Interband Retrieval and Classification Using the Multilabeled Sentinel-2 BigEarthNet Archive

Ushasi Chaudhuri, *Student Member, IEEE*, Subhadip Dey, *Student Member, IEEE*, Mihai Datcu, *Fellow, IEEE*, Biplab Banerjee, *Member, IEEE*, and Avik Bhattacharya, *Senior Member, IEEE*

***Abstract*—Conventional remote sensing data analysistechniques have a significant bottleneck of operating on a selectively chosen small-scale dataset. Availability of an enormous volume of data demands handling large-scale, diverse data, which have been made possible with neural network-based architectures. This article exploits the contextual information capturing ability of deep neural networks, particularly investigating multispectral band properties from Sentinel-2 image patches. Besides, an increase in the spatial resolution often leads to nonlinear mixing of land-cover types within a target resolution cell. We recognize this fact and group the bands according to their spatial resolutions, and propose a classification and retrieval framework. We design a representation learning framework for classifying the multispectral data by first utilizing all the bands and then using the grouped bands according to their spatial resolutions. We also propose a novel triplet-loss function for multilabeled images and use it to design an interband group retrieval framework. We demonstrate its effectiveness over the conventional triplet-loss function. Finally, we present a comprehensive discussion of the obtained results. We thoroughly analyze the performance of the band groups on various land-cover and land-use areas from agro-forestry regions, water bodies, and human-made structures. Experimental results for the classification and retrieval framework on the benchmarked BigEarthNet dataset exhibit marked improvements over existing studies.**

***Index Terms*—Interband retrieval, multilabel classification, multilabel cross triplet loss, multimodal classification, Sentinel-2, land-cover classification.**

I. Introduction

IMAGES from multispectral and hyperspectral sensors have found wide applications, ranging from mining [1], oceanography [2], agriculture [3], meteorological studies [4], geological observations [5], to name a few. Multispectral satellites consist of several spectral bands, which image the land surface with multiple spatial resolutions. Each spectral band essentially captures specific physical information from these distinctive land surface covers. This information essentially depends on the interaction of the electromagnetic waves of particular wavelengths with the physical and geochemical characteristics of the land cover surface within a sensor resolution cell. Therefore, this information helps in efficiently characterizing different land cover classes, such as vegetation and water-bodies.

Understanding the data has become crucial using neural networks, followed by advances in various deep learning frameworks. While using a conventional dense network, we usually neglect the information about the neighborhood pixel. This information holds vital information about the change of pixel characteristics [6]. For example, let us consider an image that consists of a water body and a beach. A conventional dense network might not focus on the boundary of these two land features. However, a convolutional neural networks (CNNs) considers the spatial heterogeneity in terms of their spatial distribution and neighborhood pixel information. Hence, CNNs will be able to differentiate these various land cover types along with their corresponding boundaries. Moreover, the main advantage of CNNs is that it automatically detects essential features from contexts [7].

Several studies aim to bridge the gap between the feature embeddings of multisensor imagery. However, as different sensors acquire images over a different time, a particular region in the acquired data may suffer from changes in the local weather or land cover (during a harvest season or post a natural disaster). In such a case, these multisensor acquired images are more fit for a change detection task rather than a fusion or cross/multimodal retrieval task. Drawing motivation from this, we aim to look into multimodal data classification and retrieval wherein there has been no change within the acquisitions. We choose to consider imageries acquired by a single satellite for this objective and propose using different bands grouped based on their spatial resolution as different data modalities. This article primarily aims to study classification and retrieval among interband groups with precisely the same region data without any externally induced change.

To utilize all the bands together, we either downsize, interpolate, or super resolve a few bands to bring them to a common spatial resolution. However, a significant drawback in multilabeled data is that we often get a nonlinear mixing of end-members within a target land-cover region with increased spatial resolution. Therefore, if we downsize the images, we end up missing a considerable amount of information. Hence, in this study, we group the bands according to the band

Manuscript received July 26, 2021; revised September 3, 2021; accepted September 8, 2021. Date of publication September 14, 2021; date of current version October 11, 2021. The work (Research internship) of Ushasi Chaudhuri was supported in part by the Conservatoire National des Arts et Métiers (CNAM), Paris, France and in part by the Campus France. *(Corresponding author: Ushasi Chaudhuri.)*

Ushasi Chaudhuri, Subhadip Dey, Biplab Banerjee, and Avik Bhattacharya are with the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai 400076, India (e-mail: ushasi2cool@gmail.com; subhadipdey23071994@gmail.com; getbiplab@.com; avikb@csre.iitb.ac.in).

Mihai Datcu is with the German Aerospace Center (DLR), D 82234 Wessling, Germany (e-mail: mihai.datcu@dlr.de).

Digital Object Identifier 10.1109/JSTARS.2021.3112209

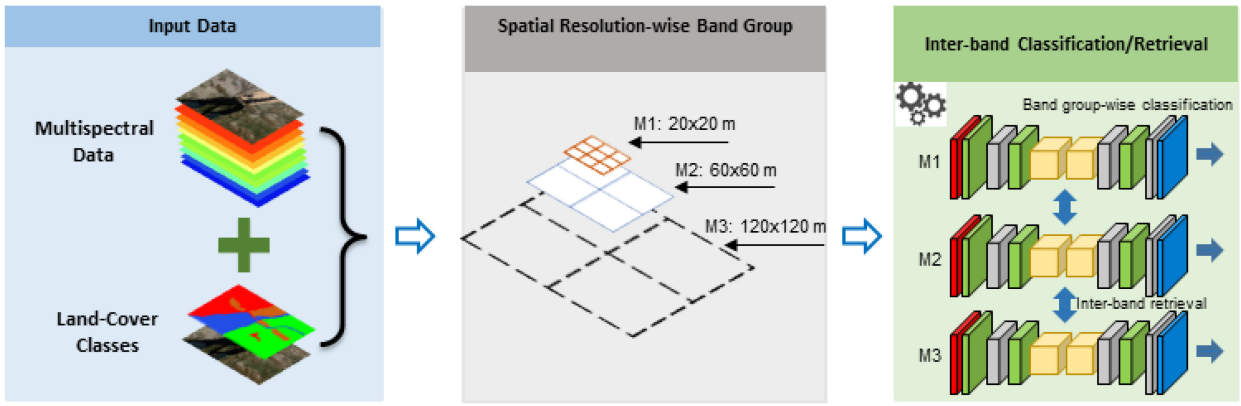

Fig. 1. Block diagram of the complete framework.

spatial resolutions and propose a classification and retrieval framework.

Several studies recommended using handcrafted feature extractors to develop a robust multilabel classification model. The normalized difference vegetation index (NDVI) is the most widely used descriptor for green vegetation region detector [8]. Likewise, there are several such conventional handcrafted indices used for detecting various land-cover regions. With the onset of CNN-based feature extracting procedures, the complexity of different classes exploded with deep neural networks predominantly handling the classification task [9], [10]. Several classification studies have been reported using multispectral satellite data [9], [11]–[13].

