Plasticity-Stability Preserving Multi-Task Learning for Remote Sensing Image Retrieval

Gencer Sumbul, *Graduate Student Member, IEEE*, and Begüm Demir, *Senior Member, IEEE*

***Abstract*— Deep learning-based multi-task learning (MTL) methods have recently attracted attention for content-based image retrieval (CBIR) applications in remote sensing (RS). For a given set of tasks (e.g., scene classification, semantic segmentation, and image reconstruction), existing MTL methods employ a joint optimization algorithm on the direct aggregation of task-specific loss functions. Such an approach may provide limited CBIR performance when: 1) tasks compete or even distract each other; 2) one of the tasks dominates the whole learning procedure; or 3) characterization of each task is underperformed compared to single-task learning. This is mainly due to the lack of: 1) plasticity condition (which is associated with sensitivity to new information) or 2) stability condition (which is associated with protection from radical disruptions by new information) of the whole learning procedure. To avoid this issue, as a first time, we propose a novel plasticity-stability preserving MTL (PLASTA-MTL) approach to ensure the plasticity and the stability conditions of the whole learning procedure independently of the number and type of tasks. This is achieved by defining two novel loss functions. The first loss function is the plasticity preserving loss (PPL) function that aims to enforce the global image representation space to be sensitive to new information learned with each task. This is achieved by minimizing the difference of gradient magnitudes for the global representation and task-specific embedding spaces. The second loss function is the stability preserving loss (SPL) function that aims to protect the global representation space radically disrupted by a new task. This is achieved by minimizing the angular distances between the task gradients over global representation space. To effectively employ the proposed loss functions, we also introduce a novel sequential optimization algorithm. Experimental results show the effectiveness of the proposed approach compared to the state-of-the-art MTL methods in the context of CBIR.**

***Index Terms*— Image retrieval, multi-task learning (MTL), remote sensing (RS), representation learning.**

I. Introduction

THE development of efficient and effective methods for content-based image retrieval (CBIR) from large-scale remote sensing (RS) image archives is one of the growing research interests in RS. CBIR aims at searching for RS images similar to a query image based on their semantic content. Traditional CBIR methods extract and exploit handcrafted image representations, which are manually designed based on the spatial and spectral characteristics of RS images [1]–[4]. In recent years, deep learning (DL)-based image representation learning has been found very effective for CBIR problems due to its capability to construct multiple levels of representations in a hierarchical manner [5]. Unlike the traditional methods, in DL-based CBIR methods, image representations are automatically learned during the optimization of an objective function based on the characteristics of a learning paradigm (i.e., task). For DL-based RS image representation learning, existing CBIR methods utilize following tasks: 1) scene classification [6]–[16]; 2) similarity learning [17]–[29]; 3) image reconstruction [30], [31]; 4) semantic segmentation [32], [33]; and 5) image captioning [34]. Each task has different objectives that lead to different optimization procedures throughout the training of the considered deep neural network (DNN). Accordingly, learned image representations have different characteristics for different tasks and, thus, carry different information to be utilized in CBIR applications. As an example, when the task is scene classification, RS image representations can be learned with convolutional neural networks (CNNs) by optimizing entropy-based loss functions. In this way, image representations are encoded to separate predefined classes that maximize interclass distances in the image representation space [35]. For the similarity learning task, on the other hand, image representations are learned to discriminate dissimilar RS images that minimize intraclass distance in the image representation space [36]. This can be achieved by employing Siamese CNNs on tuples of RS images to optimize triplet or contrastive loss functions. If the task is chosen as the image reconstruction, autoencoder neural networks can be used first to construct the representations and then to recover RS images with reconstruction loss. In this way, resulting image representations are robust to noise in RS images [37]. In RS, it is common to use the abovementioned tasks in the framework of single-task learning (STL) for CBIR applications.

However, using a single task may not be sufficient to describe the complex content of RS images in CBIR problems. To address this issue, multiple tasks can be jointly utilized for image representation learning. When image representation learning is achieved based on multiple tasks, the resulting latent space can better represent the complex semantic content of RS images. Accordingly, few DL-based multi-task learning (MTL) methods have been recently introduced in RS for CBIR applications. As an example, in [38], RS image similarity learning based on triplet loss is combined with the

Manuscript received July 22, 2021; revised December 1, 2021; accepted February 26, 2022. Date of publication March 16, 2022; date of current version April 12, 2022. This work was supported by the European Research Council (ERC) through the ERC-2017-STG BigEarth Project under Grant 759764. *(Corresponding author: Gencer Sumbul.)*

The authors are with the Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, 10623 Berlin, Germany (e-mail: gencer.suembuel@tu-berlin.de; demir@tu-berlin.de).

Digital Object Identifier 10.1109/TGRS.2022.3160097

scene classification task. In this method, task-specific heads are combined with the CNN backbone shared by two tasks, while the joint optimization of task-specific loss functions is employed by minimizing the summation of them. In this way, MTL is regarded as a joint optimization problem based on the aggregation of task-specific loss functions. This is followed by most of the MTL methods in RS.

Due to the complexity of the MTL problem, it is common that: 1) tasks may compete or even distract each other during training; 2) one of the tasks may dominate the whole learning procedure; or 3) characterization of each task can be underperformed compared to STL [39]. These problems undermine the effectiveness of the whole representation learning procedure [40]. These issues occur due to the stability-plasticity constraint of MTL [41]. MTL methods require being sensitive to new information learned from each task that allows the contribution of each task to further improve modeling the image characterization. This condition is known as plasticity [41]. If there is a lack of plasticity condition in response to new information, an image representation space will be slightly affected while learning a new task and, thus, will merely reflect different characteristics of representations learned via different tasks. If the considered DNN suffers from the lack of plasticity condition, information specific to each task will be only encoded in the corresponding task-specific head. The possible drawbacks of this issue are twofold. First, only the general features of RS images can be encoded in the CNN backbone, and thus, image features extracted from the considered DNN will have a lower discrimination capability compared to STL. Second, one of the tasks can dominate the global image representation space. In this case, all tasks except the one, which dominates the image representation space learned via the backbone, will not significantly affect the image features. For MTL, during the learning process of a new task, new information encoded in the considered DNN should not radically disrupt what is already characterized based on the other tasks. This condition is known as stability. When there is a lack of stability condition in response to new information captured via a new task, there is a risk that previous information encoded by the considered DNN can be forgotten. Thus, a global image representation space will be mainly characterized based on the characteristics of representations learned via a single task. This risk is more evident when some of the tasks compete with each other. In this case, since every task aims to radically change the global image representation space compared to other tasks, tasks may distract each other that leads to less accurate RS image characterization for MTL compared to STL.

The MTL formulation of the existing DL-based CBIR methods (which is based on joint optimization) is limited to control the learning of each task. Thus, it does not allow the controlling plasticity and stability of the whole learning procedure. It is also worth noting that, in the abovementioned MTL formulation, the whole learning procedure is sensitive to the proper selection of loss function weight for each task that generally requires a grid search (which is computationally demanding) [42]. Thus, MTL methods that can effectively combine multiple tasks without the need for selection of loss weights while considering the stability-plasticity problem are needed in the context of CBIR to accurately apply RS image representation learning.

To avoid the abovementioned problems, as a first time, we propose a novel **PLA**sticity-**STA**bility preserving MTL (PLASTA-MTL) approach. The PLASTA-MTL approach aims to preserve: 1) the plasticity for each task and 2) the stability in between learning consecutive tasks for the whole learning procedure independently of the number of tasks and the type of tasks. To this end, we introduce novel plasticity preserving and stability preserving loss (SPL) functions. The plasticity preserving loss (PPL) function enforces the global image representation space (which is shared by all the tasks) to be sensitive to new information learned with each task during training. This is achieved by minimizing the gradient magnitude differences between global image representation and task-specific embedding spaces. The SPL function protects the image representation space radically disrupted by each task during training. This is achieved by minimizing the angular distances between task gradients over global image representation space. To effectively apply these two loss functions, unlike most of the existing MTL methods, we also propose a sequential optimization algorithm. The proposed algorithm aims to adaptively adjust the interactions between task-specific learning procedures, allowing to ensure plasticity and stability conditions for all the tasks. To this end, instead of joint optimization of all loss functions, task-specific objectives together with the PPL function are sequentially optimized. By this algorithm, the SPL function is optimized at the end of the task sequence for all the considered tasks.

The novelty of the proposed PLASTA-MTL approach consists of: 1) the adaptive adjustment of interactions between task-specific learning procedures by the proposed sequential optimization algorithm; 2) the protection of image representation space from radical disruptions that occurred due to each task by the proposed SPL function; and 3) the sensitivity assurance of the image representation space to new information from each task by the proposed PPL function. Due to the proposed sequential optimization algorithm, our PLASTA-MTL approach does not need to select loss function weights for each task. Due to its stability and plasticity preserving capabilities, our PLASTA-MTL approach overcomes the abovementioned MTL problems of joint optimization algorithm, which are mainly conflicts between tasks, the dominance of one of the tasks, and underperformance of tasks compared to STL. It is worth noting that the proposed PLASTA-MTL approach is independent of the number of considered tasks and their types. The code of the proposed approach is publicly available at https://git.tu-berlin.de/rsim/PLASTA-MTL.

The rest of this article is organized as follows. Section II provides the related works. Section III presents the proposed PLASTA-MTL approach. Section IV describes the considered datasets and the experimental setup. Section IV provides the experimental results, while Section VI concludes this article.

II. Related Work

In this section, we initially present the recent advances in single-task-driven CBIR methods in RS and then survey the existing DL-based MTL methods for RS CBIR.

A. Single-Task-Driven CBIR Methods

In the context of DL-based single-task-driven CBIR, an objective function is usually selected on the basis of the characteristics of the considered learning paradigm (i.e., task), and thus, image features are automatically learned during the optimization of this objective function. We categorize the existing methods into five groups based on the tasks that they utilize and survey in the following.

1) Scene Classification-Driven CBIR Methods: The task of scene classification aims at automatically assigning single-labels or multilabels to image scenes. In [15], land-use class probabilities obtained by a CNN are exploited for weighting the distance between a query image and the archive images obtained by conventional distance metrics. In [8], a distance between the image and its land-use class is used to apply reranking on the order of retrieved images. In [11], aggregated deep local features are utilized for query-sensitive CBIR on RS images. To this end, vectors of locally aggregated descriptors obtained via multiplicative and additive attention mechanisms are used to construct memory vectors for expanded query description. In [7], to retrieve similar images to a query image, the fuzzy distance calculation is introduced based on fuzzy rules and image descriptors extracted from a CNN. In [9], query-adaptive feature fusion technique is introduced to employ different hierarchical image representations from a CNN in the context of CBIR.

2) Similarity Learning-Driven CBIR Methods: DL-based similarity learning aims to automatically identify image similarity based on an image representation space, where semantically similar images are located close to each other. In [23], a twin CNN is introduced for the prediction of pairwise image similarity during the hash code generation of RS images. In [17], a triplet deep metric learning network (TDMLN) is introduced for RS image similarity learning. TDMLN utilizes three CNNs with shared model parameters that allow learning RS image similarity through triplet loss function on image triplets, each of which includes anchor, positive, and negative images. TDMLN aims at learning a metric space where the distance between an anchor and its positive image is minimized, and that between the anchor and its negative image is maximized. In [19], a Siamese graph convolutional network is proposed to employ region adjacency graph-based image descriptors for the characterization of pairwise image similarity with a contrastive loss function. In [29], RS image similarity learning based on image triplets is utilized for hash code generation of RS images in the context of CBIR. In [18], the distribution consistency loss function is proposed in the context of deep metric learning to make use of multiple positive and negative images for each anchor image, unlike the triplet loss function. In [24], a quantized DL to hash approach is introduced for efficient CBIR. In this approach, DNN weights and activation functions are binarized, while pairwise image similarity characterization is used for hash code generation of RS images. In [26], the generative adversarial network regularization-based deep metric learning method is introduced to model pairwise image similarity, while a generative adversarial network is used to mitigate the overfitting problem. In [27], a global optimization algorithm is introduced to jointly employ different metric learning-based loss functions on image representations and the retrieval results for the consistency between the loss reduction direction and the optimization direction. In [28], the weighted Wasserstein ordinal loss function is proposed for Siamese CNNs to formulate the image similarity learning problem as an unsupervised deep ordinal classification problem. In [21], the dual-anchor triplet loss function is introduced to make use of more than one anchor for each image triplet (which is achieved by considering the positive image as the second anchor).

