrieval accuracy by explicitly optimizing for discriminative and compact feature embeddings (Cao et al. 2020; Zhang et al. 2021; Regan et al. 2023). The triplet graph CNN (CNN-TL) approach (Regan et al. 2023) extends this concept by integrating triplet loss with graph convolution to model contextual relationships between image regions, thereby improving the semantic consistency of retrieved results. Owing to its strong reported performance, CNN-TL is also considered as a baseline in our experimental comparisons. Enhancements in original triplet loss, including hard negative mining (Wu et al. 2017; Kert'esz 2022), adaptive margin strategies (Ha and Blanz 2021), and label distribution smoothing (Yuan et al. 2021), have further improved the robustness of the learned embeddings.

Recent advances in deep metric learning have emphasized the importance of designing loss functions and sampling strategies that better reflect semantic similarity in RS imagery. Standard triplet loss approaches depend on selecting effective anchor-positive-negative triplets, which can be challenging due to high intra-class variability and inter-class similarity (Wang et al. 2014; Schroff et al. 2015). Hard negative mining has been widely adopted to address this, by focusing training on negatives that are most similar to the anchor, thereby increasing the discriminative power of the learned embeddings (Kert'esz 2022; Wu et al. 2017). Adaptive margin strategies further improve this process by dynamically adjusting the margin based on the difficulty of each triplet (Ha and Blanz 2021).

Although adaptive margin strategies such as Adaptive Margin Triplet Loss (AMTL) (Ha and Blanz 2021) dynamically adjust the margin based on triplet difficulty, several limitations remain. Many existing approaches adjust the margin only at a coarse level and do not sufficiently amplify the distinction between moderately hard and extremely hard triplets. As a result, highly informative triplets may not receive adequate emphasis during training. In addition, these methods typically do not explicitly modulate the gradient contribution of samples according to their hardness, allowing trivial triplets to still influence the optimization process. These limitations motivate the need for a hardness-aware adaptive margin strategy that more effectively prioritizes difficult samples and improves discriminative feature learning.

Graph-based extensions of deep metric learning, such as triplet graph CNNs, have shown promise in RS CBIR (Regan et al. 2023). By incorporating relationships among multiple regions of an image, graph-based methods can capture contextual information that standard triplet loss approaches might overlook. Recent studies have also explored size-aware graph embedding approaches that incorporate object-level relationships and relative size information to improve semantic understanding in remote sensing imagery, highlighting the importance of modelling structural attributes in scene representation (Ni et al. 2026). Proxy-based metric learning addresses scalability concerns in large RS datasets by representing each class with a learnable proxy vector instead of computing distances between all possible sample pairs (Sohn 2016; Movshovitz-Attias et al. 2017; Kim et al. 2020). Hybrid loss functions that combine triplet loss with proxy-based objectives or multi-similarity loss have demonstrated additional improvements in retrieval accuracy (Harwood et al. 2017; Wang et al. 2019; Teh et al. 2020).

Another important aspect is the extraction of multi-scale features. RS images often contain objects of varying sizes, ranging from small vehicles to large buildings. Capturing both local and global structures in embeddings is critical for accurately representing these scenes, enhancing retrieval precision, especially for heterogeneous landscapes (Ha and Blanz 2021; Regan et al. 2023). In addition to methodological improvements, practical considerations have been explored to optimize deep metric learning in RS CBIR. Efficient triplet or pair sampling strategies help focus training on the most informative examples, addressing challenges posed by high intra-class variability and visually similar classes (Schroff et al. 2015; Hermans et al. 2017). Incorporating spatial and contextual information enables embeddings to reflect both local textures and global scene layout, improving consistency across complex RS images (Ha and Blanz 2021; Regan et al. 2023). Multi-scale feature modelling is particularly beneficial in remote sensing, where objects range from small roads and vehicles to large industrial or forested regions. Another recent direction within metric learning focuses on proxy-based formulations for improving representation robustness. Cai et al. (2024) introduced a proxy-based rotation-invariant DML framework (PBFFN), which employs proxy Euclidean contrast loss and multi-layer feature fusion to enhance global feature discrimination under arbitrary rotations. Their method achieves highly competitive retrieval performance on the PatternNet dataset, demonstrating the effectiveness of proxy-driven embedding optimization. However, proxy-based approaches introduce additional overhead for managing class-level proxies and require careful training stabilization, whereas our hardness-aware adaptive margin formulation directly optimizes sample relationships without relying on external proxy parameters.

In addition to these methodological advances, recent works have emphasized the use of contrastive learning approaches for RS CBIR. SimCLR-style contrastive learning has also been adapted to remote sensing, with Ayush et al. (2021) introducing a geography-aware self-supervised framework that improves representation robustness across diverse geographic regions. Although SimCLR does not report retrieval metrics on RSIR benchmarks, it represents a mainstream contrastive learning direction relevant to CBIR and has inspired RSIR-specific extensions such as DMBSS (Liu et al. 2024). These techniques alleviate the dependence on large annotated datasets by pretraining models with instance discrimination or augmentation based consistency objectives, which are later fine tuned for retrieval tasks (Chowdhury et al. 2021; Jiang et al. 2023; Yang et al. 2023). Such pretraining strategies improve generalization across diverse geographic regions and acquisition conditions. Another direction of research focuses on multi-label retrieval, where each image may contain several semantic categories. Embedding models that incorporate multi-label information better capture the inherent complexity of RS scenes and improve retrieval precision when compared to single-label representations. Furthermore, scalability concerns in large scale archives have motivated the development of efficiency-focused techniques such as hashing based retrieval (Liang et al. 2021; Guo et al. 2022). In addition to traditional hashing frameworks, Wang et al. (2025) recently introduced an Automatic Weight Allocation based Contrastive Hashing approach (AWCLoss), which integrates a Gaussian weighted contrastive loss with a dynamic hard-negative suppression mechanism. Their method further incorporates a hashing layer to produce compact binary codes, enabling efficient large-scale retrieval. AWCLoss demonstrates strong mAP, ANMRR, and Precision@K performance on UCM, AID, and NWPU-RESISC45 datasets, making it one of the most recent state-of-the-art contrastive learning baselines for RS CBIR. Despite these advances, challenges remain in handling high intra-class variability, visually similar inter-class samples, and computational efficiency. These limitations highlight the relevance of our proposed hardness-aware adaptive margin strategy, which dynamically prioritizes informative triplets to enhance discriminative learning.