While classification has remained a classical problem description in remote sensing (RS), retrieval tasks have found more attention recently with extensive data acquisition by various satellite missions using different sensor technologies. Therefore, this challenges having an interband group retrievable framework [14]. Since each band provides distinct land-cover reflectance properties, it is imperative to have an interband group retrieval framework. Recently, we have seen a lot of focus on cross-sensor/cross-modal retrieval techniques in RS using various learning techniques [15], [16]. Some of the notable works in this domain are presented in [17]–[20]. Several literary works in this domain have tried to exploit the conventional triplet loss or the Siamese loss function to discriminate classes within a fine-grained dataset. One major shortcoming in using this for a multilabeled dataset ignores the presence of common classes between a positive class image and a negative class image to the anchor image. This condition leads to the learning up of scattered clusters for each class in the embedding space. To overcome this concern, we propose a modified triplet loss for multilabeled images.

Once we establish an interband group retrieval framework, it is easier to group certain bands based on their spatial resolution and consequently study the band properties of those modalities. To better understanding channel properties in SAR RS applications, various works have been proposed (e.g., [21], [22]). However, to the best of authors knowledge, no such work has been carried out on multispectral sensors, which uncovers many more widespread applications. We elaborate on the classification and retrieval performance of our framework on various band groups and comprehend the overall band group properties. We show an overview of the problem statement in Fig. 1.

How are we different? In this work, we perform a comprehensive study of the band properties of the Sentinel-2 multispectral data. First, we design a multilabel classification network for classifying the BigEarthNet dataset. The proposed network is better than the state-of-the-art (SOTA) and yields a good performance.

We then split the data into three groups based on the spatial resolution of each band. For each of these groups of bands, we test the interband multilabel classification performance. We utilize the same model, which outperforms the SOTA in classification with all the bands combined for this purpose. While keeping the network architecture intact, we vary the number of convolution kernel channels depending on the number of bands in the corresponding modality. The proposed method produces good discriminative features to describe the multilabeled land cover classes with lower dimensions.

Furthermore, we propose an interband multilabel retrieval architecture using a novel triplet loss function. We demonstrate that this function is better than the conventional triplet loss function for multilabeled images. This study provides an assessment of bands that effectively contribute toward better land cover retrieval tasks. Finally, we also study the properties of each band and explore their contribution in classifying each land use/land cover class. None of the literary works using multispectral satellites have studied its band properties to the best of our knowledge.

We summarize the main contributions of this article as follows.

1) We propose a multilabel classification network using representation learning using all the bands of the Sentinel-2 that performs comparably to the SOTA performance on the large-scale benchmarked BigEarthNet dataset [23].
2) Using the abovementioned framework, we analyze the multilabel classification performance of the band group-wise (bands clubbed together by their spatial resolution). We also find class-wise identification of each land-cover class in different band groups.
3) We propose a novel modified cross-triplet loss-based metric learning technique for retrieving multilabeled images

and demonstrate its efficacy over the conventional triplet loss in the experimental section.
4) Using the proposed modified cross-triplet loss, we design an interband group retrieval framework among these modalities.

II. Related Works

With the onset of deep learning technologies, the past decade has successfully handled various heterogeneous applications in RS. The deep learning-based approach primarily has demonstrated its superiority in feature extraction in several types of satellite images. This strategy has opened up a new research extent in RS for data classification and retrieval. The following sections discuss current relevant work using deep learning in:
1) classification;
2) retrieval;
3) understanding band properties, respectively.

A. Classification in RS

In RS, deep learning techniques have been successively and successfully utilized in classifying images acquired from different sensors. Different sensors capture different information, which necessitates having robust classifiers compatible with various forms of RS data. For multispectral satellite data classification, various studies have been conducted [9], [11]–[13]. Chaib *et al.* [9] proposed a simple framework by exploiting the features constructed by a VGG pretrained network [24]. They perform a discriminant correlation analysis using these features for refining the original features using information fusion, enabling them to obtain a good scene classification framework. Xia *et al.* [11] proposed a benchmarked dataset for aerial scene classification. In [12], the authors reported a similar very-high resolution land cover classification by augmenting the RS data and using a standard transfer learning from a pretrained network. Cheng *et al.* [13] reviewed the important literature in RS for scene classification while also proposing another benchmarked dataset. Similarly, quite a few studies have also reported hyperspectral image classification [25]–[28], synthetic aperture radar (SAR) classification [29], [30], polarimetric SAR (PolSAR) classification [21], [31], RS object detection [16], etc.

Xu *et al.* [32] classified land cover types using the features derived from a CNN architecture. They utilized the derived features directly into the support vector machine classifier for the classification purpose. According to their study, these derived features enhanced the classification accuracy by 2.65% as compared to other traditional methods. In another study, Runyu *et al.* [33] proposed a semisupervised multi-CNN ensemble learning method to classify different land cover types. Their proposed method outperforms other existing methods by 3% to 4% overall accuracy score. Therefore, these extracted features are essential in classifying several land cover classes within a study area [34]–[39]. Moreover, CNN has gained importance in land cover classification due to its flexibility and adaptation capability in several land cover scenarios for different sensing platforms.

B. Retrieval in RS

There exists a plethora of work in the literature on image retrieval in RS [40]–[43]. Most retrieval frameworks create a hashed feature space, as it is much faster to search for the nearest neighbor using a hamming distance measure. In [41], the authors proposed a kernel-based nonlinear hashing technique for retrieving large-scale RS archives. They leverage the semantic similarity of the annotated images in the dataset to construct the hashed space. Xia *et al.* [44] provide a detailed review of all the important literary works in RS unimodal data retrieval.

While most of the literary work is on unimodal data retrieval, exploiting the robust feature extraction capability of CNNs, new studies are appearing that address cross-modal data retrieval. The need for multimodal approaches is particularly more evident in the area of RS data analysis due to the availability of a large number of satellite missions with various sensors and complementary information obtained by multiple cross-sensor acquisitions [14], [19]. A similar study [15] has addressed this strategy using SAR and multispectral images and thereby has proposed the SEN12MS dataset, which consists of Sentinel-1 and Sentinel-2 images over the same patch across various geographical locations. SAR images provide backscatter information of a target at microwave wavelength that penetrates the atmospheric layer and often provides added information than multispectral data. These cross-sensor retrieval techniques also extend to retrieving unseen class images upon deployment, commonly called zero-shot cross-modal retrieval [45]. As one of the major bottlenecks of solving RS problems using deep learning is the lack of annotated samples for training, these cross-sensor zero-shot retrieval has received a lot of attention recently [18].

C. Physical Parameter Estimation or Understanding Band Properties

Various studies have been proposed using multiple sensors to interpret physical parameters from RS data. For example, one uses the double-bounce scattering mechanism to classify urban areas and city blocks. In contrast, volume scattering helps detect forests and dense vegetation areas from PolSAR data. Similarly, surface scattering characterizes flat terrains and ocean covers. Zhao *et al.* [21] used a contrastive-regulated CNN to learn the physical parameters from PolSAR images. In [22], the authors utilized SAR images to learn the spatial texture and backscattering patterns from target areas.