3) Image Reconstruction-Driven CBIR Methods: DL-based image reconstruction task aims at automatically reconstructing input images based on unsupervised image representation learning. In [30], a deep bag-of-word method is introduced for CBIR problems. In this method, a convolutional autoencoder (CAE) is utilized to: 1) encode the RS image local areas into a representation space and 2) decode local descriptors to image space. A reconstruction loss function is employed between an image local area and the CAE output, while k-means clustering is used with the bag-of-word approach to define the global image representation. In [31], residual-dyad units (which is the combination of full preactivation block and a convolutional shortcut block) are proposed for CAEs to avoid diminishing feature reuse problem of conventional residual connections.

4) Semantic Segmentation-Driven CBIR Methods: The semantic segmentation task aims to automatically identify pixel-based class labels, which are associated with RS images. In [32], a multilabel CBIR approach based on a fully convolutional network (FCN) is proposed to apply CBIR on local areas of multilabel RS images. The FCN is first trained to predict land-cover maps of RS images, which are then used to characterize convolutional descriptors of image local areas. The set of final local descriptors is utilized for region-based RS image matching. In [33], a graph-theoretic deep representation learning method is introduced to characterize multilabel co-occurrence relationships associated with each RS image in an archive. To this end, a CNN is employed for the automatic prediction of graph-driven region-based image representation with a region representation learning loss function.

5) Image Captioning-Driven CBIR Methods: The image captioning task aims to automatically identify the textual descriptions (i.e., captions) of image scenes. A few DL-based methods employ RS image captioning for CBIR problems. As an example, in [34], image representations are encoded to generate captions of RS images to describe the relationships between the objects in the ground and their attributes, while CBIR is performed by comparing the predicted image captions.

B. Multi-Task-Driven CBIR Methods

MTL aims at enhancing the effectiveness of image representation learning and the prediction accuracy of each task compared to using a separate learning procedure for each task [43]. To this end, the DL-based MTL problem is formulated as learning the model parameters of a DNN with respect to multiple loss functions, each of which is associated with a task. In RS, DL-based MTL has been applied to various applications (e.g., motion deblurring [44], building damage mapping [45], change detection [46], and road extraction [47]). In the context of CBIR, few DL-based MTL methods have been recently proposed in RS and these methods only combine two tasks: 1) scene classification and 2) similarity learning. In [48], a wide-context attention network is introduced to learn the correlation of local descriptors with wide context information by employing channel dependence- and spatial context-attention modules. In [38], a center-metric learning method, which employs the positive–negative center loss function for modeling metric space, is proposed to characterize within-class variations. In [49], a discriminative distillation network is introduced to increase the interclass variations and to reduce the intraclass differences. In [50], a deep hashing CNN is employed for simultaneously generating hash codes and predicting land-use classes of RS images. All the abovementioned deep MTL methods in RS utilize a CNN backbone (which is shared by all tasks) followed by task-specific heads, while image representation learning is done by jointly optimizing the aggregation of task-specific loss functions. Although the main problems of this MTL formulation are separately addressed by automatically selecting loss weights with gradient adjustment strategies in the computer vision domain (e.g., [43] and [51]–[53]), they are still based on the joint optimization algorithm.

III. Proposed Plasticity-Stability Preserving Multi-Task Learning Approach

Let $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$ be an archive that includes M images, where $\boldsymbol{x}_i$ is the ith RS image in the archive $\mathcal{X}$. CBIR methods aim at retrieving images from the archive $\mathcal{X}$ similar to a given query image $\boldsymbol{x}_q$ based on the distances in image representation (i.e., feature) space. Let $\phi : \theta, \mathcal{X} \mapsto \mathbb{R}^\gamma$ be any type of DNN that maps the image $\boldsymbol{x}_i$ to γ-dimensional image descriptor $\phi(\boldsymbol{x}_i; \theta)$, where θ is the set of DNN parameters. Let $\mathcal{T} = \{T_1, \ldots, T_N\}$ be a set of N tasks, where the ith task T_i is associated with a loss function $\mathcal{L}_{T_i}$. When image representation learning is achieved based on multiple tasks, the objective function consists of multiple loss functions $\{\mathcal{L}_{T_i}\}_{i=1}^N$. In this article, MTL is performed by hard parameter sharing technique [39] that allows characterizing a global descriptor for each image based on the multiple tasks. In this way, considered DNN typically includes an encoder (i.e., a CNN backbone), which is shared by all the tasks, and task-specific heads, which are branched out from the CNN backbone. Each task-specific head characterizes the task-specific embedding space based on the characteristics of each task. The CNN backbone models global image representation space. Let $G \in \theta$ be the set of DNN parameters that are used for defining global image representation space. G is chosen as the parameters of the last layer of the CNN backbone shared by all the tasks. Let $E_{T_i} \in \theta$ be the set of parameters that is used to construct the task-specific embedding for the ith task T_i. Accordingly, after learning DNN parameters θ, G is used to extract image descriptors that are utilized to perform CBIR.

In the standard MTL formulation (which is based on joint optimization algorithm), all the model parameters θ, including G and $\{E_{T_i}\}_{i=1}^N$, are simultaneously updated based on the gradients of aggregated loss functions ($\nabla_\theta \sum_i \mathcal{L}_{T_i}$). This MTL formulation is limited to control the learning process of each task and, thus, the plasticity and stability conditions of the whole learning procedure. This leads to the problems, which are discussed in Section I. To avoid these problems by preserving the plasticity and stability capabilities for all the considered tasks, the proposed PLASTA-MTL approach is characterized by two novel loss functions and a novel optimization algorithm. By the proposed PPL function, the PLASTA-MTL approach minimizes the gradient magnitude differences between global image representation space and task-specific embedding spaces for the sensitivity of the global image representation space to new information learned via each task. By the proposed SPL function, the PLASTA-MTL approach minimizes angular distances between task gradients over global image representation space to protect it from radical disruptions by each task. To accurately apply these loss functions, the proposed optimization algorithm sequentially optimizes task-specific objectives together with the PPL function. In our algorithm, the SPL function is optimized at the end of the task sequence for all the tasks. In the following, we initially explain in detail the proposed PPL and SPL functions and then introduce the proposed sequential optimization algorithm.

A. Plasticity Preservation by the Proposed PPL Function

The proposed PLASTA-MTL approach aims to control the level of plasticity for each task in the context of MTL and, thus, to ensure the sensitivity to new information learned via each task. The level of plasticity for each task is controlled by what extent information encoded in task-specific embedding space is also encoded in the global image representation space. To this end, we define the plasticity condition for the ith task T_i as how much change is occurred in G compared to that of E_{T_i}, while learning T_i is based on the corresponding loss function $\mathcal{L}_{T_i}$. To measure the change occurred in G and E_{T_i} for T_i, we utilize the gradients of $\mathcal{L}_{T_i}$ with respect to the global image representation and task-specific embedding parameters [$\nabla_G \mathcal{L}_{T_i}(\theta)$ and $\nabla_{E_{T_i}} \mathcal{L}_{T_i}(\theta)$]. Then, the gradient magnitude difference between global image representation space G and task-specific embedding space E_{T_i} for task T_i represents the change occurred in G and E_{T_i} as follows: $\|\nabla_G \mathcal{L}_{T_i}\| - \|\nabla_{E_{T_i}} \mathcal{L}_{T_i}\|$. When this difference increases throughout the learning procedure, information specific to task T_i is only encoded by task-specific embedding space. Then, the considered DNN suffers from the lack of plasticity condition for global image representation space. Accordingly, to minimize the degree of changes in global image representation space G and the task-specific embedding space E_{T_i}, we define the PPL function $\mathcal{L}_{\text{PPL}}^{T_i}$ for the task T_i as follows:

$$\mathcal{L}_{\text{PPL}}^{T_i} = \left| \frac{\|\nabla_G \mathcal{L}_{T_i}(\theta)\|}{\dim(\nabla_G \mathcal{L}_{T_i}(\theta))} - \frac{\|\nabla_{E_{T_i}} \mathcal{L}_{T_i}(\theta)\|}{\dim(\nabla_{E_{T_i}} \mathcal{L}_{T_i}(\theta))} \right| \qquad (1)$$

where dim function gives the dimensions of the gradient vectors that are used to normalize the gradient magnitude difference. Since each task is associated with a separate set of task-specific embedding parameters, PPL is defined for each task. In detail, we define the PPL objective based on the gradients of a task-specific loss function. It is worth noting that defining loss functions based on the task-specific gradients is often considered in the framework of MTL (e.g., [42], [52], [54], and [55]) to control the effect of each task on the weight update of a DNN [39].

Due to our PPL function, the proposed PLASTA-MTL approach keeps the gradient magnitudes of G and E_{T_i} on the same scale while modeling the task T_i. This leads the task-specific information to be characterized in both global image representation space and task-specific embedding space. Thus, the global image representation space (which is shared by all the tasks) is enforced to be sensitive to new information learned with each task during training. Accordingly, the proposed PLASTA-MTL approach prevents the considered DNN from the lack of plasticity condition for each considered task. It is worth noting that, when a joint optimization algorithm is employed on the aggregation of all task-specific loss functions, the application of our PPL function for all tasks can increase the complexity of the whole learning procedure. In this case, the gradient magnitude of G is forced to simultaneously have the same scale with that of E_{T_i} for each $i \in \{1, \ldots, N\}$ that can exacerbate confusion for the whole learning procedure.

B. Stability Preservation by the Proposed SPL Function

The proposed PLASTA-MTL approach aims to adjust the level of stability in between consecutive tasks in the context of MTL and, thus, to prevent the whole learning procedure from radical disruptions while learning multiple tasks. The level of stability in between learning different tasks is characterized by the degree of change (which is occurred in global image representation space) due to a new task with respect to that of previous tasks. Accordingly, the level of stability condition for all the tasks $\{T_1, \ldots, T_N\}$ can be defined as how much change is occurred in G in-between learning consecutive tasks based on their corresponding loss functions $\{\mathcal{L}_{T_1}, \ldots, \mathcal{L}_{T_N}\}$. To this end, we define the relative change in G between learning two consecutive tasks T_i and T_{i+1} as the angular distance between the gradients of the associated loss functions $\nabla_G \mathcal{L}_{T_i}(\theta)$ and $\nabla_G \mathcal{L}_{T_{i+1}}(\theta)$. If this angular distance between two gradient vectors (that is associated with two consecutive tasks) becomes extremely high throughout the learning procedure, the gradient of the latter task enforces global image representation to change into a very different direction compared to the former task. In this way, the latter task radically changes the global image representation space. This may lead to a lack of stability for the considered learning. Accordingly, to minimize the angular distances, each of which is between the gradients of each consecutive task, we define the SPL function as follows:

$$\mathcal{L}_{\text{SPL}} = \frac{1}{N-1} \sum_{i=1}^{N-1} \arccos\left(\frac{\nabla_G \mathcal{L}_{T_i}(\theta) \cdot \nabla_G \mathcal{L}_{T_{i+1}}(\theta)}{\|\nabla_G \mathcal{L}_{T_i}(\theta)\| \|\nabla_G \mathcal{L}_{T_{i+1}}(\theta)\|} \right) \quad (2)$$

where $\arccos((a \cdot b)/(||a||||b||))$ measures the angle between the vectors a and b. To ensure the stability condition for all the tasks $\{T_1, \ldots, T_N\}$, the proposed SPL function considers the angular distances between all consecutive pairs in the task sequence.