Scalability and efficiency are critical for operational RS CBIR. Proxy-based and hybrid loss strategies reduce the computational cost of large datasets (Sohn 2016; Harwood et al. 2017), while lightweight networks and knowledge distillation techniques maintain retrieval performance with smaller models (Tan and Le 2019; Feng et al. 2020). Evaluation metrics have also evolved to include mean Average Precision (mAP), Normalized Modified Retrieval Rank (NMRR), and top-k retrieval accuracy, providing more comprehensive assessments of embedding quality (Cao et al. 2020; Zhang et al. 2021). Domain adaptation and transfer learning address dataset diversity and sensor variation (Cheng et al. 2018; Cao et al. 2020; Zhang et al. 2021). Fine-tuning pre-trained models remains effective for improving generalization across different geographical areas and acquisition conditions (Cheng et al. 2017; Zhou et al. 2018). Several recent surveys have reviewed these developments, identifying key challenges and emerging opportunities (Chowdhury et al. 2021; Jiang et al. 2023; Yang et al.

2023). These include improving interpretability of retrieval models, robustness to domain shifts, and the creation of large-scale, diverse benchmark datasets. There is also growing interest in self-supervised learning and unsupervised domain adaptation to reduce reliance on labelled training data (Chowdhury et al. 2021; Jiang et al. 2023; Yang et al. 2023).

Overall, deep metric learning in remote sensing CBIR has matured from simple triplet-based methods to sophisticated frameworks combining adaptive margins, hard negative mining, graph representations, proxy-based objectives, and multi-scale feature modelling. These developments have contributed to more discriminative embeddings and higher retrieval accuracy, particularly in large and complex RS datasets.

3. Methods

3.1. Problem formulation

Let $\mathbf{X} = \{X_i\}_{i=1}^M$ be the set of M RS images in the archive where X_i represents i -th image. Given a query image X_q , the proposed approach aims to retrieve the top k similar images from the archive. Consider a triplet image sample $(A, P, N)_i$ where A_i, P_i, N_i correspond to the anchor, positive and negative image of the i^{th} sample, respectively. Figure 2 illustrates the HA²TL-CBIR approach, which mainly consists of four modules mainly: i) triplet selection; ii) convolutional neural network (CNN) based backbone network to obtain the feature embeddings; iii) hardness-aware adaptive triplet learning approach which includes computing hardness coefficient and adaptive margin (see section II. B and II. C); and iv) image retrieval based on similarity matching. The proposed method randomly selects a set of triplets from a mini batch $\mathbf{B} \subset \mathbf{X}_T$, where $\mathbf{X}_T \subset \mathbf{X}$ is a set of T training images. From each triplet sets, an anchor is chosen using random selection approach which ensures representative selection with minimal bias. This is followed by a *hardness* calculation of the selected triplets. When *hard triplets* are assigned with larger margins, their contribution to the loss becomes more significant. This encourages the model to focus on refining embeddings in these challenging cases, improving the discriminative power of the feature space. In contrast, *trivial triplets* are deemphasized, as the margin remains at a base value, preventing unnecessary penalization and saving computational resources. This dynamic approach prevents fast learning convergence due to zero loss, which is often noticed in traditional methods. By prioritizing *hard triplets* and reducing emphasis on easy ones, the adaptive margin approach provides a more robust, discriminative, and efficient learning process. The learned feature embeddings can be used to perform image retrieval based on similarity matching between the query and archive images. To this end, the proposed novel triplet training approach with adaptive margin dynamically adjusts the margin based on the hardness of the triplets, providing a significant improvement in performance over traditional methods.

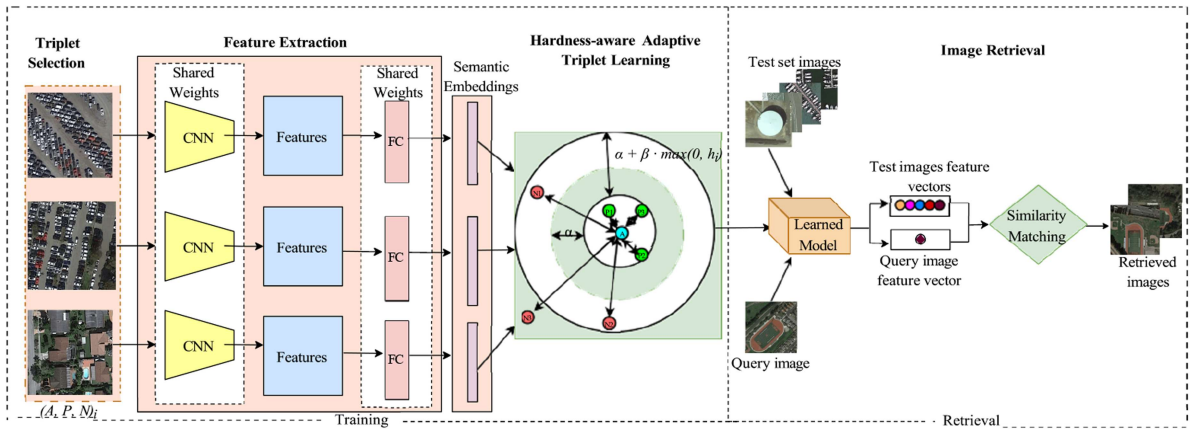


Figure 2. HA²TL-CBIR: Adaptive triplet learning approach for content-based image retrieval.

3.2. Hardness calculation

To compute the hardness, we obtain the feature vectors $\mathbf{F}^i$, $\mathbf{F}^i$ and $\mathbf{F}^i$ of the triplet $A P N$ set $(A, P, N)_i$, respectively, through a CNN-based backbone network. To calculate the hardness coefficient (h) corresponding to the triplet selected, we initially compute the Euclidean distance (d) between feature vectors that corresponds to $(A, P)_i$ and $(A, N)_i$ image pairs as,

$$d_{AP}^i = \frac{q}{F_A^i - F_P^{i^2}} \quad (1)$$

$$d_{AN}^i = \frac{q}{F_A^i - F_N^{i^2}} \quad (2)$$

Then, the hardness coefficient is obtained as,

$$h^i = d_{AP}^i - d_{AN}^i \quad (3)$$

where h^i represent the hardness corresponding to the i^{th} triplet $(A, P, N)_i$ selected from the mini-batch $\mathbf{B}$. A positive h^i indicates that the negative image is closer to the anchor compared to the positive, making the triplet difficult and more informative. In contrast, a negative or zero h^i implies the model has already learned to separate the samples and such *trivial triplets* offer little learning value. This hardness measure serves as a measure to assess the relative difficulty and relevance of triplets for effective learning.