Several studies have proposed classification by exploiting the physical properties of each spectral band and the corresponding target behaviors using multispectral data [46]. However, most of these studies directly utilize these bands for specific tasks of classification and retrieval. For example, one uses thermal infrared bands to measure the land surface temperature changes. Similarly, bands three, five, and seven from Landsat-8 are used for snow cover detection [47]. Likewise, each band has its unique physical characteristics, which are suitably exploited for various

TABLE I
EXAMPLES OF FEW TYPICAL APPLICATIONS OF EACH SPECTRAL BAND OF SENTINEL 2

Band	**Res.** (m)	**Imagery**	**Utility**
Band 1	60	Coastal aerosol	Coastal and aerosol studies.
Band 2	10	Blue	Bathymetric mapping, distinguishing soil from vegetation and deciduous from coniferous vegetation.
Band 3	10	Green	Emphasizes peak vegetation, helps in assessing plant vigor.
Band 4	10	Red	Discriminates vegetation slopes.
Band 5	20	Vegetation Red Edge	Useful estimator of green Leaf Area Index [8], Emphasizes biomass content and shorelines.
Band 6	20	Vegetation Red Edge	Useful for estimating canopy chlorophyll [8]. Discriminates moisture content of soil and vegetation.
Band 7	20	Vegetation Red Edge	Applications similar to Band 5 and 6.
Band 8	10	Near Infra Red	Detect healthy/unhealthy vegetation and discriminates dense canopy from urban and water bodies.
Band 8A	20	Vegetation Red Edge	Detects bare soil and built-up areas, and along with Band B12, helps in calculating moisture indexes.
Band 9	60	Water Vapour	Primarily used for water vapour detection.
Band 10	60	Short Wave IR	Detection of Cirrus clouds
Band 11	20	Short Wave IR	Discriminates moisture content of soil and vegetation, monitor health of crops, thermal mapping.
Band 12	20	Short Wave IR	Discriminates soil and vegetation moisture, geological faults, features and formations, and lithology.

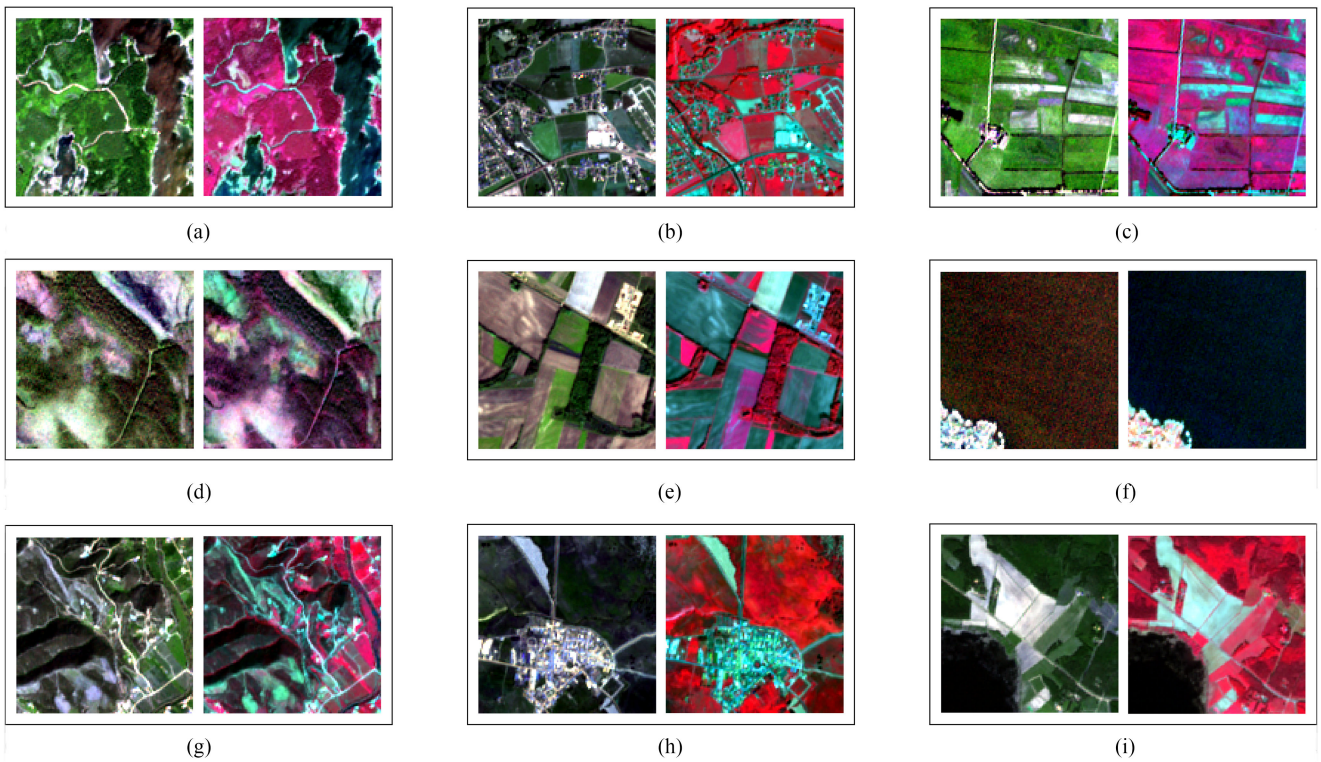

Fig. 2. Sample instances of TCC and FCC of a few multilabeled image patches. FCC constructed using band 8, 4, and 3 as RGB, while TCC images are synthesized by using the bands 4, 3, and 2 as RGB. (a) *Coniferous forest*, mixed forest, sea and ocean, (b) Discontinuous urban fabric, industrial or commercial units, nonirrigated arable land, (c) Nonirrigated arable land, *Pastures*, complex cultivation patterns, (d) Coniferous forest, transitional woodland/shrub, Peatbogs, (e) Discontinuous urban fabric, nonirrigated arable land, broad-leaved forest, (f) Bare rock, sea and ocean, (g) Fruit trees and berry plantations, Sclerophyllous vegetation, transitional woodland/shrub, (h) Continuous urban fabric, (i) Nonirrigated arable land, *agriculture and natural vegetation*, mixed forest, inland marshes, water bodies.

applications. We tabulate specific critical applications using each spectral band of Sentinel-2 as reported in the literature in Table I.

Robinson *et al.* [48] proposed a multiresolution data fusion method for high-resolution land cover mapping. They identify the challenges of deep learning-based land cover mapping and propose techniques to overcome these challenges. They primarily tackle a multimodal/multiresolution data fusion task to develop an efficient land cover mapping framework. To this end, the authors use a large-scale database comprising over eight trillion pixels to train the model.