Due to our SPL function, the proposed PLASTA-MTL approach keeps the angular distances between different task gradients minimum while learning all the tasks $\{T_1, \ldots, T_N\}$. Thus, the directions of task gradients over global image representation space are forced to be stable throughout the whole learning procedure. This prevents radical changes in global image representation space due to learning any task. Accordingly, the proposed PLASTA-MTL approach prevents the considered DNN from the lack of stability condition for all the tasks. We would like to point out that, if the conventional optimization algorithm of MTL is applied, the optimization of all loss functions is applied simultaneously. In this way, there is a single change in G based on the gradient of aggregated loss functions of all tasks. Then, it is hard to model relative changes in G with respect to different tasks.

C. Proposed Sequential Optimization Algorithm

For the whole learning procedure, the proposed sequential optimization algorithm aims to adaptively adjust the interactions between task-specific learning procedures and, thus, allows the proposed PLASTA-MTL approach to ensure plasticity and stability conditions for all the tasks. As in most of the DL-based MTL methods, learning the parameters of the considered DNN for the tasks $\{T_i\}_{i=1}^N$ can be achieved based on the following empirical risk minimization formulation:

$$\min_{\theta} \sum_{i=1}^{N} \lambda_i \mathcal{L}_{T_i}(\theta) \quad (3)$$

where λ_i is the weight parameter of the task T_i. In this formulation, for a given minibatch of training images, there is one optimization procedure, where all the model parameters are jointly updated to minimize the aggregation of all loss functions. This formulation limits to control plasticity and stability conditions for each task, as explained in Sections I and II. Unlike the existing MTL methods, in the proposed sequential optimization algorithm, there is one optimization procedure for each task-specific loss function together with the corresponding PPL function. At the end of the task sequence, this algorithm applies one more additional optimization procedure for SPL by considering all the tasks. To this end, we first formulate (3) as a multilevel optimization problem as follows:

$$\begin{aligned} &\min_{G, \theta^{T_N}} \ \mathcal{L}_{T_N}(G, \theta^{T_N}) \\ &\text{s.t. } G \in \arg\min_{G, \theta^{T_{N-1}}} \ \mathcal{L}_{T_{N-1}}(G, \theta^{T_{N-1}}) \\ &\quad \ldots \\ &\quad \text{s.t. } G \in \arg\min_{G, \theta^{T_1}} \mathcal{L}_{T_1}(G, \theta^{T_1}) \end{aligned} \quad (4)$$

where $\theta^{T_i} \in \theta$ is the set of task-specific parameters associated with the task T_i (i.e., task-specific head parameters). The reader is referred to [56] for the details of multilevel optimization formulation. For (4), the set of all tasks $\mathcal{T}$ is regarded as a sequence $\langle T_i \mid i \in \{1, \ldots, N\}\rangle$. Accordingly, instead of jointly optimizing all the tasks, every task T_i in the sequence is optimized sequentially. In this way, global image representation space (which is defined by G) is always affected by the optimization of last task in the sequence. This allows to adaptively adjust the interactions between task-specific learning procedures and, thus, to integrate the plasticity and stability preserving capabilities of the proposed PLASTA-MTL approach into the whole learning procedure. To this end, for each task, we minimize the corresponding PPL function $\mathcal{L}^{T_i}_{\text{PPL}}$ with the task-specific loss function $\mathcal{L}_{T_i}$ by integrating multiobjective optimization of two loss functions to (4) as follows:

$$\begin{aligned}&\min_{G,\theta^{T_N}} \left(\mathcal{L}_{T_N}(G, \theta^{T_N}), \mathcal{L}^{T_N}_{\text{PPL}}(\nabla_G \mathcal{L}_{T_N}, \nabla_{E_{T_N}} \mathcal{L}_{T_N})\right)\\ &\text{s.t. } G \in \operatorname*{arg\,min}_{G,\theta^{T_{N-1}}} \left(\mathcal{L}_{T_{N-1}}, \mathcal{L}^{T_{N-1}}_{\text{PPL}}\right)\\ &\quad\cdots\\ &\qquad \text{s.t. } G \in \operatorname*{arg\,min}_{G,\theta^{T_1}} \left(\mathcal{L}_{T_1}, \mathcal{L}^{T_1}_{\text{PPL}}\right). \end{aligned} \tag{5}$$

It is worth noting that, during the optimization of $\mathcal{L}^{T_i}_{\text{PPL}}$, $\nabla_{E_{T_i}} \mathcal{L}_{T_i}$ is regarded as constant. Due to this, global image representation space (which is defined by G) is affected by the optimization of the last task in the sequence with the corresponding PPL function. Since the SPL function $\mathcal{L}_{\text{SPL}}$ is applied for all the tasks, it is optimized at the end of the sequence as follows:

$$\begin{aligned}&\min_{G} \ \mathcal{L}_{\text{SPL}}\left(\{\nabla_G \mathcal{L}_{T_i}\}_{i=1}^{N}\right)\\ &\text{s.t. } G \in \operatorname*{arg\,min}_{G,\theta^{T_N}} \left(\mathcal{L}_{T_N}, \mathcal{L}^{T_N}_{\text{PPL}}\right)\\ &\quad \text{s.t. } G \in \operatorname*{arg\,min}_{G,\theta^{T_{N-1}}} \left(\mathcal{L}_{T_{N-1}}, \mathcal{L}^{T_{N-1}}_{\text{PPL}}\right)\\ &\qquad\cdots\\ &\qquad\quad \text{s.t. } G \in \operatorname*{arg\,min}_{G,\theta^{T_1}} \left(\mathcal{L}_{T_1}, \mathcal{L}^{T_1}_{\text{PPL}}\right) \end{aligned} \tag{6}$$

where $\nabla_G \mathcal{L}_{T_i}$ is stored in each minimization step to be utilized for the optimization of $\mathcal{L}_{\text{SPL}}$.

Algorithm 1 Proposed Sequential Optimization Algorithm to Train the Proposed PLASTA-MTL Approach

```
Require: Mini-batch $\mathcal{B} \in \mathcal{X}$, set of tasks $\mathcal{T}=\{T_1, \ldots, T_N\}$,
    set of model parameters $\theta$, $\alpha$, $\beta$
for $i \leftarrow 1$ to N do
    Compute $\mathcal{L}_{T_i}(\theta)$
    Compute $\nabla_\theta \mathcal{L}_{T_i}(\theta)$, $\nabla_G \mathcal{L}_{T_i}(\theta)$ and $\nabla_{E_{T_i}} \mathcal{L}_{T_i}(\theta)$
    Compute $\mathcal{L}^{T_i}_{PPL} = |\frac{\|\nabla_G \mathcal{L}_{T_i}(\theta)\|}{dim(\nabla_G \mathcal{L}_{T_i}(\theta))} - \frac{\|\nabla_{E_{T_i}} \mathcal{L}_{T_i}(\theta)\|}{dim(\nabla_{E_{T_i}} \mathcal{L}_{T_i}(\theta))}|$
    Compute $\nabla_G \mathcal{L}^{T_i}_{PPL}$
    Update $\theta$ using $\nabla_\theta \mathcal{L}_{T_i}(\theta)$
    if $\|\nabla_G \mathcal{L}_{SPL}\| < \alpha$ then
        Update $G$ using $\nabla_G \mathcal{L}^{T_i}_{PPL}$
    end if
end for
Compute $\mathcal{L}_{SPL} = \frac{1}{N-1} \sum_{i=1}^{N-1} \arccos(\frac{\nabla_G \mathcal{L}_{T_i}(\theta) \cdot \nabla_G \mathcal{L}_{T_{i+1}}(\theta)}{\|\nabla_G \mathcal{L}_{T_i}(\theta)\| \|\nabla_G \mathcal{L}_{T_{i+1}}(\theta)\|})$
Compute $\nabla_G \mathcal{L}_{SPL}$
if $\|\nabla_G \mathcal{L}_{SPL}\| > \beta$ then
    Update $G$ using $\nabla_G \mathcal{L}_{SPL}$
end if
```

It is worth noting that, depending on the selection of tasks, the assurance of the stability condition for the considered DNN may decrease the level of plasticity condition and vice versa. In this way, the lack of one of the stability and plasticity conditions is associated with the excess of the other condition. As an example, if some of the considered tasks are in heavy competition during training and one of the tasks can distract the other tasks, there is a lack of stability condition. This is also due to the excess plasticity condition. In this way, increasing the level of stability condition results in a decrease in the plasticity condition that leads to the lack of stability condition. Under such conditions, the stability-plasticity constraint of a DNN is defined as a dilemma between these two capabilities of the DNN. If there is this dilemma, it can be misleading to address both stability and plasticity capabilities at the same time. This may lead to the ineffective characterization of one of the conditions. The drawback of this can be more evident if preserving one of the capabilities is more important than the other one. Accordingly, in the proposed PLASTA-MTL approach, we aim to automatically detect which capability should be preserved if there is a need for selecting only one of them. To this end, we define the importance level of stability condition for the considered DNN and the tasks based on the L^2-norm of the gradient of SPL. Accordingly, for a given set of tasks, we define the set of all the loss functions to be considered based on the two different levels of importance for L_{SPL} as follows:

$$\mathfrak{L} : \begin{cases} \mathcal{L}_{T_1} \ldots \mathcal{L}_{T_N}, \mathcal{L}_{\text{SPL}}, & \text{if } \|\nabla_G \mathcal{L}_{\text{SPL}}\| \geq \alpha \\ \mathcal{L}_{T_1}, \mathcal{L}^{T_1}_{\text{PPL}} \ldots \mathcal{L}_{T_N}, \mathcal{L}^{T_N}_{\text{PPL}}, & \text{if } \|\nabla_G \mathcal{L}_{\text{SPL}}\| \leq \beta \\ \mathcal{L}_{T_1}, \mathcal{L}^{T_1}_{\text{PPL}} \ldots \mathcal{L}_{T_N}, \mathcal{L}^{T_N}_{\text{PPL}}, \mathcal{L}_{\text{SPL}}, & \text{otherwise} \end{cases} \tag{7}$$

where α and β control the importance limits, while $\alpha > \beta$. If L^2-norm of the gradient $\nabla_G \mathcal{L}_{\text{SPL}}$ is significantly high (higher than α), we assume that there is no need to apply $\mathcal{L}_{\text{PPL}}$. This applies to $\mathcal{L}_{\text{SPL}}$ if L^2-norm of the gradient $\nabla_G \mathcal{L}_{\text{SPL}}$ is significantly low (lower than β). If the L^2-norm is in between α and β, we define this interval as the condition where stability-plasticity constraint is not a dilemma anymore, and thus, both of the capabilities can be preserved in the proposed PLASTA-MTL approach. It is worth noting that, since $\nabla_G \mathcal{L}_{\text{SPL}}$ depends on the normalized gradients of consecutive task-specific loss functions [see (2)], it is mostly affected by which tasks are jointly considered. However, it is less affected by the considered dataset since the input samples indirectly changes the gradient of the SPL function. The proposed sequential algorithm automatically decides to apply PPL, SPL, or both loss functions together depending on the parameters α and β. Accordingly, (5) is used to apply only PPL function, (6) is used without $\mathcal{L}^{T_i}_{\text{PPL}}$ to apply SPL function, and (6) is used to apply both loss functions together. In practice, this decision can be made at the end of the first epoch of the training based on the parameters of α and β. The proposed sequential optimization algorithm is summarized in Algorithm 1. To better understand the applied operations in it,