3.3. Adaptive margin calculation

To facilitate adaptive margin computation, we propose two factors: i) a base margin; and ii) the scaling factor. Base margin (α) sets the minimum margin between different triplets whereas the scaling factor (β) refers to the margin that needs to be considered based on the *hardness* of the triplet. We compute the adaptive margin ($m_{adaptive}$) as,

$$m_{adaptive} = \alpha + \beta \max(0, h^i) \quad (4)$$

For *hard triplets*, where h^i is positive, the margin increases proportionally to the hardness, while for *trivial triplets* where h^i is zero or negative, the margin remains at the base margin value (α). This formulation ensures that harder triplets receive proportionally larger margins, amplifying their contribution to the loss. *Trivial triplets*, where $h^i \leq 0$, retain the base margin α and contribute minimally to the training process. This adaptive behaviour helps avoid the dominance of trivial examples and encourages the model to focus on refining embeddings in challenging regions of the feature space. The ranges for the base margin (α) and scaling factor (β) were selected based on prior deep metric learning literature, where effective margins typically lie between 0.1 and 1.0, ensuring that the chosen values remain a stable and meaningful proportion of the embedding distances and support stable gradient behaviour during training (Schroff et al. 2015; Hermans et al. 2017; Ha and Blanz 2021). In addition to prior literature, the specific grid search ranges used in this work were determined through preliminary empirical analysis. When the base margin was set to values $\alpha < 0.2$, the model rapidly converged to zero loss, indicating insufficient separation between positive and negative pairs. Conversely, using $\alpha > 0.6$ or $\beta > 0.5$ led to unstable training behaviour with fluctuating gradients. These patterns are consistent with observations reported in adaptive-margin and triplet-learning studies (Ha and Blanz 2021), which highlight that both α and β must remain moderate fractions of the embedding distances to ensure stable optimization. Therefore, the selected grid search ranges $\alpha \in [0.2, 0.6]$ and $\beta \in [0.1, 0.5]$ represent a rational choice that balances stability, gradient behaviour, and discriminative learning capacity.

3.4. Proposed triplet loss function

The triplet loss function (L) optimizes the embedding space so that similar samples are closer together and dissimilar samples are farther apart by a specified margin (Ha and Blanz 2021). Here, a triplet loss function integrated with adaptive margin calculation is obtained as,

$$L = \max(0, h^i + m) \quad (5)$$

where m represents the fixed margin. If $h^i + m > 0$, then the network will be able to learn, while if $h^i + m \leq 0$, then the loss is zero, because it does not contribute any further information required to train the model. Triplet loss with adaptive margin is as,

$$L_{\text{adaptive}} = \max 0, h^i + m_{\text{adaptive}} \quad (6)$$

where L_{adaptive} and m_{adaptive} are the adaptive triplet loss and adaptive margin, respectively. The adaptive margin reduces the dominance of trivial triplets that typically produce zero loss in fixed-margin triplet training. By prioritizing *hard triplets* and reducing emphasis on easy ones, the adaptive margin triplet loss leads to robust, discriminative, and efficient learning. Unlike traditional fixed-margin triplet loss, where all triplets are penalized equally, the proposed adaptive loss dynamically scales the margin, effectively modulating the gradient strength based on the sample's hardness. This can be interpreted as a form of implicit hard triplet weighting, which improves training efficiency and generalization. The adaptive margin implicitly prioritizes difficult samples by increasing the loss for *hard triplets* and suppressing updates from trivial ones. This behaviour encourages stronger intraclass similarity and interclass separability. Furthermore, by avoiding the need for explicit hard negative mining, it ensures stability and robustness during training, which is particularly valuable for high-resolution RS image archives with subtle semantic distinctions. The main steps involved in the proposed method are summarized in Algorithm 1.

Algorithm 1. Adaptive triplet learning approach for image retrieval.

Require: t test images $X_{\text{test}}^i \subset \mathbf{X}$, $i = 1, 2, \dots, t$; batch size B

Ensure: Top k similar images to query image X_{test}^i

- 1: Initialize the backbone network model M with a 128-dimensional embedding output layer.
- 2: Repeat Steps 3–7 until convergence.
- 3: Create a subset of training image samples: sample a mini-batch B of b triplets $(A, P, N)_i$ using random selection.
- 4: Compute feature embeddings: train the model M using triplets $(A, P, N)_i$ from the batch to obtain 128-dimensional embeddings.
- 5: Perform hardness calculation using $h^i = d_{AP}^i - d_{AN}^i$.
- 6: Compute adaptive margin as $m_{\text{adaptive}} = \alpha + \beta \cdot \max(0, h^i)$.
- 7: Compute batch loss: for all triplets in the batch, compute the triplet loss:

$$L_{\text{adaptive}} = \max 0, h^i + m_{\text{adaptive}} .$$

- 8: Backpropagate the gradients and update the learnable parameters ϑ of model M using the Adam optimizer.
 - 9: Given a query image $X_q \subset \mathbf{X}_{\text{test}}$, retrieve the top k most similar images.
-

4. Experiments

4.1. Dataset description

To evaluate the effectiveness of the proposed approach, we carried out extensive experiments on two benchmark RS image archives. The first one is the PatternNet (Zhou et al. 2018) archive where we consider a subset of 12 classes which includes 9600 images of varying land-use categories (e.g. airfield, anchorage, beach, dense residential, farm, flyover, forest, game space, parking space, river, sparse residential, and storage cisterns). Each class includes 800 images of size 256×256 pixels with a varying spatial resolution between 0.062 m and 4.693 m. The archive includes images of high intra-class dissimilarity and inter-class similarity.