III. DATA PREPARATION AND GROUND TRUTH GENERATION

A. Data Preparation

The Sentinel-2 satellite has 13 bands with different spatial resolutions and different sizes of images. Sumbul *et al.* [23] reported the BigEarthNet dataset that was created using the Sentinel-2 images. Here, we demonstrate our work on this BigEarthNet dataset. Fig. 2 shows the true color composites (TCCs) and false-color composites (FCCs) of a few sample images from the BigEarthNet dataset over a diverse subset of

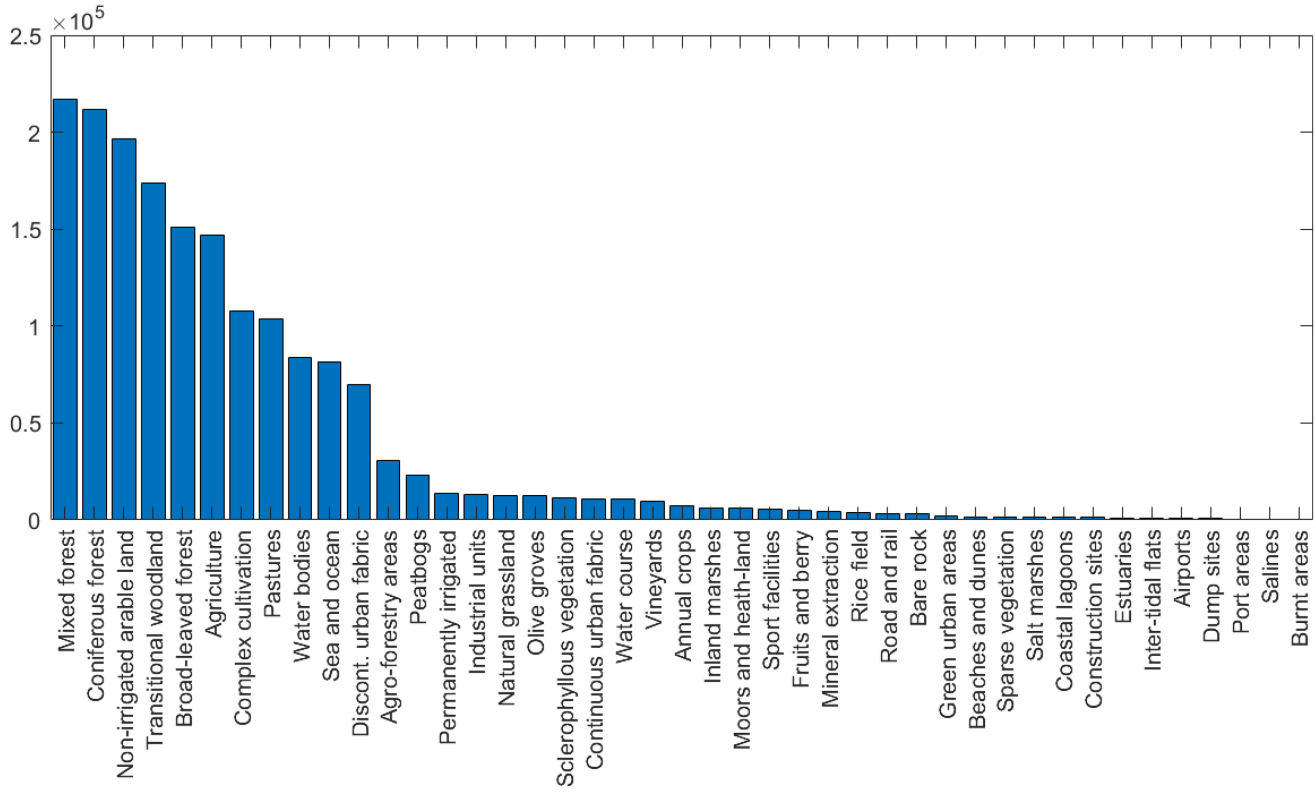

Fig. 3. Land-cover classes versus the number of image patches of each of these categories. Plot distribution highlights heavy data imbalance.

land-cover classes. The FCCs are made using bands 8, 4, and 3, while the TCCs are made using bands 4, 3, and 2.

In the literature, the spatial resolution is shown to be one of the critical parameters for the classification of diverse land cover types [49], [50]. The effectiveness of any classification or retrieval problem primarily depends on the texture information in the data and the existence of pure end-member within a pixel. With an increase in the spatial resolution, the possibility of detecting a pure end-member in a pixel reduces due to the high chance of nonlinear mixing of several other end-members within the resolution cell. This characteristic becomes especially more crucial and challenging for multilabeled RS datasets.

It is noteworthy to say that, due to the tradeoff between spectral and spatial resolutions, lower frequency, and high spectral resolution bands require coarser spatial resolution. The BigEarthNet multilabeled data obtained from the Sentinel-2 sensor have three different spatial resolutions of bands at our disposition depending on their bandwidth and central frequencies. Therefore, we have considered grouping the bands based on their spatial resolution from the available bands in Sentinel-2 data in this study. We split the bands according to their spatial resolutions (viz., 10 m, 20 m, and 60 m), and analyze their contributions to segregate different land-cover classes. The database contains image patches of size:

1) 120×120 pixels in the 10 m bands;
2) 60×60 pixels in the 20 m bands;
3) 20×20 pixels in the 60 m bands.

The band number 10 (out of 13 bans) was excluded from the dataset due to the lack of surface information. Therefore, we have two bands with 60 m resolution, i.e., band number 1 (coastal aerosol) and 9 (water vapor), which we refer to as M1. There are six 20 m bands, i.e., band number 5 (vegetation red edge), 6 (vegetation red edge), 7 (vegetation red edge), 8 A (vegetation red edge), 11 (short wave infrared), and 12 (short wave infrared), which we refer to as M2. There are four 10 m resolution, i.e., band 2 (blue), 3 (green), 4 (red), and 8 (near-infrared), which we refer to as M3.

For the initial classification, we combined band groups (i.e., M1, M2, and M3). We simultaneously obtain the three groups of data from the same satellite (i.e., same sensor). In this work, we also use these different groups, i.e., M1, M2, and M3, separately to analyze their contribution to each land-use/land-cover class. Further investigations involve group-wise classification and interband group retrieval along with band properties interpretation.

The BigEarthNet is a multilabel dataset with labels from 43 different land cover categories. These labels were used from the Corine land cover database. The dataset consists of a total of 590 326 image patches generated from 125 Sentinel-2 image tiles. These tiles were selected from data acquired from June 2017 to May 2018.

B. *Training Data Imbalance*

Fig. 3 shows the bar graph of the number of image instances corresponding to each land-cover class. One can see that the difference between the most and the least frequently occurring classes varies in the range of approximately more than 200 000 samples. This disparity causes a large bias on the network to

learn the more commonly occurring classes. The less frequently occurring classes are seldom learned. So in the training sequence, we need to carefully take care of this fact by either using the weighted average of the number of samples in each class distribution or cleverly select the batches while training the network.

Some of the commonly used strategies for handling data imbalance in the literature are based on rebalancing the dataset and cost-sensitive learning of classifier [51]. Naive-over and undersampling [52], selective decontamination [53], SMOTE [54], GAN-based augmentation [55] are some of the commonly used rebalancing techniques. The cons-sensitive learning approach involves adding focal loss [56] and diversity regularizers [51].

IV. Methodology

Preliminaries: The idea behind the multilabel classification task is to find a mapping function that can help get the labels given an input image. We conduct our studies through the following steps.

1) Multilabel image classification by individual band groups (M1, M2, and M3) and studying the importance of each band group for each class.
2) Designed an interband group, multilabel image retrieval network.
3) Understanding the band properties and their analysis.

We explain these steps in the following sections in detail.