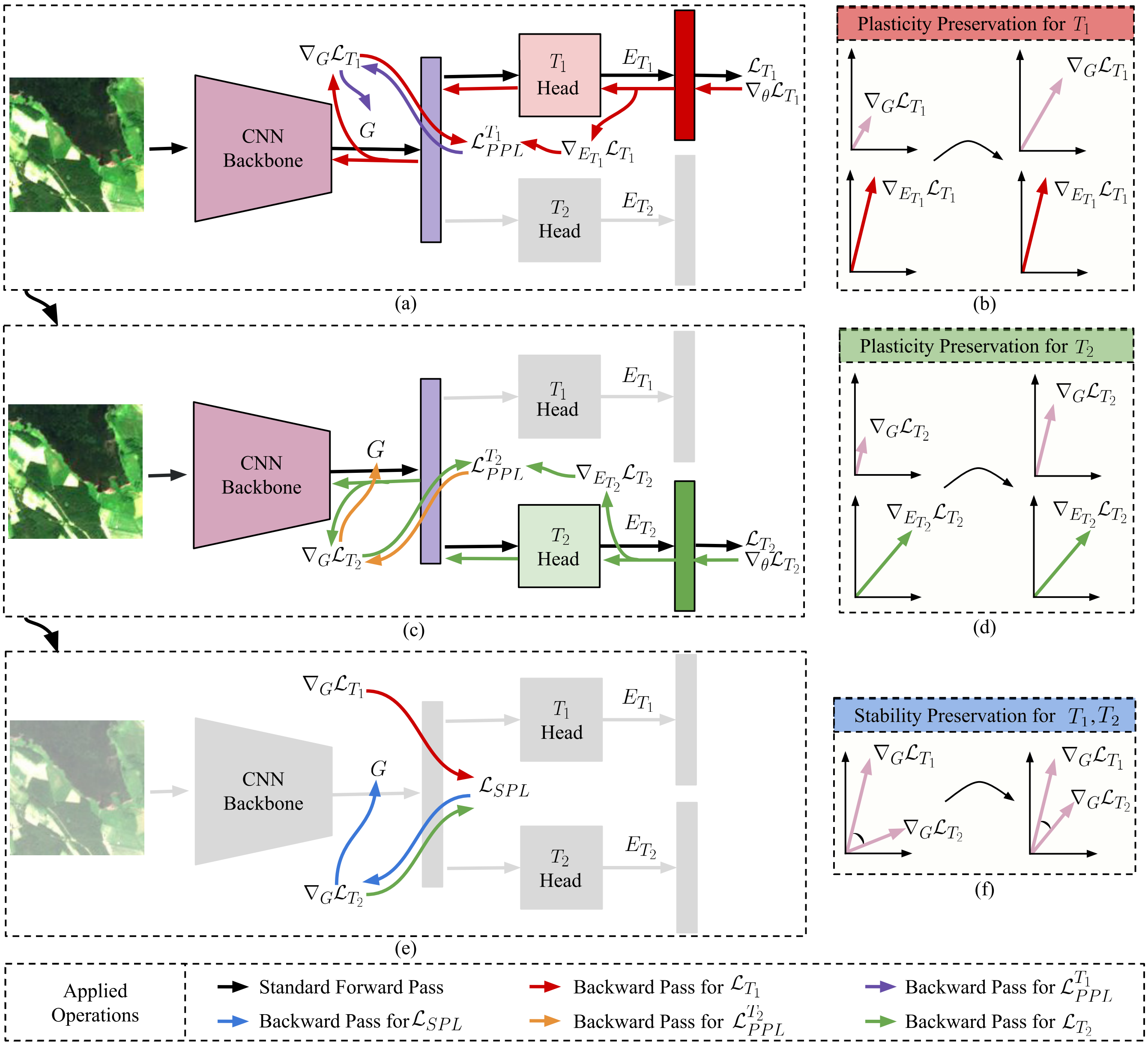

$\nabla_G \mathcal{L}_{T_1}$
CNN Backbone
G
T_1 Head
E_{T_1}
$\mathcal{L}_{T_1}$
$\nabla_\theta \mathcal{L}_{T_1}$
$\mathcal{L}^{T_1}_{PPL}$
$\nabla_{E_{T_1}} \mathcal{L}_{T_1}$
T_2 Head
E_{T_2}
(a)
Plasticity Preservation for T_1
$\nabla_G \mathcal{L}_{T_1}$
$\nabla_{E_{T_1}} \mathcal{L}_{T_1}$
(b)
CNN Backbone
G
T_1 Head
E_{T_1}
$\mathcal{L}^{T_2}_{PPL}$
$\nabla_{E_{T_2}} \mathcal{L}_{T_2}$
$\nabla_G \mathcal{L}_{T_2}$
T_2 Head
E_{T_2}
$\mathcal{L}_{T_2}$
$\nabla_\theta \mathcal{L}_{T_2}$
(c)
Plasticity Preservation for T_2
$\nabla_G \mathcal{L}_{T_2}$
$\nabla_{E_{T_2}} \mathcal{L}_{T_2}$
(d)
$\nabla_G \mathcal{L}_{T_1}$
CNN Backbone
G
$\mathcal{L}_{SPL}$
T_1 Head
E_{T_1}
T_2 Head
E_{T_2}
$\nabla_G \mathcal{L}_{T_2}$
(e)
Stability Preservation for T_1, T_2
$\nabla_G \mathcal{L}_{T_1}$
$\nabla_G \mathcal{L}_{T_2}$
(f)
Applied Operations
Standard Forward Pass
Backward Pass for $\mathcal{L}_{SPL}$
Backward Pass for $\mathcal{L}_{T_1}$
Backward Pass for $\mathcal{L}^{T_2}_{PPL}$
Backward Pass for $\mathcal{L}^{T_1}_{PPL}$
Backward Pass for $\mathcal{L}_{T_2}$

Fig. 1. Illustration of the proposed PLASTA-MTL approach training when two tasks T_1 and T_2 are considered. Standard and plasticity preservation backward passes for (a) T_1 and (c) T_2 are shown, while the changes over the gradient vectors (b) $\nabla_G \mathcal{L}_{T_1}$ and (d) $\nabla_G \mathcal{L}_{T_2}$ during the plasticity preservation of these tasks are visualized. (e) Backward pass for stability preservation of all the tasks are given with (f) illustration of changes over their gradient vectors.

Fig. 1 shows an illustration of the proposed PLASTA-MTL approach training with the proposed optimization algorithm. It is noted that, for simplicity, forward and backward passes applied in our optimization algorithm are visualized for two tasks. For the first task, while $\nabla_\theta \mathcal{L}_{T_1}$ is propagated back [which is visualized with red arrows in Fig. 1(a)], $\mathcal{L}^{T_1}_{\text{PPL}}$ is calculated. Then, backward pass for $\mathcal{L}^{T_1}_{\text{PPL}}$ is applied [which is illustrated with purple arrows in Fig. 1(a)]. During the plasticity preservation for the first task, the change over the gradient vector $\nabla_G \mathcal{L}_{T_1}$ is visualized in Fig. 1(b). The same steps are also presented for the second task in Fig. 1(c) and (d). After the plasticity preservation is employed for both tasks, $\mathcal{L}_{\text{SPL}}$ is calculated [see Fig. 1(e)]. At the end, the backward pass for the SPL function is applied (which is visualized with blue arrows). During the stability preservation for both tasks, the changes over the gradient vectors of both tasks are presented in Fig. 1(f).

Since the proposed algorithm allows applying a task-specific optimization procedure for each task unlike the joint optimization algorithm, the PLASTA-MTL approach is capable of effectively preserving plasticity and stability capabilities for each task in the context of MTL. We would like to point out that this algorithm does not require the selection of any loss function weights (which generally requires a computationally demanding grid search in the joint optimization algorithm). It is also worth noting that the proposed algorithm works independently of the number of considered tasks and the type of tasks.

IV. Dataset Description and Experimental Setup

A. Dataset Description

The experiments were performed on the DLRSD [57] and the BigEarthNet-S2 [58] benchmark archives. The DLRSD archive includes the same images as the UC Merced archive [59] that consists of 2100 aerial images, each of which has the size of 256 × 256 pixels with a spatial resolution of 30 cm. In the DLRSD archive, the images are associated with the multilabels and the pixel-based labels, where the set of class labels is defined in [4]. We utilized the Serbia subset of the BigEarthNet-S2 benchmark archive, where images are acquired during the summer season. This subset includes 14 832 Sentinel-2 images, each of which is a section of: 1) 120 × 120 pixels for 10-m bands; 2) 60 × 60 pixels for 20-m bands; and 3) 20 × 20 pixels for 60-m bands. For the experiments, we applied bicubic interpolation to 20-m bands and excluded 60-m bands. Each image is annotated with multiple classes provided by the CORINE Land Cover Map (CLC) database of the year 2018. For the experiments, we utilized the 19 class nomenclature presented in [58]. For the tasks that require the availability of land-cover maps, we extracted the CLC land cover map of each image.

To perform experiments, we divided the DLRSD and the BigEarthNet-S2 archives into training, validation, and test sets with the ratios of 70%, 10%, and 20% and 52%, 24%, and 24%, respectively. To apply CBIR, the training set of the DLRSD archive and the validation set of the BigEarthNet-S2 archive were used for selecting query images, while images were retrieved from the test set for both archives.

B. Experimental Setup

In the experiments, we utilized the DenseNet-121 CNN architecture [60] as the MTL backbone shared by all the tasks. To perform the experiments, we utilized the different combinations of four tasks: 1) supervised scene classification; 2) supervised similarity learning; 3) supervised multilabel co-occurrence prediction; and 4) unsupervised similarity learning. For each task, we added a task-specific head to the CNN backbone. Each task head includes a fully connected (FC) layer that: 1) takes the global image representation from the CNN backbone and 2) produces a 64-D task-specific embedding. Supervised scene classification task (which is denoted as T_1) aims to automatically assign multilabels to image scenes. To this end, the task head of T_1 also includes a classification layer that produces multilabel class probabilities. For this task, the task-specific loss function $\mathcal{L}_{T_1}$ is selected as a cross-entropy loss function. For the details of this task, the reader is referred to [61]. The supervised similarity learning task (which is denoted as T_2) aims to automatically identify image similarities. To this end, we selected a triplet loss function as the task-specific loss function $\mathcal{L}_{T_2}$. The triplet loss function directly operates on the task-specific embeddings and requires the availability of image triplets (each of which includes anchor, positive, and negative images). For this task, image triplets are selected by using the hard triplet sampling technique based on the multilabel similarities. The reader is referred to [62] and [63] for the details of the triplet loss function and the triplet sampling techniques. The supervised multilabel co-occurrence prediction task (which is denoted as T_3) aims to predict co-occurrence relationships of multiple classes present in an image. To this end, by following the method presented in [33], the task head of T_3 also includes an FC layer that takes task-specific embeddings and produces the prediction for graph-driven region-based image representations. For this task, the region representation learning loss function [33] is selected as the task-specific loss function $\mathcal{L}_{T_3}$. It minimizes the prediction error of the task-specific head with comparison to the image graphs, which are obtained based on the image land-cover maps. The unsupervised similarity learning task (which is denoted as T_4) aims at learning image representations by maximizing the similarity between different views of the same image without relying on any ground-truth information. To this end, by following the strategy of self-supervised contrastive learning presented in [64], we used a set of data augmentation techniques to generate different views of each training image. Then, the task-specific loss function $\mathcal{L}_{T_4}$ is selected as a contrastive loss function, which operates on the task-specific embeddings of two different augmented views of each image. It allows maximizing the similarity between the augmented views of images with respect to the rest of the images. The reader is referred to [64] for the details of contrastive loss and the set of data augmentation techniques, which is applied to generate different views of images.

We trained the proposed PLASTA-MTL approach for 100 epochs. For training, we utilized the Adam variant of stochastic gradient descent with the initial learning rate of 10^{-3}. All the experiments were performed on four NVIDIA Tesla V100 GPUs. After training is finished by employing the abovementioned tasks in the context of MTL, we extracted the features of query and archive images from the last layer of the CNN backbone. To apply CBIR, we applied similarity matching of the extracted image features based on the χ^2-distance measure. CBIR results are provided in terms of two evaluation metrics: 1) normalized discounted cumulative gains (NDCGs) and 2) mean average precision (mAP). For the details of these metrics, the reader is referred to [65].