The second one is NWPU-RESISC45 (Cheng et al. 2017) image archive that comprises of a subset of 12 scene classes which includes 8400 images. The subset of classes used in our work are airfield, anchorage,

beach, dense residential, farm, flyover, forest, game space, parking space, river, sparse residential, and storage cisterns. Each class includes 700 images each of size 256×256 pixels of varying spatial resolution 30 m to 0.2 m.

4.2. Design of experiments and sensitivity analysis

To perform experiments, we chose three convolutional neural network (CNN) architectures as the backbone network to integrate with our proposed HA²TL-CBIR model: ResNet18 (He et al., 2016), DenseNet121 (Huang et al. 2017) and EfficientNet-B0 (Tan and Le 2019). To improve model generalization, data enhancement techniques such as random cropping and horizontal flipping were applied during training. The batch size is 32. The model was trained using the Adam optimizer with its default momentum parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$). A fixed learning rate of 0.0001 was used throughout training without any decay schedule.

Early stopping was not used; instead, the network was trained for a fixed 50 epochs, as stable convergence was consistently observed. All experiments were performed on a GPU NVIDIA P100. For our experiments, we considered a training, validation as well as test ratio of 80%, 10% and 10%, respectively. To quantitatively evaluate HA²TL-CBIR, we used F1-score, recall, mean average precision (mAP), average normalized modified retrieval rank (ANMRR), and Precision@K (P@K) metrics where K is the K retrieved images. The query images were selected from the test set and retrieved top 50 most similar images for each query based on cosine similarity in the embedding space. To determine the optimal values of the adaptive margin parameters—base margin (α) and scaling factor (β)—we conducted a grid-based sensitivity analysis. We evaluated combinations of $\alpha \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$ and $\beta \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ using validation loss as the criterion across all backbone-dataset configurations. For instance, the combination $\alpha = 0.2$, $\beta = 0.3$ achieved the lowest validation loss of 0.0642 for DenseNet121 (Huang et al. 2017) on PatternNet, while $\alpha = 0.2$, $\beta = 0.2$ achieved a loss of 0.0147 for ResNet18 on the same dataset (see Table 1). In contrast, EfficientNet-B0 demonstrated less stable performance, especially on NWPU-RESISC45. These results highlight that assigning moderate margins to hard triplets helps the model focus on informative examples, while avoiding over-penalization from trivial triplets. The optimal α - β values from this analysis were used in all subsequent experiments.

5. Results

In this section, we present the performance of the proposed HA²TL-CBIR model with varying backbone DL architectures for both PatternNet and NWPU-RESISC45 archives qualitatively as well as quantitatively (see Table 2). The DenseNet121 DL architecture outperformed the other networks and is therefore considered as the backbone network for our subsequent experiments. On NWPU-RESISC45, DenseNet121 improves mAP by 6%, recall by 1%, and F1-score by 2% compared to ResNet18. In addition, DenseNet121 achieved a lower ANMRR score, indicating that it consistently retrieves more relevant images within the top ranks. These results suggest that DenseNet121 is more effective in capturing both fine-grained local details and broader contextual patterns in RS images, making it particularly well-suited for retrieval tasks involving high intra-class variability.

Figure 3 presents the qualitative results of the proposed HA²TL-CBIR model when the query images considered are selected from the class: *Dense Residential* and *Storage Cisterns*. The figure illustrates the top-50 retrieved images for each query, allowing a clear comparison of retrieval consistency. Each retrieved sample is also annotated with its cosine similarity score relative to the query to improve interpretability. By

Table 1. Optimal α - β combinations and validation loss across networks and datasets.

Backbone	Dataset	α	β	Validation Loss
DenseNet121	PatternNet	0.2	0.3	0.0642
	NWPU-RESISC45	0.2	0.1	0.1843
ResNet18	PatternNet	0.2	0.2	0.0147
	NWPU-RESISC45	0.3	0.2	0.0244
EfficientNet-B0	PatternNet	0.2	0.1	0.0876
	NWPU-RESISC45	0.4	0.4	0.7078

Table 2. Performance metrics for different backbone networks in HA²TL-CBIR.

Archive	Backbone Network	P@5	P@10	P@50	mAP	Recall	F1-Score	ANMRR
NWPU-RESISC45	ResNet18 (He et al. 2016)	0.80	0.80	0.76	0.86	0.79	0.78	0.03
	EfficientNet-B0 (Tan and Le 2019)	0.53	0.53	0.47	0.70	0.52	0.48	0.03
	DenseNet121 (Huang et al. 2017)	1.00	0.89	0.88	0.92	0.80	0.80	0.02
	ResNet18 (He et al. 2016)	1.00	0.90	0.88	0.99	0.98	0.97	0.02
PatternNet	EfficientNet-B0 (Tan and Le 2019)	0.95	0.91	0.92	0.92	0.92	0.91	0.02
	DenseNet121 (Huang et al. 2017)	1.00	1.00	0.99	0.99	0.99	0.99	0.02

Note: Bold values indicate the best-performing results among the compared methods for each evaluation metric.

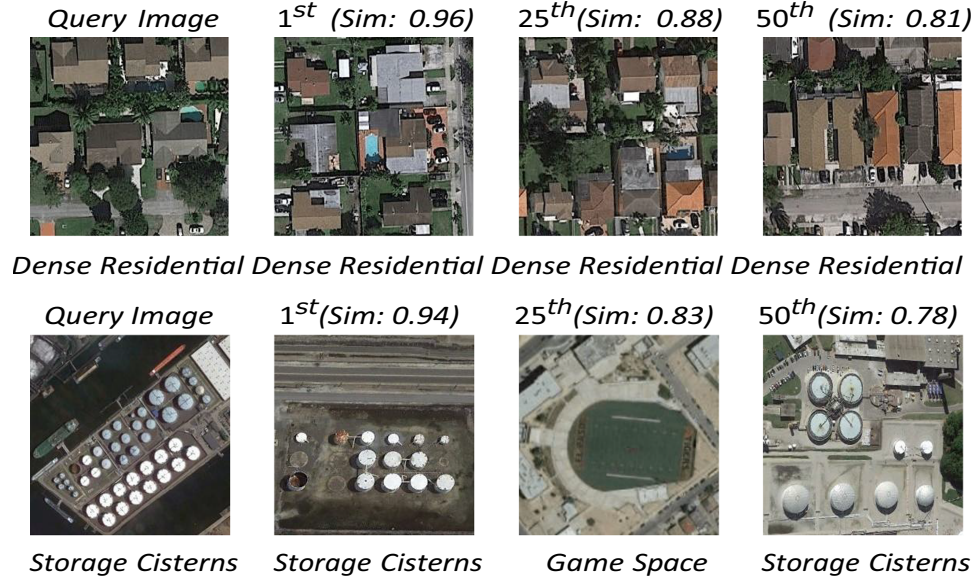


Figure 3. Qualitative retrieval results of the proposed HA²TL-CBIR using DenseNet121 on the NWPU-RESISC45 archive. Each retrieved image is annotated with its cosine similarity score relative to the query. Semantic relevance is interpreted based on (i) class-level consistency, (ii) spatial-structural similarity, and (iii) visual resemblance of key land-use patterns.