A. Interband Classification

Multilabel image classification is a more challenging problem than single-label image classification algorithms. This is mainly due to preserving the accountability of every minute detail from different categories of images. For a dataset consisting of $\mathcal{Y} = \{1, 2, \ldots, L\}$ distinct land-cover classes, we get $Y_i \subseteq 2^L$ corresponding combinations of possible land-cover multilabels. Here, 2^L is the power set of the set of all labels. This shows how the challenge of classifying multilabel images increases in many folds.

Ideally, we use a softmax function after the last layer of the neural network for a single-label classification problem. After the last layer, we use a sigmoid activation function for a multilabel classification framework to get the class labels. Whether a class is present or not is given by an indicator function, which we set high only when the probability of getting a class is higher than some preset threshold. Formally we define this as given in equation 1. Here, A_n is the indicator function for class n, p_n is the probability for nth class, and threshold α

$$A_n = I[p_k > \alpha]. \tag{1}$$

We try to keep the precision to recall ratio as close to one as possible. If the ratio is less than one, we increase the threshold constant α. Similarly, if the ratio is greater than one, we reduce the value of the threshold constant. We tune the threshold value in this way by a grid-search-based method.

We use the three spatial resolution-wise compiled groups of the data (M1, M2, and M3) from the same satellite for this set of experiments. Since we consider all the three modalities from the same satellite, all the groups capture the image simultaneously. There is no time delay between them. Here, we train the previously designed network for a group-wise multilabel classification to identify the land-cover classes of the image patches.

Training: We design a representation learning network for tackling the problem of multilabel classification. We apply a cubic-interpolation to the 20 m and the 60 m bands of each image to bring them to the same image dimension as the 120-pixel images. We choose to interpolate the data to be consistent with the other state-of-the-art architectures [23], [57], [58] while preserving comparison fairness among them.

We use four sets of convolution—pooling—nonlinearity blocks with the filter sizes as $(3 \times 3 \times \text{channels} \times 32)$, $(3 \times 3 \times 32 \times 64)$, $(3 \times 3 \times 64 \times 128)$, and $(3 \times 3 \times 128 \times 256)$. The size of the initial convolution kernel depends on the group that we are using. For M1, we use four channels, six channels for M2, and two channels for M3. We use the leaky_ReLU(.) function to inject nonlinearity into the designed model. We also use batch normalization after every convolution and dropouts after every pooling layer. Instead of using only the max-pooling layer, we perform max-pooling on half the channels and average pooling on another half of the channel.

This is then followed by adding deconvolution layers and up-sampling. The middle layer that gets created acts as the bottleneck layer that consists of most of the propagated information at a much smaller dimension. We add an auxiliary classifier at this bottleneck layer as used in InceptionNet [59]. The auxiliary classifier is attached to intermediate layers of the network, and it helps in improved convergence during training by combating the vanishing gradient problem. The auxiliary classifier prevents the bottleneck weights from dying out. Besides, two layers of deconvolution layers of the dimensions are similar to the convolution layers. We perform up-sampling before each deconvolution layer. The final layer is followed by two fully connected layers of dimensions 256 and 128. Finally, this is followed by a sigmoid activation function after the last layer to get the multiclass labels. The final loss is a sum of the auxiliary classifier and the final layer multilabel classification loss.

We split the dataset into a 70:30 train:test ratio. To account for the considerable class imbalance and ensure proper training, we made training batches comprising of at least one instance of each class in every batch. This ensures that all the classes have the same contribution to the model in the training process. We initiate the model using Xavier weights. Since we do not have much information about the data, Xavier assigns weights from a Gaussian distribution with zero mean and some finite variance. Xavier weights keep the variance the same in each passing layer, preventing vanishing, or exploding gradients problems. We train the network using the standard back-propagation algorithm using a mini-batch stochastic gradient approach and minimizing a momentum optimizer.

B. Interband Group Retrieval

This section aims to learn a shared latent space equivalently representing all three band groups, M1, M2, and M3. The main

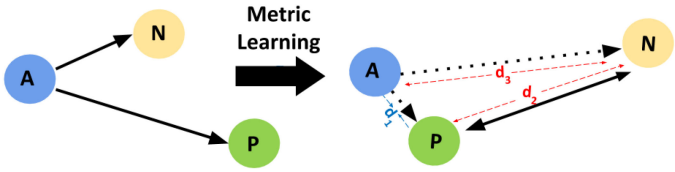

Fig. 4. Overall pipeline of the proposed interband retrieval framework using the modified-triplet function and cross-entropy loss. The inference phase from the synthesized shared latent space is shown in the right.

idea behind designing this unified latent space is to bring the similar class instances of different band groups nearby while pushing diverse class instances of different band groups far apart from each other. This effectively implies that we reduce the intergroup distance and increase the intragroup distances within each class.

To synthesize this common shared space, we design a pipeline, as shown in Fig. 4. We extract the 128-d features from the previously trained network for multilabel, band group-wise classification. We save this final layer feature weights corresponding to each image instance from the corresponding trained classifier. We add a series of fully connected neural network layers from the three pretrained band groups. For this part, we stack three fully connected layers of dimensions 128, 128, and 64.

Modified-cross-triplets loss: The conventional triplet loss is a type of similarity learning loss function. This work uses the basic idea of training a multilabeled triplet loss in conjunction with satellite images from different band groups. We aim to design a domain-agnostic latent space for an interband group retrieval setup. To achieve this, we employ three branches of fully connected networks from the input data stream. Each of these data is derived from the pretrained weights of each instance. We use the proposed modified cross-triplet loss-based network to train this network with learnable parameters represented by θ. We train the network until it reaches a very low loss value of ϵ.

For this purpose, we sample the positive exemplars, by considering image samples from different band groups comprising the same classes. For an image $m_1^n \in M_1$ comprising of N multilabels $c_1, c_2, \ldots, c_n \in C$, any image from $m_2^n \in M_2$ or $m_3^n \in M_s$, having either the same ground-truth labels or any subset of the ground-truth labels $c_1, c_2, \ldots c_n$ is considered as a positive exemplar. An instance which has labels apart from $c_1, c_2, \ldots c_n$ along with a subset of these labels are also considered as a valid positive exemplar.