We carried out various experiments to: 1) perform a sensitivity analysis of the proposed PLASTA-MTL approach and 2) compare our approach with state-of-the-art MTL methods in the context of CBIR. For the sensitivity analysis, we assessed: 1) the effectiveness of the selection of plasticity and stability preserving capabilities; 2) the effect of task sequence order on the proposed sequential optimization algorithm; 3) computational complexity of the PLASTA-MTL approach; and 4) the comparison of utilizing multiple tasks in our approach with separately employing each task (that is based on STL).

We compared the proposed approach with: 1) conventional MTL (equal weighting); 2) MTL using uncertainty to weigh losses (uncertainty weighting) [43]; 3) projecting conflicting gradients (PCGrads) [51]; 4) gradient normalization for adaptive loss balancing in deep multi-task networks (GradNorm) [52]; and 5) dynamic weight average (DWA) [53]. For all the methods, we used the same CNN backbone and task-specific heads with our approach. For the first method, we applied joint optimization on the summation of task-specific loss functions with equal weights. For the other four methods, we used the same method-specific parameters given in [43] and [51]–[53].

V. Experimental Results

We performed different kinds of experiments in order to: 1) carry out sensitivity analysis and 2) compare the effectiveness of the proposed PLASTA-MTL approach with the state-of-the-art MTL methods in the framework of CBIR.

A. Sensitivity Analysis of the Proposed Approach

In this section, we performed the sensitivity analysis of the proposed PLASTA-MTL approach in terms of: 1) the effectiveness of the automatic selection of plasticity and stability preserving capabilities; 2) the task sequence order utilized in our approach; 3) the computational complexity; and 4) the comparison with STL.

In the first set of trials, we analyzed the effectiveness of automatically detecting the preservation of plasticity and stability capabilities in the proposed PLASTA-MTL approach. Table I shows the mAP scores for the DLRSD archive when different combinations of the tasks $\{T_1, T_2, T_3, T_4\}$ are utilized with the different combinations of plasticity and stability preserving capabilities in the PLASTA-MTL approach. By assessing the table, one can observe that the selection of which capabilities are preserved in our PLASTA-MTL approach is one of the most important factors affecting the overall CBIR performance. This issue becomes more evident under two scenarios. First, if some of the considered tasks are in competition during training, the preservation of both capabilities at the same time leads to the ineffective characterization of either stability or plasticity conditions. This results in lower mAP scores compared to preserving only one of the capabilities. As an example, when the considered tasks include T_1 and T_2, employing only either PPL or SPL leads to 1.7% and 1% higher mAP scores, respectively, compared to utilizing both loss functions together in the proposed PLASTA-MTL approach. This is due to the fact that learning the task T_1 (which is supervised scene classification) enforces to maximize interclass distances in the global image representation space, while learning the task T_2 (which is supervised similarity learning) enforces to minimize intraclass distance. These learning characteristics can easily result in the competition between the two tasks. However, when the considered tasks include T_2 and T_3 (which are not in competition during training), preserving each capability further improves the CBIR performance. Second, when the number of considered tasks decreases, the effect of selecting one of the plasticity and stability preserving capabilities on mAP scores increases. As an example, when the considered tasks include only T_1 and T_4, the difference of mAP scores between preserving plasticity and stability capabilities is more than 4%. However, when all the tasks are considered, including T_1, T_2, T_3, and T_4, this difference is less than 1%. These two scenarios show that the accurate selection of which capabilities are preserved in our PLASTA-MTL approach is crucial for accurate CBIR performance. The proposed sequential optimization strategy automatically detects which capabilities are preserved by controlling the importance level of stability condition, which is defined based on the L^2-norm of the gradient of the SPL function. Table I also includes the average gradient norm values that are obtained in the first epoch of the training. By the analyzing the table, one can observe that, when the norm value is significantly high (e.g., $\mathcal{T} = \{T_1, T_2\}$ and $\mathcal{T} = \{T_1, T_3\}$), preserving only stability capability in the PLASTA-MTL approach provides the highest mAP scores. When the norm value is significantly low (e.g., $\mathcal{T} = \{T_1, T_2, T_4\}$ and $\mathcal{T} = \{T_1, T_3, T_4\}$), preserving only plasticity capability in the PLASTA-MTL approach provides the highest mAP scores. This shows the effectiveness of the automatic detection strategy of the proposed sequential optimization algorithm, which is utilized to identify which capabilities are preserved in our PLASTA-MTL approach. The average gradient norm values given in Table I show that two importance levels of stability condition can be defined as $\alpha = 0.3$ and $\beta = 0.1$. Accordingly, we used these parameters in the proposed sequential optimization algorithm for the rest of the experiments.

TABLE I

mAP Scores Associated With the Different Combinations of Tasks With Different Capabilities of the PLASTA-MTL Approach Are Utilized (the DLRSD Archive)

Tasks				**PLASTA-MTL**			
T_1	T_2	T_3	T_4	$\mathcal{L}_{PPL}$	$\mathcal{L}_{SPL}$	$\|\nabla_G \mathcal{L}_{SPL}\|$	**mAP (%)**
✓	✓	✗	✗	✓	✗	0.33	95.0
				✗	✓		95.7
				✓	✓		94.0
✓	✗	✓	✗	✓	✗	0.43	96.0
				✗	✓		97.6
				✓	✓		96.7
✓	✗	✗	✓	✓	✗	0.13	95.2
				✗	✓		91.0
				✓	✓		96.0
✗	✓	✓	✗	✓	✗	0.18	94.8
				✗	✓		95.2
				✓	✓		95.5
✗	✓	✗	✓	✓	✗	0.09	86.1
				✗	✓		84.5
				✓	✓		85.4
✗	✗	✓	✓	✓	✗	0.09	95.4
				✗	✓		94.8
				✓	✓		93.8
✓	✓	✓	✗	✓	✗	0.12	96.5
				✗	✓		96.3
				✓	✓		97.2
✓	✓	✗	✓	✓	✗	0.04	96.7
				✗	✓		94.7
				✓	✓		94.8
✓	✗	✓	✓	✓	✗	0.05	97.0
				✗	✓		94.5
				✓	✓		96.8
✗	✓	✓	✓	✓	✗	0.06	95.5
				✗	✓		93.4
				✓	✓		95.2
✓	✓	✓	✓	✓	✗	0.13	97.5
				✗	✓		97.0
				✓	✓		97.6

TABLE II

mAP Scores When the Tasks T_1, T_2, and T_3 Are Utilized in Different Orders for the PLASTA-MTL Approach (the DLRSD Archive)

Task Order	mAP (%)
$T_1 \rightarrow T_2 \rightarrow T_3$	97.2
$T_1 \rightarrow T_3 \rightarrow T_2$	97.0
$T_2 \rightarrow T_1 \rightarrow T_3$	97.1
$T_2 \rightarrow T_3 \rightarrow T_1$	96.8
$T_3 \rightarrow T_1 \rightarrow T_2$	97.7
$T_3 \rightarrow T_2 \rightarrow T_1$	97.5

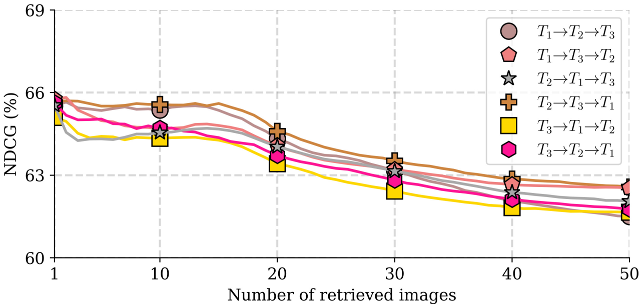


Fig. 2. NDCGs versus the number of retrieved images obtained for the DLRSD archive when the tasks T_1, T_2, and T_3 are utilized in different orders for the PLASTA-MTL approach.

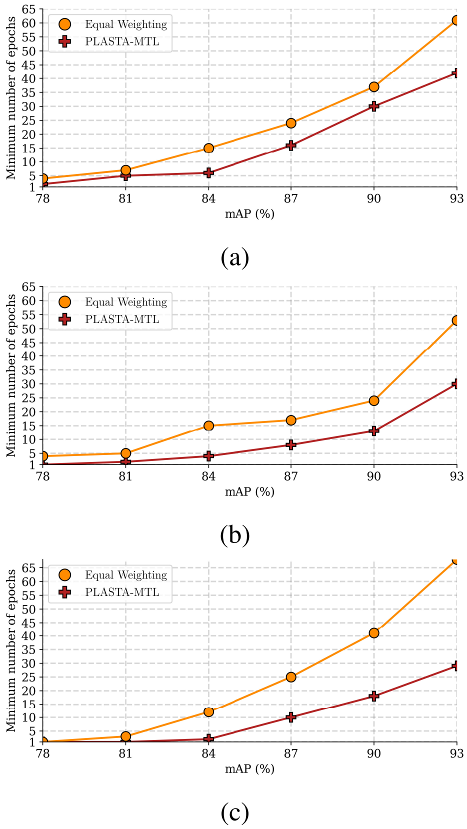
TABLE III

Training Times per Epoch on the DLRSD Archive When the Different Combinations of Tasks Are Utilized for the Proposed PLASTA-MTL Approach and Equal Weighting

Tasks				Method	Training Time per Epoch (sec)
T_1	T_2	T_3	T_4		
✓	✓	✗	✗	Equal Weighting	9.3
				PLASTA-MTL	18.0
✓	✗	✓	✗	Equal Weighting	18.3
				PLASTA-MTL	24.5
✓	✗	✗	✓	Equal Weighting	57.6
				PLASTA-MTL	62.9
✗	✓	✓	✗	Equal Weighting	17.7
				PLASTA-MTL	24.6
✗	✓	✗	✓	Equal Weighting	60.2
				PLASTA-MTL	64.9
✗	✗	✓	✓	Equal Weighting	66.4
				PLASTA-MTL	70.9
✓	✓	✓	✗	Equal Weighting	15.7
				PLASTA-MTL	32.6
✓	✓	✗	✓	Equal Weighting	57.9
				PLASTA-MTL	73.6
✓	✗	✓	✓	Equal Weighting	69.1
				PLASTA-MTL	77.7
✗	✓	✓	✓	Equal Weighting	67.8
				PLASTA-MTL	77.4
✓	✓	✓	✓	Equal Weighting	64.5
				PLASTA-MTL	88.4

In the second set of trials, we analyzed the effect of the task sequence order utilized in the proposed PLASTA-MTL approach. Table II shows the mAP scores for the DLRSD archive when the tasks $\{T_1, T_2, T_3\}$ are utilized with all the possible orders in the task sequence of our approach. By analyzing the table, one can see that, when the order of the considered tasks is changed, the proposed PLASTA-MTL approach provides different mAP scores. This is due to the fact that, since all the tasks are learned sequentially in the proposed optimization algorithm, different task sequence orders lead to changes in the whole learning procedure. However, from the table, one can also observe that the differences between the mAP scores of different task orders are not significantly high. The difference between the highest mAP score (which is obtained with the task order of $T_3 \rightarrow T_1 \rightarrow T_2$) and the lowest mAP score (which is obtained with the task order of $T_2 \rightarrow T_3 \rightarrow T_1$) is less than 1%. Fig. 2 shows the NDCG scores of the same tasks, and their orders for the DLRSD archive under different numbers of retrieved images. From the figure, one can see that increasing the number of retrieved images does not change our conclusion. These results show that utilizing different task orders does not significantly affect the CBIR performance of the proposed PLASTA-MTL approach. For the rest of the experiments, we employed the numerical order of tasks (i.e., $T_1 \rightarrow T_2 \rightarrow T_3 \rightarrow T_4$) for the proposed PLASTA-MTL approach.