analysing the retrieval results obtained for each class separately, one may observe that the proposed method retrieves similar images with challenging semantic contents. For *Storage Cisterns*, one can observe that the retrieved images are semantically relevant to the provided query image, while in the case of *Dense Residential* scenes, the results include houses, roads, and other urban structures that closely resemble the spatial arrangement of the query. This highlights the ability of the method to capture subtle but discriminative features. Semantic relevance in these qualitative results is assessed based on three criteria: (i) class-level consistency between the query and retrieved samples, (ii) similarity in spatial and structural scene patterns, and (iii) visual resemblance of key land-use attributes such as object distribution, texture, and colour.

All experiments were conducted on an NVIDIA Tesla P100 GPU. To enable a fair comparison independent of implementation details, we report the per-epoch training time for fixed-margin and adaptive-margin triplet learning, followed by the total time for 50 epochs. For the PatternNet dataset (7,680 training images), fixed-margin training required 86.0 seconds per epoch, whereas the proposed HA²TL-CBIR required 90.4 seconds per epoch. Over 50 epochs, this corresponds to ≈ 71.7 minutes (fixed) and ≈ 75.3 minutes (adaptive), reflecting an overhead of only $\approx 5.0\%$. For the NWPU-RESISC45 dataset (6,720 training images), fixed-margin training required 78.0 seconds per epoch, while adaptive-margin training required 82.1 seconds per epoch, giving total training times of ≈ 65.0 minutes (fixed) and ≈ 68.4 minutes (adaptive), an overhead of $\approx 5.2\%$.

The adaptive margin introduces only lightweight scalar computations (distance subtraction, max operator, and margin scaling; see Equations (3,4), Page 6) and therefore adds no additional learnable

parameters. Importantly, inference speed remains unchanged: DenseNet121 produces a 128-D embedding in ≈ 3.1 ms, and cosine similarity matching adds < 1 ms, yielding a total retrieval time of ≈ 4 ms per query. These results demonstrate that HA²TL-CBIR maintains computational efficiency very close to fixed-margin training while providing significantly improved retrieval performance, confirming its practicality for large-scale RS archives.

These qualitative outcomes further support the quantitative metrics, confirming that the proposed adaptive margin strategy improves both retrieval accuracy and semantic consistency.

We assessed the effectiveness of the HA²TL-CBIR model using nine state-of-the-art methods: (i) DML using triplet loss (DML-TL) (Cao et al. 2020); (ii) the triplet non-local neural network with dual-anchor triplet loss (T-NLNN+DATL) (Zhang et al. 2021); (iii) the active learning method with weighted subspace clustering (ALM-WS) (Sudha and Aji, 2022); (iv) the triplet graph CNN (CNN-TL) (Regan et al. 2023); (v) adaptive margin triplet loss (AMTL) (Ha and Blanz 2021); (vi) the SimCLR-based contrastive learning method DMBSS (Liu et al. 2024); (vii) the AWCloss-Hashing contrastive framework (Wang et al. 2025); (viii) TDO-Net (Hou et al. 2022), an attention-enhanced end-to-end discriminative network with multi-scale feature learning; and (ix) the Proxy-Based Rotation-Invariant Deep Metric Learning framework (PBFFN) (Cai et al. 2024).

Among these methods, DML-TL, T-NLNN+DATL, CNN-TL, and TDO-Net follow DML paradigms, with DML-TL relying on a fixed-margin triplet formulation. AMTL employs an adaptive margin strategy, while T-NLNN+DATL incorporates non local feature aggregation to strengthen global context modelling, and TDO-Net enhances discriminative capability through multiscale contextual feature learning. In contrast, DMBSS and AWCloss-Hashing represent recent contrastive-learning approaches. DMBSS adapts SimCLR for remote sensing retrieval using dynamic manifold-based sample selection, whereas AWCloss-Hashing improves contrastive representation learning by assigning Gaussian-based weights to positive and negative pairs and suppressing extremely hard negatives via automatic weight allocation. PBFFN serves as a recent proxy-based DML baseline, combining proxy Euclidean contrast loss with multi-layer feature fusion to achieve rotation-invariant embeddings.

Table 3 shows the comparison of the proposed HA²TL-CBIR method with the state-of-the-art approaches. The proposed HA²TL-CBIR achieves state-of-the-art retrieval performance, matching or exceeding the strongest recent baselines on both PatternNet and NWPU-RESISC45. On PatternNet, HA²TL-CBIR achieves P@50 = 0.99, matching the strongest recent baselines. Figure 4 presents a qualitative comparison between the DML-TL baseline and the proposed HA²TL-CBIR method for the storage cisterns class. The performance on the PatternNet archive is higher due to the larger spatial resolution of its images. Classical triplet-learning methods typically assume a fixed margin across varying triplet pairs, limiting their adaptability to samples of different hardness. Although AMTL introduces an adaptive margin, its formulation does not sufficiently amplify hardness-based differences between positive and negative pairs, thereby restricting the discriminative information available to the model. The results demonstrate the importance of adopting a hardness-aware adaptive triplet learning strategy to enhance the discriminative capability of the embedding space.

Table 3. Comparison of HA²TL-CBIR with state-of-the-art methods.