Conversely, to select the negative examples, an instance from a separate band group, having any labels apart from $c_1, c_2, \ldots, c_n$ is considered a negative example for that image. Therefore, while selecting the negative exemplar for a given training sample, we ensure that $y_i^p \cap y_i^n = \varnothing$. The standard triplet loss is defined as (2), where A denotes the anchor image, P represents the positive exemplar, and N denotes a negative example

$$\mathcal{L}_{(A,P,N)}^{\text{triplet}} = \max(\|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + \alpha, 0). \quad (2)$$

If the distance between the anchor and the positive pair is more than the distance between the anchor and its negative pair, we update the weights by minimizing the loss function. We try to bring the positive sample close to the anchor and push the negative sample farther from the anchor beyond a margin α. In this case, we do not consider the distance between the positive and the negative pairs. Since the data has multiple labels, we modify the loss function to make the learning more robust. The proposed similarity loss is given in (3). Here, the first loss term is similar to the conventional triplet loss function. However, we use the anchor, positive, and negative data instances from three different band groups, contrary to the conventional triplet loss. This term pushes the negative data instance away from the positive and the anchor data beyond a margin of α. The second term is added to push the positive and the negative exemplars apart from each other. This is conditional because only if the classes of the positive and negative exemplars are completely nonoverlapping, this part of the loss is to be considered. To make this conditional, we use an indicator function $\mathbb{K}$

$$\begin{aligned}\mathcal{L}_{(A,P,N)} &= \max(\|f(A) - f(P)\|^2 \\ &\quad - \|f(A) - f(N)\|^2 + \alpha, 0) \\ &\quad + \mathbb{K}(P,N)\max(-\|f(P) - f(N)\| + \beta, 0) \quad (3)\end{aligned}$$

where $\mathbb{K} = \mathbf{1}$, if $P \cap N \neq \{\varnothing\}$, and α and β are the two margin values. To extend this to a interband setup, we choose A, P, and N from different band groups. We train the network with all six combination of triads. The distribution gap between the two band groups are further moved apart by using a decoder network. An illustration of the proposed modified-cross-triplet metric learning is provided in Fig. 5

$$\begin{aligned}\mathcal{L}_{(A,P,N)} &= \max(\|f(A)^{m_i} - f(P)^{m_j}\|^2 \\ &\quad - \|f(A)^{m_i} - f(N)^{m_k}\|^2 + \alpha, 0) \\ &\quad + \mathbb{K}(P,N)\max(-\|f(P)^{m_j} - f(N)^{m_k}\| + \beta, 0). \quad (4)\end{aligned}$$

Selection of positive and negative pairs: We need to feed the network with triads, with inputs from different band groups to train the network. For a multilabeled dataset, it is evident that the number of possible combinations of negative pairs is far more than the number of possible positive pairs. This could lead the network to learn the embedding features with a large intraclass distance in different band groups. To avoid this, it is crucial how we choose the triads during the training process. For this purpose, we feed the network with a similar number of all the six combinations of triads in each mini-batch. The higher euclidean distance classes between their embedding features are farther apart from each other in the shared-embedding space. For example, classes comprising vegetation covers would be farther from classes containing water. Similarly, a few classes are closer to each other in the embedding space, as their mutual euclidean distance is also much smaller. These are classes such as sea, ocean, water bodies, and water courses. These classes are required to fine-tune the boundaries of the representation of instances from each class in the embedding space.

Objective function: In the experiments, we noticed that solely minimizing the triplet loss is insufficient to train the network

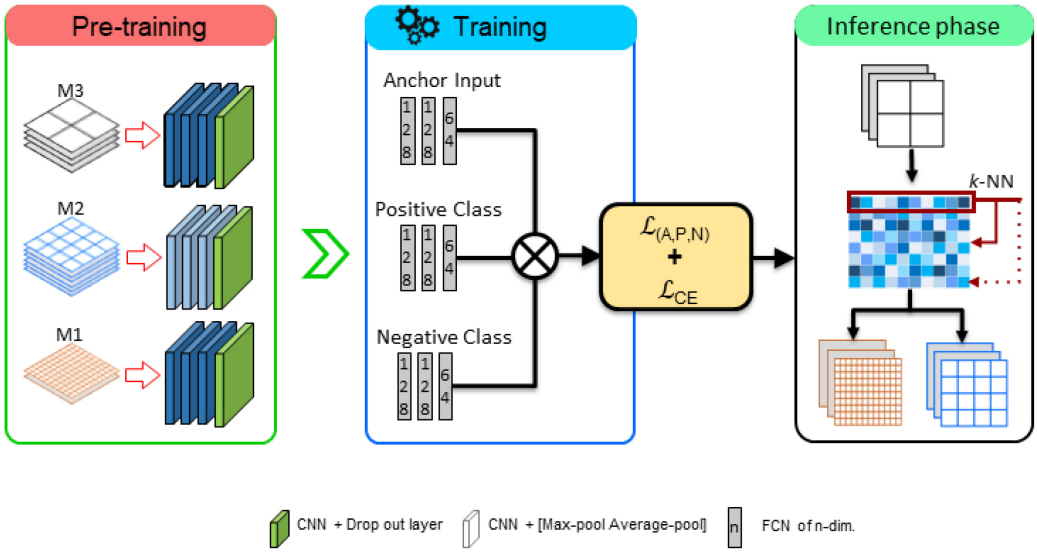

Fig. 5. Modified cross-triplet metric learning.

Algorithm 1: Inter-band data retrieval using the proposed modified cross triplet loss.

Data: $\{M_1, M_2, M_3\}$
Randomly initialize θ.
$m_1^n \in M_1$, $m_2^n \in M_2$, and $m_3^n \in M_3$
Construct sets of A, P, and N separately.
while $\mathcal{L} > \epsilon$ **do**
  $T_s \longleftarrow \{a^c_{m_i}, p^c_{m_j}, n^{c'}_{m_k}\}$; $a \in A$, $p \in P$, $n \in N$;
  $\min_\theta \{\mathcal{L}_{(A,P,N)} + \mathcal{L}_{\text{CE}}\}$;
end
Result: Learnt weights θ;

for such a fine-grained multilabeled dataset. While pushing two different class samples apart, we also need to ensure that the class, which is driven away does get cluttered with some other class. For this purpose, we also add a sigmoid layer to minimize the cross-entropy loss in the network to encode the class information (5). The cross-entropy loss helps maintain differentiating attributes among each class within a group, while the cross-triplet loss aids in bridging the domain gap between the different groups. Algorithm 1 demonstrates the overall working framework

$$\mathcal{L}_{\text{CE}} = \sum_{i=1}^{3} \text{CE}(m_i). \tag{5}$$

The overall loss function is the sum of the cross-triplet loss and the cross-entropy loss functions. We refer to the cumulative loss as $\mathcal{L}$ and define it as (6)

$$\mathcal{L} = \mathcal{L}_{(A,P,N)} + \mathcal{L}_{\text{CE}}. \tag{6}$$

C. Band Group Properties

It is universally acknowledged in RS that different bands of a multispectral sensor are essential in capturing diverse land-use/land-cover regions owing to their absorption characteristics in that band. However, to the best of authors' knowledge, no study has hitherto shown the essential contribution of each of these bands in capturing their band properties. This article uses a modified pairwise-triplet similarity loss-based architecture for interband retrieval while a representation learning network-based architecture for classifying the different groups.

To study the contribution of each band group in the detection of each land-use/land-cover class, we examine the class-wise precision of each land-cover category from the multilabel network trained using each group individually. Finding the accuracy of recognizing each land-cover class using the three band groups would provide us an insight into the properties of the bands in that group. We throw more light to this in the discussion results and discussion Section V.

V. Experimental Results and Analysis

A. Experimental Setup

1) Evaluation Metrics: To evaluate the performance of the multilabel classification for both group-wise and cumulative framework, we use the conventional precision and recall measures. Precision is defined as the proportion of the truly positive to the predicted positives (true positive + false positives). Likewise, recall is defined as the proportion of actual positives that are correctly classified (true positive + false negatives). Conventionally, there is a tradeoff between achieving high precision and a high recall.