In the third set of trials, we assessed the computational complexity of the proposed PLASTA-MTL approach. To this end, in Table III, we compared our approach with the equal weighting method in terms of the training time required per epoch when the different combinations of the tasks $\{T_1, T_2, T_3, T_4\}$ are utilized on the DLRSD archive. It is worth noting that the equal weighting method jointly optimizes all the loss functions without the need for any other steps that may increase the computational complexity. Accordingly, this method can be regarded as one of the MTL methods, which are associated with the lowest computational complexity. By assessing the table, one can observe that our approach requires higher training time per epoch compared to the equal weighting method for each task combination. This is due to the fact that the sequential optimization applied in the proposed PLASTA-MTL approach requires a higher number of forward and backward passes of the considered DNN compared to the joint optimization algorithm. This increases the required training time per epoch for our approach. This becomes more evident if the same batches of training images are used for all the tasks (e.g., $\mathcal{T} = \{T_1, T_2\}$). In this condition, the equal weighting method requires one forward pass and one backward pass for each batch, while our approach requires at least two forward and backward passes depending on the number of tasks. When some of the considered tasks require different batches of training images that lead to more than one forward pass, the computational complexity of the equal weighting method increases. However, it does not affect the computational complexity of our proposed approach. As an example, when the tasks $\{T_1, T_2\}$ are utilized, the training time per epoch of our approach is almost twice as large as that of the equal weighting method. However, when the tasks $\{T_1, T_4\}$ are utilized, the task T_4 requires to feed the augmented views of images into the considered DNN that costs an additional forward pass step. In this case, the required training time per epoch of the proposed PLASTA-MTL approach is almost the same as that of the equal weighting method. It is worth noting



(a)

(b)

(c)

Fig. 3. mAP versus the minimum number of training epochs for the DLRSD archive when the tasks: (a) T_2 and T_3; (b) T_1, T_2, and T_3; and (c) T_1, T_2, T_3, and T_4 are utilized for the PLASTA-MTL approach and the equal weighting method.

that the overall computational complexity is also affected by the total number of epochs in addition to the training time per epoch. Accordingly, Fig. 3 shows the minimum numbers of training epochs at which the proposed PLASTA-MTL approach and the equal weighting method reach a range of mAP scores when the different number of tasks are considered. By analyzing the figure, one can see that our approach is able to achieve the same mAP scores with less number of training epochs compared to the equal weighting method. As an example, when the tasks $\{T_1, T_2, T_3\}$ are considered, our approach achieves a 93% mAP score with 25 fewer training epochs compared to the equal weighting method. This leads to less total training time for our approach although the corresponding training time per epoch is higher than the equal weighting method. This issue becomes more evident when the number of considered tasks increases. As an example, when all the tasks are utilized, the total training time of our approach to reach 93% mAP score is significantly less than that of the equal weighing method. These results show that the learning efficiency of the proposed PLASTA-MTL approach is significantly higher than the equal weighting method. This leads to the reduction of total training time (which is required to reach a high CBIR performance) for the proposed PLASTA-MTL approach.

In the fourth set of trials, we analyzed the effectiveness of the proposed PLASTA-MTL approach compared to separately employing each task of the considered task set (that is based on STL). For the DLRSD archive, Table IV shows the mAP scores of the PLASTA-MTL approach for the different combinations of the tasks $\{T_1, T_2, T_3, T_4\}$ and the STL for each task. By analyzing the table, one can observe that, for each combination, our approach provides higher mAP scores compared to separately learning each task. As an example, when the tasks $\{T_1, T_2, T_4\}$ are considered, the proposed PLASTA-MTL approach provides almost 2%, 15%, and 14% higher mAP scores compared to applying separate learning procedures for T_1, T_2, and T_4, respectively. This shows that our approach effectively combines multiple tasks together, which leads to more accurate image representation learning compared to utilizing a single task.

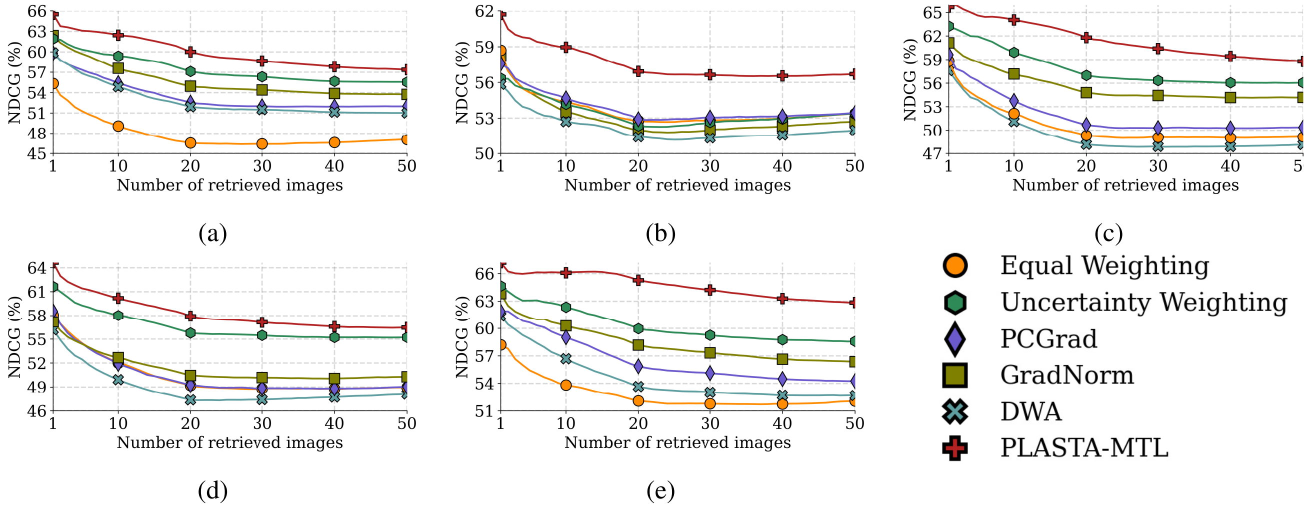
TABLE IV

mAP Scores When the Different Combinations of Tasks Are Utilized in the PLASTA-MTL Approach Compared to Single Task Learning (the DLRSD Archive)

Tasks				Method	mAP (%)
T_1	T_2	T_3	T_4		
✓	✗	✗	✗	STL	94.9
✗	✓	✗	✗		81.8
✗	✗	✓	✗		95.4
✗	✗	✗	✓		83.2
✓	✓	✗	✗	PLASTA-MTL	**95.7**
✓	✗	✓	✗		**97.6**
✓	✗	✗	✓		**96.0**
✗	✓	✓	✗		**95.5**
✗	✓	✗	✓		**86.1**
✗	✗	✓	✓		**95.4**
✓	✓	✓	✗		**97.2**
✓	✓	✗	✓		**96.7**
✓	✗	✓	✓		**97.0**
✗	✓	✓	✓		**95.5**
✓	✓	✓	✓		**97.6**

B. Comparison Among the Sate-of-the-Art MTL Methods

In the fifth set of trials, we analyzed the effectiveness of the proposed PLASTA-MTL approach compared to the state-of-the-art MTL methods in the context of CBIR under various combinations of the considered four tasks. These methods are equal weighting, uncertainty weighting [43], PCGrad [51], GradNorm [52], and DWA [53]. Tables V and VI show the corresponding mAP scores on the DLRSD and the BigEarthNet-S2 archives, respectively. By assessing the tables, one can observe that the proposed PLASTA-MTL approach leads to the highest mAP scores on each task combination for both archives compared to the state-of-the-art MTL methods. As an example, the proposed PLASTA-MTL approach outperforms the PCGrad by more than 4% for the DLRSD archive and more than 8% for the BigEarthNet-S2 archive when the tasks $\{T_2, T_3, T_4\}$ are utilized. When all the tasks $\{T_1, T_2, T_3, T_4\}$ are used, our approach provides almost 3% higher mAP scores for both archives compared to the GradNorm. We observed similar behaviors while comparing the methods of equal weighting, uncertainty weighting, and DWA with our approach. This shows that the proposed PLASTA-MTL approach provides more accurate RS image representations that lead to more effective CBIR compared to other methods. This is due to the plasticity and



(a)

(b)

(c)

(d)

(e)

Fig. 4. NDCGs versus the number of retrieved images obtained for the DLRSD archive when the tasks: (a) T_1 and T_4; (b) T_2 and T_3; (c) T_1, T_2, and T_4; (d) T_2, T_3, and T_4; and (e) T_1, T_2, T_3, and T_4 are used in the context of MTL.

TABLE V

mAP Scores Associated With the Different Combinations of Tasks (the DLRSD Archive)

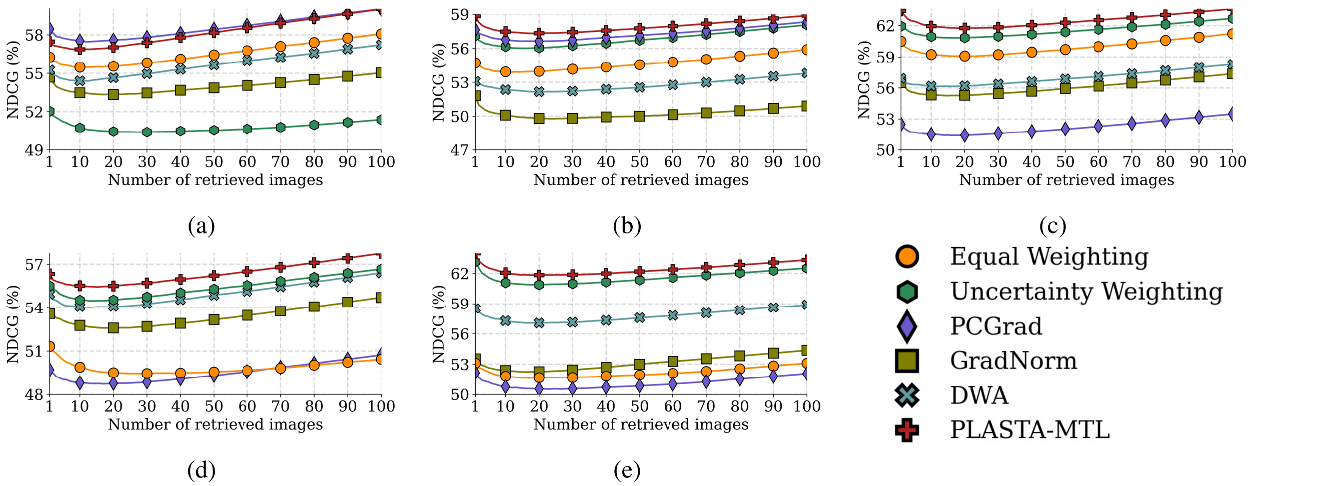
Tasks				Method	mAP (%)
T_1	T_2	T_3	T_4		
✓	✗	✗	✓	Equal Weighting	90.1
				Uncertainty Weighting [43]	94.4
				PCGrad [51]	92.7
				GradNorm [52]	94.3
				DWA [53]	93.0
				PLASTA-MTL	**96.0**
✗	✓	✓	✗	Equal Weighting	93.2
				Uncertainty Weighting [43]	94.0
				PCGrad [51]	92.6
				GradNorm [52]	92.9
				DWA [53]	92.5
				PLASTA-MTL	**95.5**
✓	✓	✗	✓	Equal Weighting	91.6
				Uncertainty Weighting [43]	95.4
				PCGrad [51]	92.9
				GradNorm [52]	93.8
				DWA [53]	91.4
				PLASTA-MTL	**96.7**
✗	✓	✓	✓	Equal Weighting	92.0
				Uncertainty Weighting [43]	95.0
				PCGrad [51]	91.2
				GradNorm [52]	91.4
				DWA [53]	90.9
				PLASTA-MTL	**95.5**
✓	✓	✓	✓	Equal Weighting	92.6
				Uncertainty Weighting [43]	95.8
				PCGrad [51]	94.9
				GradNorm [52]	95.0
				DWA [53]	93.7
				PLASTA-MTL	**97.6**