Archive	Method	P@5	P@10	P@50
NWPU-RESISC45	ALM-WS (Sudha and Aji 2022)	0.66	0.62	0.59
	CNN-TL (Regan et al. 2023)	0.72	0.70	0.68
	DML-TL (Cao et al. 2020)	0.68	0.65	0.59
	TDO-Net (Hou et al. 2022)	0.88	0.88	0.88
	AWCloss-Hashing (Wang et al. 2025)	0.87	0.84	0.78
	HA²TL-CBIR	1.00	0.89	0.88
	ALM-WS (Sudha and Aji 2022)	0.92	0.92	0.93
PatternNet	CNN-TL (Regan et al. 2023)	0.93	0.93	0.92
	DML-TL (Cao et al. 2020)	0.94	0.93	0.94
	AMTL (Ha and Blanz 2021)	0.96	0.95	0.91
	T-NLNN+DATL (Zhang et al. 2021)	0.99	0.99	0.98
	TDO-Net (Hou et al. 2022)	0.99	0.99	0.99
	PBFFN (Cai et al. 2024)	0.99	0.99	0.99
	DMBSS (Liu et al. 2024)	–	0.98	0.97
	HA²TL-CBIR	1.00	1.00	0.99

Note: Bold values indicate the best-performing results among the compared methods for each evaluation metric.

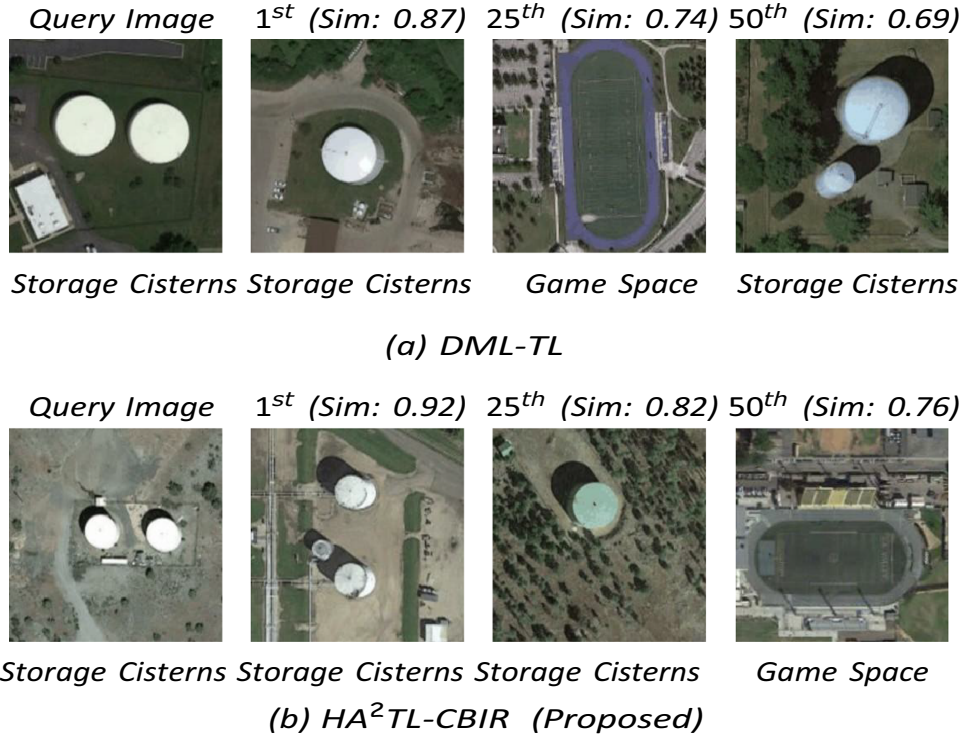


Figure 4. Qualitative comparison of (a) DML-TL (Cao et al. 2020) and (b) HA²TL-CBIR on the PatternNet archive. For each method, the 1st, 25th and 50th retrieved images with respect to the query are shown and annotated with their cosine similarity scores (Sim) relative to the query. Semantic relevance is interpreted based on (i) class-level consistency, (ii) spatial-structural similarity, and (iii) visual resemblance of key land-use patterns.

Both PatternNet and NWPU-RESISC45 differ substantially in spatial resolution (PatternNet: 0.062–4.693 m; NWPU-RESISC45: 0.2–30 m), which provides an implicit cross-resolution evaluation setting. The proposed HA²TL-CBIR model exhibits consistently strong retrieval accuracy on both datasets, suggesting that the adaptive-margin mechanism remains robust to variations in spatial detail. Higher-resolution images in PatternNet allow the model to capture finer local structures, while the adaptive margin helps maintain discriminative embeddings even for the coarser-resolution NWPU-RESISC45 imagery. This indicates that the model generalizes effectively across different resolutions without requiring explicit cross-resolution training.

To assess the discriminative structure of the learned embeddings, we visualize the 128-dimensional feature space using t-SNE and compare the fixed-margin and adaptive-margin triplet models (Figure 5). The adaptive-margin model produces visibly tighter intra-class clusters and clearer inter-class boundaries. To quantitatively validate this observation, we compute two standard cluster-validity indices: the silhouette score (higher is better) (Rousseeuw 1987) and the Davies–Bouldin index (lower is better) (Davies and Bouldin 1979). The adaptive-margin model achieves a higher silhouette score (0.4835) than the fixed-margin model (0.4736) and a lower Davies–Bouldin index (0.8301 vs. 0.8595). These improvements confirm that the proposed hardness-aware adaptive margin encourages a more compact and discriminative embedding space.

6. Conclusion

This paper presented the proposed HA²TL-CBIR approach, a content-based image retrieval framework for remote sensing imagery that leverages triplet loss with an adaptive margin strategy. Unlike traditional triplet loss, which relies on a fixed margin, HA²TL-CBIR dynamically adjusts the margin according to the hardness of triplets, enabling the model to prioritize challenging examples during training. This hardness-aware mechanism strengthens inter-class separability, reduces intra-class variation, and results in more discriminative feature embeddings. Experimental evaluation on benchmark datasets demonstrates that the