2) Parameter Settings: While training the multilabel classification network, we choose the threshold $\alpha = 0.5$. Since there are many classes, we did not intend to set a very high threshold value. For the interband retrieval network, we again chose a value

TABLE II
MULTILABEL CLASSIFICATION PERFORMANCE OF THE PROPOSED FRAMEWORK

Bands	Band group	BigEarthNet		BigEarthNet-S2	
		P@10	Recall	P@10	Recall
All	S-CNN-All [23]	69.93	77.10	-	-
	VGGNet-All	72.76	68.65	74.32	73.78
	Weakly-supervised [64]	36.77	31.51	-	-
	K-Branch CNN [17]	-	-	71.61	78.96
	DenseNet-121 [65]	-	-	79.00	74.60
	Inception-v2 [59]	48.23	56.79	54.38	62.07
	ResNet50-All	74.54	72.20	76.08	77.60
	Regularized ResNet [66]	**85**	77	-	-
	Proposed – All	75.65	**77.98**	**79.32**	**80.54**
RGB	VGGNet-RGB [24]	47.87	49.85	52.54	55.73
	ResNet50-RGB [60]	51.74	50.28	54.50	52.09
	S-CNN-RGB [23]	65.05	75.57	-	-
	Proposed – RGB	58.23	64.95	60.38	67.43
M1	ResNet50–M1	15.32	14.90	15.42	16.00
	VGGNet–M1	14.04	9.65	13.65	12.54
	Proposed – 60 m	14.26	16.30	14.83	17.11
M2	ResNet50–M2	58.54	59.85	58.11	61.78
	VGGNet–M2	55.76	61.86	57.99	64.63
	Proposed – 20 m	57.54	62.57	58.43	64.51
M3	ResNet50–M3	71.21	72.87	71.30	71.38
	VGGNet–M3	70.00	65.87	71.07	69.72
	Proposed – 10 m	71.34	69.13	73.49	72.75

The bold signifies the best performance.

of 0.5 for both the margins α and β. Too high value often leads to dispersed clusters, while too low value does not sufficiently separate two classes.

3) Implementation Details: Following the experimental protocol of [23], we split the entire dataset into 60:20:20 train:val:test split. For all the subproblems, we choose a batch size of 50 for training the network. We select the batches to have at least one image instance of each of the 43 labels in each batch. For training the classification networks, we used a momentum optimizer with a small learning rate of 0.001. We trained the network for about 1000 epochs until the losses converged. We saved the model after every 20 epochs and loaded the best-trained model for the final test. We chose an even lower learning rate of 0.0001 on stochastic gradient descent optimizer for the interband retrieval network. We trained it for about 2000 epochs before saving the best model. We constructed the triplets as mentioned in Section IV-B and ensured that each batch contained at least one anchor image from one of the 43 distinct land-cover classes.

Some of the Sentinel-2 image patches contain a considerable amount of seasonal snow. Also, while most of the images in this dataset are selected from regions with less than 1% could cover, the cloud cover is localized within some patches in some cases. This includes a substantial number of patches (13%). We conducted another set of experiments by eliminating these patches from the dataset and refer to this data as BigEarthNet–subset (S2) subsequently.

Table II shows the comparison of various existing works in the literature on the BigEarthNet dataset and the BigEarthNet–subset datasets. Sumbul *et al.* [57] report their results on two variants of their model, where they train the network using the RGB channels and using all the channels. For our experiments as well, we have reported the performance of both variants. There is a tradeoff between precision and recall values, and maximizing one can lead to a fall in the other. Hence, it is important to maximize precision and recall optimally to attain a high F_1 score. In addition to this, we also compare our results with [23], [24], [59], [60]. Although the work in [23] attains a high recall value, its corresponding precision value falls considerable, taking a toll on their F_1 score. Some existing literary works like [17], [58], [61]–[63] utilize variants/subsets of the BigEarthNet dataset with either different experimental protocol or different aim than classification/retrieval tasks (e.g., colorization, noisy label detection, interband retrieval). Hence, to maintain fairness in comparison, we do not directly compare with their results.

B. Band Group Classification Results

We chose the train-test samples randomly to avoid training bias. Typically for classification problems, it is more common to report the results in terms of accuracy. However, in multilabel classification, there comes ambiguity in categorizing a subset of labels or detecting all the labels and a few more incorrect ones. To avoid this, precision, recall, and mean average precision (mAP) values are considered for multilabel classification. We report the performance of the network in Table II in terms of precision at top-10 (P@10) and recall values. Since the all-channel multilabel classification model outperforms the literary works on this data, we can state that the current network is suitable for group-wise classification without loss of generality. Some of the experiments from the literature have reported their performance on only one of the variants of the BigEarthNet dataset; hence, the alternate variant results are unavailable.

Furthermore, from Table II, we see that the classification performance using just the 20×20 pixel (60 m) band group yields inferior results. This is primarily because there are too few bands in this group. Moreover, the spatial resolution of the bands is very low. This also majorly affects the classification performance of the images. Finally, and most importantly, this group comprises the coastal aerosol and the water vapor band. If we study the land-cover classes in this dataset carefully from Fig. 3, there are a few land-cover categories that we can distinctly recognize using this band group. The classes that gave the highest precision using this band group are burnt areas and continuous/discontinuous urban fabrics.

The classification performance using just the 60×60 pixels/20 m spatial resolution band group yields slightly inferior results to the 120×120 pixels/10 m band group. Even though this group has the most number of bands, i.e., six, the quantity of information contained in these bands seems lesser than 10 m bands. This band group comprises the four vegetation red edge bands and two short wave IR bands. From Fig. 3, it can be seen that there are plenty of vegetation and forest cover classes in the dataset. The M2 band group can classify within these classes much more robustly than the other groups. However, when it comes to the other nonvegetation cover classes, such as airports, salines, burnt areas, to name a few, the classification performance is drastically affected. The classes that gave this

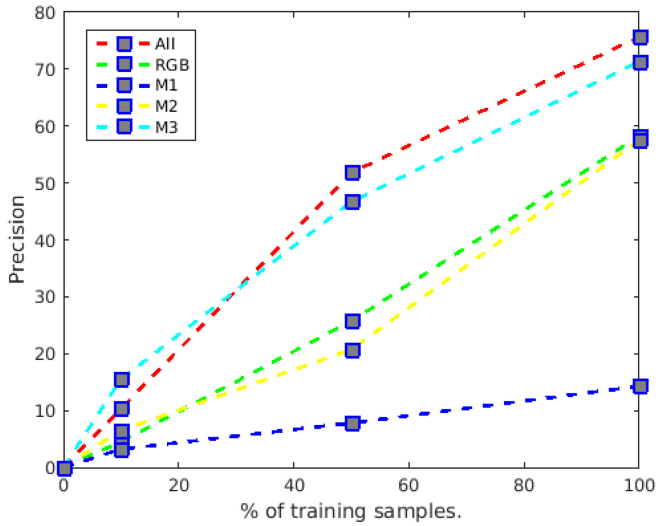

Fig. 6. Model performance with 10%, 50%, and 100% of the training data, reported in terms of precision values.

group the highest precision are coniferous forests, nonirrigated arable lands, and mixed forests. Typically, mixed forests are a combination of coniferous and broad-leaved deciduous forests.