TABLE VI

mAP Scores Associated With the Different Combinations of Tasks (the BigEarthNet-S2 Archive)

Tasks				Method	mAP (%)
T_1	T_2	T_3	T_4		
✗	✓	✓	✗	Equal Weighting	95.9
				Uncertainty Weighting [43]	83.8
				PCGrad [51]	96.3
				GradNorm [52]	90.4
				DWA [53]	94.7
				PLASTA-MTL	**97.2**
✗	✓	✗	✓	Equal Weighting	87.7
				Uncertainty Weighting [43]	92.0
				PCGrad [51]	92.1
				GradNorm [52]	84.0
				DWA [53]	88.2
				PLASTA-MTL	**93.4**
✓	✗	✓	✓	Equal Weighting	95.7
				Uncertainty Weighting [43]	96.3
				PCGrad [51]	87.0
				GradNorm [52]	92.6
				DWA [53]	94.7
				PLASTA-MTL	**97.4**
✗	✓	✓	✓	Equal Weighting	80.4
				Uncertainty Weighting [43]	90.7
				PCGrad [51]	85.5
				GradNorm [52]	89.4
				DWA [53]	90.7
				PLASTA-MTL	**93.8**
✓	✓	✓	✓	Equal Weighting	94.8
				Uncertainty Weighting [43]	97.3
				PCGrad [51]	93.9
				GradNorm [52]	95.0
				DWA [53]	95.2
				PLASTA-MTL	**97.7**

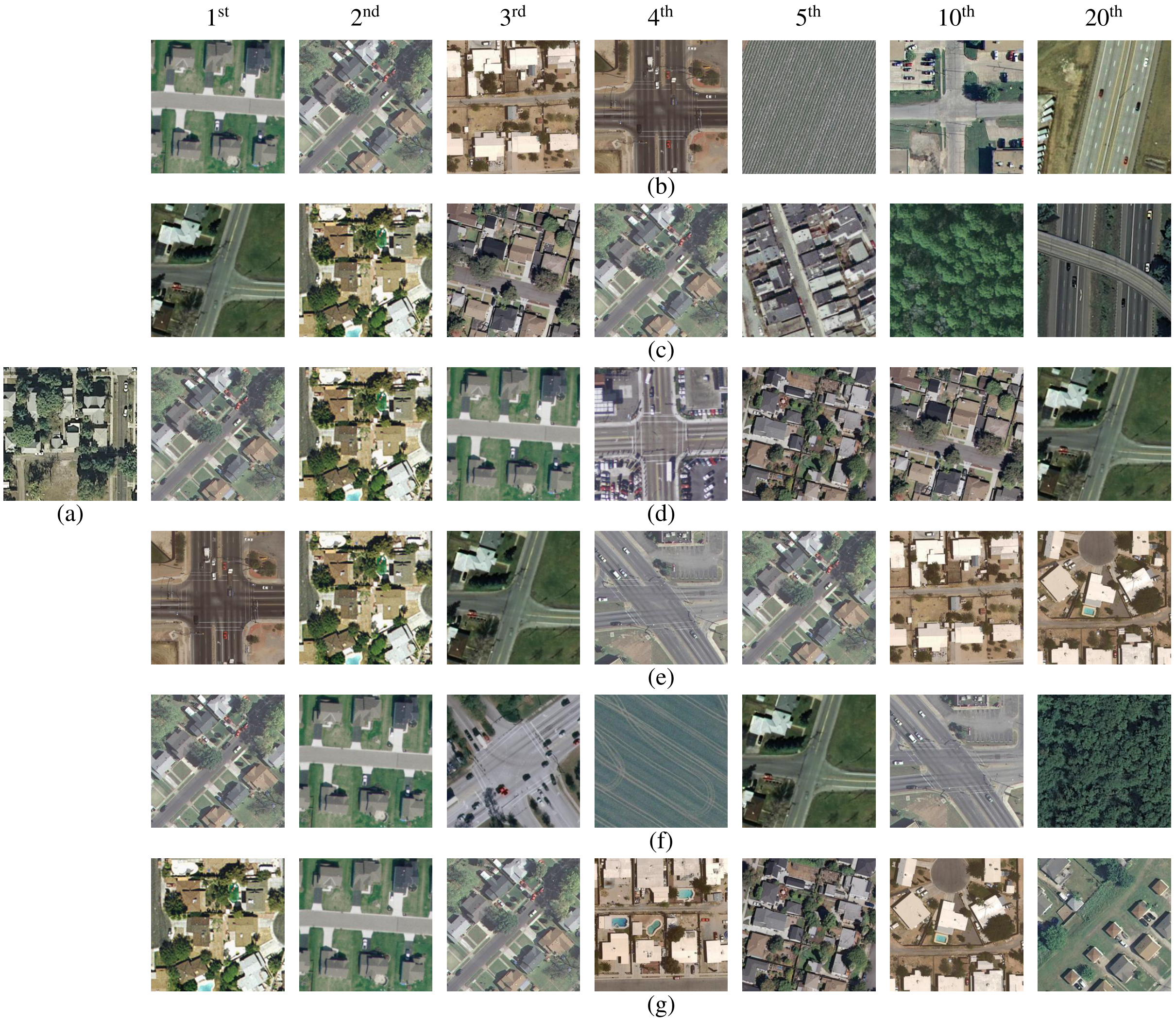
stability preserving capabilities of our approach that overcomes the well-known problems of MTL. Figs. 4 and 5 show the NDCG scores of the considered state-of-the-art methods and our approach under different combinations of the tasks $\{T_1, T_2, T_3, T_4\}$ and different numbers of retrieved images for the DLRSD and the BigEarthNet-S2 archives, respectively. From the figures, one can see that, when the number of retrieved images are increased (from 1 to 50 for DLRSD and 1 to 100 for BigEarthNet-S2), the proposed PLASTA-MTL approach provides the highest NDCG scores for almost all



(a) (b) (c)

(d) (e)

Fig. 5. NDCGs versus the number of retrieved images obtained for the BigEarthNet-S2 archive when the tasks: (a) T_2 and T_3; (b) T_2 and T_4; (c) T_1, T_3, and T_4; (d) T_2, T_3, and T_4; and (e) T_1, T_2, T_3, and T_4 are used in the context of MTL.



(b)

(c)

(a) (d)

(e)

(f)

(g)

Fig. 6. (a) Query image. Images retrieved by using (b) equal weighting, (c) uncertainty weighting, (d) PCGrad, (e) GradNorm, (f) DWA, and (g) proposed PLASTA-MTL approach when the tasks: T_1, T_2, T_3, and T_4 are utilized for the DLRSD archive.

task combinations at each number of retrieved images on both archives. For the DLRSD archive, Fig. 6 shows an example of a query image and the retrieved images by these methods and our approach when all the tasks are utilized and query image contains the classes of *buildings*, *cars*, *grass*, *pavement*, and *trees*. The retrieval orders of images are given above the figure. By assessing the figure, one can observe that the proposed PLASTA-MTL approach leads to retrieval of similar images at all retrieval orders [see Fig. 6(g)]. However, by using state-of-the-art MTL methods, retrieved images contain classes that are not present in the query image. As an example, the equal weighting and the DWA methods lead to retrieval of the image, which includes only *field* class, at the fifth and fourth retrieval orders, respectively [see Fig. 6(b) and (d)]. We observed similar behaviors of these methods for the BigEarthNet-S2 archive. We would like to point out that these methods employ different gradient adjustment strategies for overcoming the well-known problems of MTL. Accordingly, their success has been proven for many MTL problems in the computer vision domain. However, since they do not consider the stability-plasticity constraint of MTL, and they are still based on the joint optimization algorithm, they are limited to solve all possible problems of MTL under various task combinations for RS images. This leads to less accurate image representations learned via these methods compared to the proposed PLASTA-MTL approach. Accordingly, the image representations learned via our approach lead to more effective CBIR results.

VI. Conclusion

In this article, we have proposed a novel PLASTA-MTL approach for CBIR applications. This approach is characterized by novel: 1) PPL function; 2) SPL function; and 3) sequential optimization algorithm. The PPL function allows our approach to minimize the differences of gradient magnitudes for the global representation space and each task-specific embedding space of the considered DNN. The use of the SPL function in the proposed PLASTA-MTL approach leads to the minimization of the angular distances between task gradients over global image representation space. The proposed optimization algorithm sequentially optimizes: 1) each task-specific objective with the corresponding PPL function and 2) the SPL function for all the considered tasks. Experimental results conducted on two benchmark archives show the effectiveness of the proposed PLASTA-MTL approach over the state-of-the-art MTL methods in the context of CBIR. The main reasons for the success of our approach are summarized as follows.

1) Due to the proposed PPL function, the PLASTA-MTL approach enforces the global image representation space to be sensitive to new information learned with each task that leads to the preservation of plasticity condition for the considered DNN.
2) Due to the proposed SPL function, the PLASTA-MTL approach protects the global image representation space radically disrupted by a new task that leads to the preservation of stability conditions for the considered DNN.
3) Due to the proposed sequential optimization algorithm, the PLASTA-MTL approach accurately characterizes: 1) the plasticity condition for each task and 2) the stability condition in between consecutive tasks.
4) Due to the effective combination of multiple tasks independently of the number and type of tasks while considering the stability-plasticity constraint of MTL without the need for selection of loss weights, the PLASTA-MTL approach prevents: 1) conflicts between tasks; 2) the dominance of one of the tasks; and 3) underperformance of tasks compared to STL. This leads to more accurate image representation learning compared to utilizing a single task and the conventional deep MTL procedures.

It is worth noting that, in this article, the PLASTA-MTL approach is considered especially for CBIR applications. Moreover, the global image representation space learned via our approach can be also used for other applications since it applies image representation learning based on the information learned via multiple tasks to represent the complex semantic content of RS images. This can be achieved by applying fine-tuning to the pretrained backbone of the PLASTA-MTL approach for downstream applications in RS. We would like to also point out that the set of all tasks is assumed to be known during the training of our approach. However, the inclusion of new tasks to the set of considered tasks after training for the PLASTA-MTL approach can further improve the characterization of RS image content. Accordingly, as future development of this work, we plan to study continual learning to include new tasks to the PLASTA-MTL approach after completing the whole learning procedure while preserving its plasticity and stability capabilities also for these tasks.

References

[1] Y. Yang and S. Newsam, “Geographic image retrieval using local invariant features,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 818–832, Feb. 2013.

[2] E. Aptoula, “Remote sensing image retrieval with global morphological texture descriptors,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 3023–3034, May 2014.

[3] O. E. Dai, B. Demir, B. Sankur, and L. Bruzzone, “A novel system for content-based retrieval of single and multi-label high-dimensional remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 7, pp. 2473–2490, Jul. 2018.

[4] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, “Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 1144–1158, Feb. 2018.

[5] Y. Bengio, A. Courville, and P. Vincent, “Representation learning: A review and new perspectives,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.

[6] M. Guo, C. Zhou, and J. Liu, “Jointly learning of visual and auditory: A new approach for RS image and audio cross-modal retrieval,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 11, pp. 4644–4654, Nov. 2019.

[7] F. Ye, W. Luo, M. Dong, D. Li, and W. Min, “Content-based remote sensing image retrieval based on fuzzy rules and a fuzzy distance,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[8] F. Ye, M. Dong, W. Luo, X. Chen, and W. Min, “A new re-ranking method based on convolutional neural network and two image-to-class distances for remote sensing image retrieval,” *IEEE Access*, vol. 7, pp. 141498–141507, 2019.

[9] F. Ye, X. Zhao, W. Luo, D. Li, and W. Min, “Query-adaptive remote sensing image retrieval based on image rank similarity and image-to-query class similarity,” *IEEE Access*, vol. 8, pp. 116824–116839, 2020.