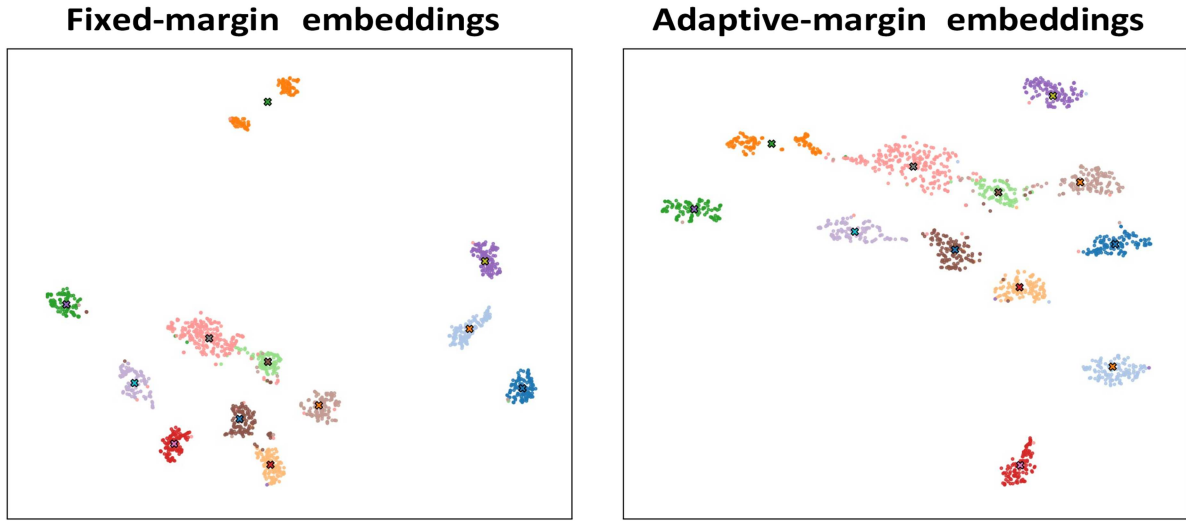


Figure 5. t-SNE visualization of the 128-dimensional embeddings learned by DenseNet121 on the PatternNet dataset. Left: fixed-margin triplet model. Right: adaptive-margin triplet model. The adaptive-margin configuration produces tighter intra-class clusters and clearer inter-class separation.

proposed method achieves superior retrieval performance in terms of precision, recall, F1-score, and ANMRR compared to conventional triplet loss and related baselines.

The results confirm the robustness and generalization capability of HA²TL-CBIR. Consistent improvements were observed across multiple datasets, particularly in scenarios characterized by high intra-class variability and visually similar inter-class samples. These improvements highlight the ability of the adaptive margin strategy to capture subtle semantic distinctions in complex remote sensing scenes. In addition to quantitative performance, qualitative retrieval results exhibit stronger semantic consistency between query and retrieved images, demonstrating the framework's effectiveness in capturing both fine-grained local features and global contextual structures.

Overall, the proposed HA²TL-CBIR framework provides an effective and scalable solution for remote sensing CBIR. By addressing key limitations of traditional triplet loss and incorporating a hardness-aware adaptive margin mechanism, it makes a significant contribution to the state of the art. These advances enhance retrieval accuracy, robustness, and efficiency, thereby reinforcing the practical value of deep metric learning for large-scale remote sensing image retrieval.

Looking forward, this work can be extended by exploring Transformer-based architectures to better capture long-range spatial dependencies, as well as graph neural networks to model structural relationships within scenes. Incorporating multi-modal data sources—such as spectral, LiDAR, or textual information may further enhance robustness across diverse remote sensing environments. Developing lightweight backbone networks can improve computational efficiency for real-time or large-scale retrieval. Finally, adapting the proposed adaptive margin strategy to multi-label, cross-resolution, or self-supervised retrieval settings represents a promising direction for broadening the applicability of the framework.

Acknowledgements

Akshara Preethy Byju conceived and supervised the study, developed the methodology alongside Aswathy K R, and oversaw project administration. Aswathy K R conducted the investigation, performed the formal analysis and validation, and prepared the original draft. Akshara Preethy Byju and Alberto S Garea contributed to the review and editing of the manuscript. Visualization was carried out by Aswathy K R, Akshara Preethy Byju, and Alberto S Garea. All authors reviewed and approved the final version of the manuscript for publication.

Author contributions

CRedit: **Aswathy K R:** Formal analysis, Investigation, Methodology, Validation, Writing – original draft; **Akshara Preethy Byju:** Conceptualization, Methodology, Supervision; **Jorge Alberto Suárez Garea:** Supervision, Visualization, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported in part by grants PID2022–141623NB–I00 and TED2021–130367B–I00 funded by MCIN/AEI/10.13039/501100011033 and by ‘European Union NextGenerationEU/PRTR’. It was also supported by Xunta de Galicia - Consellería de Educación, Ciencia, Universidades e Formación Profesional [Centro de investigación de Galicia accreditation 2024-2027 ED431G-2023/04 and Reference Competitive Group accreditation, ED431C-2022/16], and by ‘ERDF/EU’.

Data availability statement

The standard benchmark archives NWPU-RESISC45 and PatternNet dataset is publicly available at doi.org/10.1109/JPROC.2017.26759984 and doi.org/10.48550/arXiv.1706.03424, respectively.

References

- Ayush K, Uz kent B, Chugh T, Burke M, Lobell D. 2021. Geography-aware self-supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). p 10152–10161.
- Cai Z, Pan Y, Jin W. 2024. Proxy-based rotation invariant deep metric learning for remote sensing image retrieval. *IEEE J Sel Top Appl Earth Obs Remote Sens.* 17:7759–7772. <https://doi.org/10.1109/JSTARS.2024.3382845>
- Cao R et al. 2020. Enhancing remote sensing image retrieval using a triplet deep metric learning network. *IJRS.* 41:740–751. <https://doi.org/10.1080/2150704X.2019.1647368>
- Cheng G, Han J, Lu X. 2017. Remote sensing image scene classification: benchmark and state of the art. *Proc IEEE.* 105:1865–1883. <https://doi.org/10.1109/JPROC.2017.2675998>
- Cheng G, Yang C, Yao X, Guo L, Han J. 2018. When deep learning meets metric learning: remote sensing image scene classification via learning discriminative CNNs. *IEEE Trans Geosci Remote Sens.* 56:2811–2821. <https://doi.org/10.1109/TGRS.2017.2783902>
- Chowdhury S, Ma L, Wang S, Zhang Y. 2021. A comprehensive review of deep learning for remote sensing image retrieval. *Remote Sens.* 13:4000.
- Davies DL, Bouldin DW. 1979. A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell.* PAMI-1(2):224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Feng Y et al. 2020. Triplet distillation for deep face recognition. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), 25–28 October 2020. IEEE: Abu Dhabi, United Arab Emirates. p 808–812.
- Guo H, Wang J, Li B, Du Q. 2022. Deep metric learning with lightweight CNNs for remote sensing image retrieval. *IEEE Geosci Remote Sens Lett.* 19:1–5.
- Ha ML, Blanz V. 2021. Deep ranking with adaptive margin triplet loss. *arXiv:2107.06187*.
- Harwood B, Kumar BGV, Carneiro G, Reid I, Drummond T. 2017. Smart mining for deep metric learning. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 22–29 October 2017. IEEE: Venice, Italy. p 2821–2829.
- He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. p 770–778.
- Hermans A, Beyer L, Leibe B. 2017. In Defense of the Triplet Loss for Person Re-Identification. *arXiv 2017.* *arXiv:1703.07737*.
- Hou J, Li J, Wang S, Ding L, Jiao L. 2022. An attention-enhanced end-to-end discriminative network with multiscale feature learning for remote sensing image retrieval. *IJSTA.* 15:8275–8290. <https://doi.org/10.1109/JSTARS.2022.3191222>
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 21–26 July 2017. Honolulu, HI, USA. p 4700–4708.
- Jiang H, Xu J, Zhao J, Ma A, Liu H. 2023. A survey on deep learning-based remote sensing image retrieval. *IEEE Access.* 11:29915–29931.
- Kertész G. 2022. Combining negative selection techniques for triplet mining in deep metric learning. In: Proceedings of the IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC). IEEE: Óbuda, Hungary. p 155–160.
- Kim S, Kim D, Cho M, Kwak S. 2020. Proxy anchor loss for deep metric learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 13–19 June 2020. Seattle, WA, USA. p 3238–3247.
- Liang J, Zhao W, Chen X, Li J. 2021. Low-altitude remote sensing opium poppy image detection based on modified YOLOv3. *Remote Sensing Image Retrieval Based on Efficient-Net and Metric Learning.* *Remote Sensing.* 13:1532. <https://doi.org/10.3390/rs13112130>

- Liu Q, Ge Y, Wang S, Wang T, Xu J. 2024. Dynamic manifold-based sample selection in contrastive learning for remote sensing image retrieval. *The Visual Computer*. 41(6):4111–4127.
- Movshovitz-Attias Y, Toshev A, Leung TK, Ioffe S, Singh S. 2017. No fuss distance metric learning using proxies. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 22–29 October 2017. Venice, Italy. p 360–368. <https://doi.org/10.1109/ICCV.2017.47>
- Ni Z, Xu Y, Zhang W, Zong Z, Ren P. 2026. A Size-aware graph embedding approach to remote sensing image captioning with object relative size information. *IEEE Trans Geosci Remote Sens*. 64:1–13. <https://doi.org/10.1109/TGRS.2026.3650788>
- Penatti O.A.B., Nogueira K, dos Santos J.A. 2015. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 7–12 June 2015. IEEE: Boston, MA, USA. p 44–51.
- Regan A, Smith J, Li H. 2023. Triplet graph convolutional network for remote sensing image retrieval. *IEEE Trans Geosci Remote Sens*. 61:1–14.
- Rousseeuw PJ. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *JCoAM*. 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Schroff F., Kalenichenko D, Philbin J. 2015. A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7–12 June 2015. IEEE: Boston, MA, USA. p 815–823.
- Sohn K. 2016. Improved deep metric learning with multi-class n-pair loss objective. In: *Advances in Neural Information Processing Systems (NeurIPS)*, 5–10 Dec 2016. Barcelona, Spain. p 1857–1865.
- Song Ho, Xiang Y, Jegelka S., Savarese S. 2016. Deep metric learning via lifted structured feature embedding. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 27–30 June 2016. IEEE: Las Vegas, NV, USA. p 4004–4012.
- Sudha S, Aji S. 2022. An active learning method with entropy weighting subspace clustering for remote sensing image retrieval. *Appl Soft Comput*. 125:109107. <https://doi.org/10.1016/j.asoc.2022.109107>
- Tan M, Le QV. 2019. EfficientNet: rethinking model scaling for convolutional neural networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*, 9–15 June 2019. Long Beach, CA, USA. p 6105–6114.
- Teh EW, DeVries T, Taylor GW. 2020. Proxynca++: revisiting proxy anchor loss for deep metric learning. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 1–5 March 2020. IEEE/CVF: Snowmass Village, CO, USA. p 3452–3460.
- Wang J et al. 2014. Learning fine-grained image similarity with deep ranking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 23–28 June 2014. IEEE: Columbus, OH, USA. p 1386–1393. <https://doi.org/10.1109/CVPR.2014.180>
- Wang X et al. 2025. Automatic weight allocation: optimizing remote sensing image retrieval from a contrastive learning perspective. *Multimedia Syst*. 31:1–15. <https://doi.org/10.1007/s00530-025-01824-3>
- Wang X, Han X, Huang W, Dong D, Scott MR. 2019. Multi-similarity loss with general pair weighting for deep metric learning. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16–20 June 2019. IEEE/CVF: Long Beach, CA, USA. p 16–20.
- Wu CY, Manmatha R, Smola AJ, Krahenbuhl P. 2017. Sampling matters in deep embedding learning. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 22–29 October 2017. IEEE: Venice, Italy. p 2840–2848.
- Yang H, Li Y, Xu Y, Jiao L, Liu F. 2023. Remote sensing image retrieval via deep metric learning: a review. *IJGI*. 12:128.
- Yuan Y, Zhang X, Yang Q. 2021. Revisiting triplet loss with label distribution smoothing for person re-identification. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 19–25 June 2021. IEEE/CVF: Nashville, TN, USA. p 12533–12542.
- Zhang M, Cheng Q, Luo F, Ye L. 2021. A triplet nonlocal neural network with dual-anchor triplet loss for high-resolution remote sensing image retrieval. *IJSTA*. 14:2711–2723. <https://doi.org/10.1109/IJSTARS.2021.3058691>
- Zhou W, Newsam S, Li C, Shao Z. 2018. PatternNet: a benchmark dataset for performance evaluation of remote sensing image retrieval. *ISPRS J Photogramm Remote Sens*. 145:197–209. <https://doi.org/10.1016/j.isprsjprs.2018.01.004>