[10] C. Liu, J. Ma, X. Tang, F. Liu, X. Zhang, and L. Jiao, “Deep hash learning for remote sensing image retrieval,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3420–3443, Apr. 2021.

[11] R. Imbriaco, C. Sebastian, E. Bondarev, and P. H. N. de With, “Aggregated deep local features for remote sensing image retrieval,” *Remote Sens.*, vol. 11, no. 5, p. 493, Feb. 2019.

[12] Y. Boualleg and M. Farah, “Enhanced interactive remote sensing image retrieval with scene classification convolutional neural networks model,” in *Proc. IEEE Intl. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 4748–4751.

[13] W. Zhou, S. Newsam, C. Li, and Z. Shao, “Learning low dimensional convolutional neural networks for high-resolution remote sensing image retrieval,” *Remote Sens.*, vol. 9, no. 5, p. 489, 2017.

[14] F. Hu, X. Tong, G.-S. Xia, and L. Zhang, “Delving into deep representations for remote sensing image retrieval,” in *Proc. IEEE 13th Int. Conf. Signal Process. (ICSP)*, Nov. 2016, pp. 198–203.

[15] F. Ye, H. Xiao, X. Zhao, M. Dong, W. Luo, and W. Min, “Remote sensing image retrieval using convolutional neural network features and weighted distance,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1535–1539, Oct. 2018.

[16] C. Ma, F. Chen, J. Yang, J. Liu, W. Xia, and X. Li, “A remote-sensing image-retrieval model based on an ensemble neural networks,” *Big Earth Data*, vol. 2, no. 4, pp. 351–367, Feb. 2018.

[17] R. Cao *et al.*, “Enhancing remote sensing image retrieval using a triplet deep metric learning network,” *Int. J. Remote Sens.*, vol. 41, no. 2, pp. 740–751, Jan. 2020.

[18] L. Fan, H. Zhao, and H. Zhao, “Distribution consistency loss for large-scale remote sensing image retrieval,” *Remote Sens.*, vol. 12, no. 1, p. 175, Jan. 2020.

[19] U. Chaudhuri, B. Banerjee, and A. Bhattacharya, “Siamese graph convolutional network for content based remote sensing image retrieval,” *Comput. Vis. Image Understand.*, vol. 184, pp. 22–30, Jul. 2019.

[20] U. Chaudhuri, B. Banerjee, A. Bhattacharya, and M. Datcu, “A zero-shot sketch-based intermodal object retrieval scheme for remote sensing images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[21] M. Zhang, Q. Cheng, F. Luo, and L. Ye, “A triplet nonlocal neural network with dual-anchor triplet loss for high-resolution remote sensing image retrieval,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2711–2723, 2021.

[22] Y. Li, Y. Zhang, X. Huang, and J. Ma, “Learning source-invariant deep hashing convolutional neural networks for cross-source remote sensing image retrieval,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6521–6536, Nov. 2018.

[23] Y. Li, Y. Zhang, X. Huang, H. Zhu, and J. Ma, “Large-scale remote sensing image retrieval by deep hashing neural networks,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 950–965, Feb. 2018.

[24] P. Li *et al.*, “Hashing nets for hashing: A quantized deep learning to hash framework for remote sensing image retrieval,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7331–7345, Oct. 2020.

[25] W. Xiong, Z. Xiong, Y. Zhang, Y. Cui, and X. Gu, “A deep cross-modality hashing network for SAR and optical remote sensing images retrieval,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5284–5296, 2020.

[26] Y. Cao *et al.*, “DML-GANR: Deep metric learning with generative adversarial network regularization for high spatial resolution remote sensing image retrieval,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8888–8904, Dec. 2020.

[27] L. Fan, H. Zhao, and H. Zhao, “Global optimization: Combining local loss with result ranking loss in remote sensing image retrieval,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 7011–7026, Aug. 2021.

[28] Y. Liu, L. Ding, C. Chen, and Y. Liu, “Similarity-based unsupervised deep transfer learning for remote sensing image retrieval,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7872–7889, Nov. 2020.

[29] S. Roy, E. Sangineto, B. Demir, and N. Sebe, “Metric-learning-based deep hashing network for content-based retrieval of remote sensing images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 226–230, Feb. 2021.

[30] X. Tang, X. Zhang, F. Liu, and L. Jiao, “Unsupervised deep feature learning for remote sensing image retrieval,” *Remote Sens.*, vol. 10, no. 8, p. 1243, Aug. 2018.

[31] N. Khurshid, M. Tharani, M. Taj, and F. Z. Qureshi, “A residual-dyad encoder discriminator network for remote sensing image matching,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2001–2014, Mar. 2020.

[32] Z. Shao, W. Zhou, X. Deng, M. Zhang, and Q. Cheng, “Multilabel remote sensing image retrieval based on fully convolutional network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 318–328, 2020.

[33] G. Sumbul and B. Demir, “A novel graph-theoretic deep representation learning method for multi-label remote sensing image retrieval,” in *Proc. IEEE Intl. Geosci. Remote Sens. Symp.*, Jul. 2021, pp. 266–269.

[34] G. Hoxha, F. Melgani, and B. Demir, “Toward remote sensing image retrieval under a deep image captioning perspective,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4462–4475, 2020.

[35] G. Sumbul, J. Kang, and B. Demir, “Deep learning for image search and retrieval in large remote sensing archives,” in *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences*. Hoboken, NJ, USA: Wiley, 2021, pp. 150–160, ch. 11.

[36] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, “When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative CNNs,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2811–2821, May 2018.

[37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[38] Y. shu Liu, Z. Han, C. Chen, L. Ding, and Y. Liu, “Eagle-eyed multitask CNNs for aerial image retrieval and scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6699–6721, Sep. 2020.

[39] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai, and L. Van Gool, “Multi-task learning for dense prediction tasks: A survey,” *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jan. 26, 2021, doi: 10.1109/TPAMI.2021.3054719.

[40] X. Zhao, H. Li, X. Shen, X. Liang, and Y. Wu, “A modulation module for multi-task learning with applications in image retrieval,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 415–432.

[41] R. M. French, “Catastrophic forgetting in connectionist networks,” *Trends Cogn. Sci.*, vol. 3, no. 4, pp. 128–135, Apr. 1999.

[42] O. Sener and V. Koltun, “Multi-task learning as multi-objective optimization,” in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 525–536.

[43] R. Cipolla, Y. Gal, and A. Kendall, “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7482–7491.

[44] J. Fang, X. Cao, D. Wang, and S. Xu, “Multitask learning mechanism for remote sensing image motion deblurring,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2184–2193, 2021.

[45] F. Chen and B. Yu, “Earthquake-induced building damage mapping based on multi-task deep learning framework,” *IEEE Access*, vol. 7, pp. 181396–181404, 2019.

[46] R. C. Daudt, B. Le Saux, A. Boulch, and Y. Gousseau, “Multitask learning for large-scale semantic change detection,” *Comput. Vis. Image Understand.*, vol. 187, Oct. 2019, Art. no. 102783.

[47] X. Lu *et al.*, “Multi-scale and multi-task deep learning framework for automatic road extraction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9362–9377, Nov. 2019.

[48] H. Wang, Z. Zhou, H. Zong, and L. Miao, “Wide-context attention network for remote sensing image retrieval,” *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 12, pp. 1–5, Dec. 2020.

[49] W. Xiong, Z. Xiong, Y. Cui, and Y. Lv, “A discriminative distillation network for cross-source remote sensing image retrieval,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1234–1247, 2020.

[50] W. Song, S. Li, and J. A. Benediktsson, “Deep hashing learning for visual and semantic retrieval of remote sensing images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9661–9672, Nov. 2021.

[51] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, “Gradient surgery for multi-task learning,” in *Proc. Intl. Conf. Neural Inf. Process. Syst.*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds., vol. 33, 2020, pp. 5824–5836.

[52] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, “GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks,” in *Proc. Int. Conf. Mach. Learn.*, vol. 80, 2018, pp. 794–803.

[53] S. Liu, E. Johns, and A. J. Davison, “End-to-end multi-task learning with attention,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2019, pp. 1871–1880.

[54] K.-K. Maninis, I. Radosavovic, and I. Kokkinos, “Attentive single-tasking of multiple tasks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1851–1860.

[55] M. Suteu and Y. Guo, “Regularizing deep multi-task networks using orthogonal gradients,” 2019, *arXiv:1912.06844*.

[56] A. Migdalas, P. Pardalos, and P. Värbrand, *Multilevel Optimization: Algorithms and Applications*. Boston, MA, USA: Springer, 2013.

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

[58] G. Sumbul *et al.*, "BigEarthNet-MM: A large scale multi-modal multi-label benchmark archive for remote sensing image classification and retrieval," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 174–180, May 2021.

[59] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, 2010, pp. 270–279.

[60] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[61] G. Sumbul and B. Demir, "A deep multi-attention driven approach for multi-label remote sensing image classification," *IEEE Access*, vol. 8, pp. 95934–95946, 2020.

[62] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 815–823.

[63] A. Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification," 2017, *arXiv:1703.07737*.

[64] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Intl. Conf. Mach. Learn.*, vol. 119, 2020, pp. 1597–1607.

[65] Z. Zhang, Q. Zou, Y. Lin, L. Chen, and S. Wang, "Improved deep hashing with soft pairwise similarity for multi-label image retrieval," *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 540–553, Feb. 2020.

Gencer Sumbul received the B.S. and M.S. degrees in computer engineering from Bilkent University, Ankara, Turkey, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree with the Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin, Berlin, Germany.

He is currently a Research Associate with the Remote Sensing Image Analysis (RSiM) Group, Technische Universität Berlin. His research interests include computer vision, pattern recognition, and machine learning, with a special interest in deep learning, large-scale image understanding, and remote sensing.

Mr. Sumbul is also a Referee for journals, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE ACCESS, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the ISPRS *Journal of Photogrammetry and Remote Sensing*, and international conferences, such as the European Conference on Computer Vision and the IEEE International Geoscience and Remote Sensing Symposium.

Begüm Demir (Senior Member, IEEE) received the B.S., M.Sc., and Ph.D. degrees in electronic and telecommunication engineering from Kocaeli University, İzmit, Turkey, in 2005, 2007, and 2010, respectively.

She is currently a Full Professor and the Founder Head of the Remote Sensing Image Analysis (RSiM) Group, Faculty of Electrical Engineering and Computer Science, Technische Universität Berlin (TU Berlin), Berlin, Germany, where she is also the Head of the Big Data Analytics for Earth Observation Research Group, Berlin Institute for the Foundations of Learning and Data (BIFOLD). Her research activities lie at the intersection of machine learning, remote sensing, and signal processing. Specifically, she performs research in the field of processing and analysis of large-scale Earth observation data acquired by airborne and satellite-borne systems.

Dr. Demir is also a Scientific Committee Member of several international conferences and workshops, such as the Conference on Content-Based Multimedia Indexing, the Conference on Big Data from Space, the Living Planet Symposium, the International Joint Urban Remote Sensing Event, the SPIE International Conference on Signal and Image Processing for Remote Sensing, and the Machine Learning for Earth Observation Workshop organized within the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD). She is also a fellow of the European Lab for Learning and Intelligent Systems (ELLIS). She was awarded the prestigious 2018 Early Career Award by the IEEE Geoscience and Remote Sensing Society for her research contributions in machine learning for information retrieval in remote sensing. In 2018, she received a Starting Grant from the European Research Council (ERC) for her project "BigEarth: Accurate and Scalable Processing of Big Data in Earth Observation." She is also a Referee for several journals, such as the PROCEEDINGS OF THE IEEE, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, *Pattern Recognition*, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, the *International Journal of Remote Sensing*, and several international conferences. She is also an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Remote Sensing* (MDPI), and *International Journal of Remote Sensing*.