Likewise, M3 was seen to classify the best among the other two. One obvious contributing factor is its high spatial resolution of 10 m (120×120 pixel images). Although this group has fewer bands (four) than M2 (six bands), the effective information content spanning different classes is higher. The intuition is that the overall interclass distance between the broad classes in the embedding space is much higher than the other groups. However, the distinction between the finer classes, such as broad-leaved forests, mixed forests, and coniferous forests, is not high and seems cluttered. This group mainly comprises the RGB and the infrared bands. We can see from Table II that this band group alone can yield more or less comparable performance to that of the full data model. The classes that gave the highest precision using this band group are again coniferous forests and mixed forests. The road and rail networks are also detected well using this band group, which provides crisp, sharp, and detailed images due to its high spatial resolution. Water courses are also detected very well using this group.

Fig. 6 illustrates the model performance with 10%, 50%, and 100% of the training data, obtained by stratified random sampling. We plot the precision values along the vertical axis and plot the model performance with all the RGB channels, M1, M2, and M3 band groups.

C. Cross/Interband Retrieval

In this set of experiments, we aim to realize a shared embedding space for the instances of all three band groups. The shared features are designed to be discriminative while reducing the intergroup domain gap. We do so by minimizing the overall objective function 6. Given a query image from any three groups, we can find the k-nearest neighbours to that query feature from the target band group. Table III reports the results of this interband retrieval in terms of P@10 and mAP values on the BigEarthNet-subset dataset. Table III shows the ablation along with each of these three losses and highlights the advantage of the proposed modified cross-triplet loss for interband data retrieval for multilabeled images.

TABLE III
INTERBAND RETRIEVAL PERFORMANCE OF THE PROPOSED FRAMEWORK

Method	triplet		cross-triplet		modified cross-triplet	
	P@10	mAP	P@10	mAP	P@10	mAP
M1–20 to M2–60	00.05	00.07	04.88	05.02	07.04	07.91
M1–20 to M3–120	00.07	00.08	08.10	07.73	09.83	14.54
M2–60 to M1–20	03.64	04.72	11.42	09.98	12.92	08.43
M2–60 to M3–120	12.86	14.32	45.98	56.87	**63.54**	**62.90**
M3–120 to M1–20	03.87	06.80	12.34	16.72	15.06	21.58
M3–120 to M2–60	15.23	18.49	59.51	63.79	**62.88**	**69.19**

One can see from Table III that the conventional triplet loss is not able to bridge the domain gap between the two band groups. In contrast, the cross-triplet loss can handle retrieval among different bands by bridging the band groups. However, for this multilabeled dataset, the proposed modified cross-triplet loss outperforms the other losses by a margin of almost 2% in most of the results. This helps to highlight the efficacy of the proposed loss.

We observe that the interband retrieval performances between the M2 and M3 band groups are the highest. Intuitively, this is due to the high spatial resolution and information content of the two groups. It is also an important observation that when the query has higher information content (M3), the retrieved instances from a lower information content group (M2) are better than the other way round.

D. Understanding Band Properties

As mentioned in Section IV-C that to study the contribution of each band group for the categorization of each land-use/land-cover class, we examine the class-wise precision of each category. The experiments were conducted on the BigEarthNet-subset dataset devoid of cloud, shadow, and snow cover. The class-wise precision values obtained on the subset data is provided in terms of bar plots in Figs. 7–9. We can see that M3 yields better a class-wise classification performance than M2 and M3. This section briefly discusses the observations by classifying the classes into three categories: agro-forestry, water bodies, and human-made areas. In addition, it can also be seen from the above plots that we have successfully addressed the training data imbalance characteristics. Moreover, the lower number of instance classes perform adequately. *Agro-forestry regions:* We identified 20 classes, which are either of the agricultural or forest type classes. We group them together and thoroughly inspect them in this section. Fig. 7 shows the bar plot of the class-wise precision obtained in each group. Theoretically, the vegetation red edge bands in this group help calculate NDVI and help indicate chlorophyll and, hence, found helpful to distinguish between healthy and unhealthy vegetation. Studies have also shown that the presence of bands 5 and 6 assists in obtaining the biophysical properties of vegetation, such as leaf area index (LAI) and biomass [8]. Band 2 (blue) often finds application

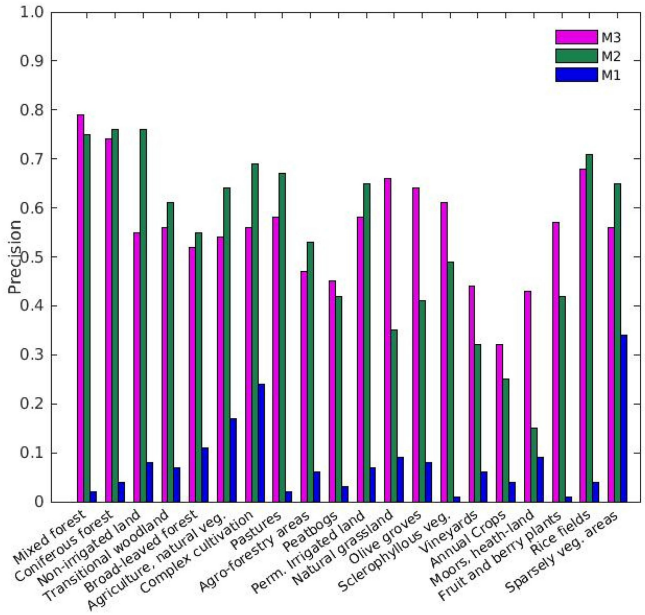

Fig. 7. Class-wise precision of the three groups for agro-forestry classes.

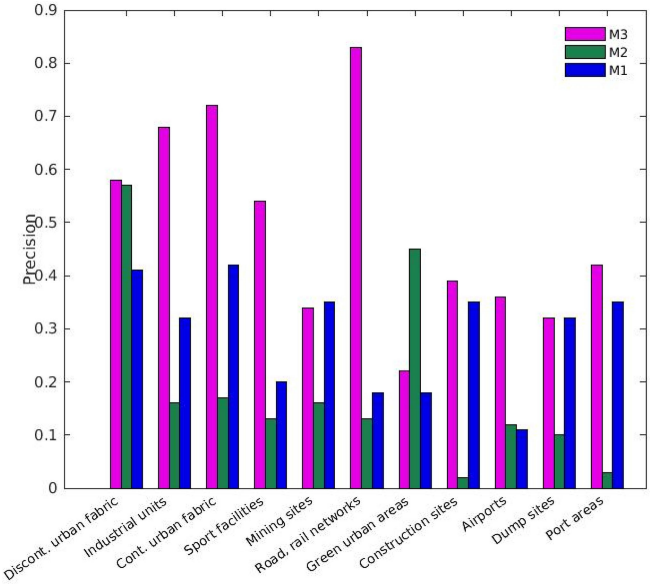

Fig. 9. Class-wise precision of the three band groups for various man-made regions.

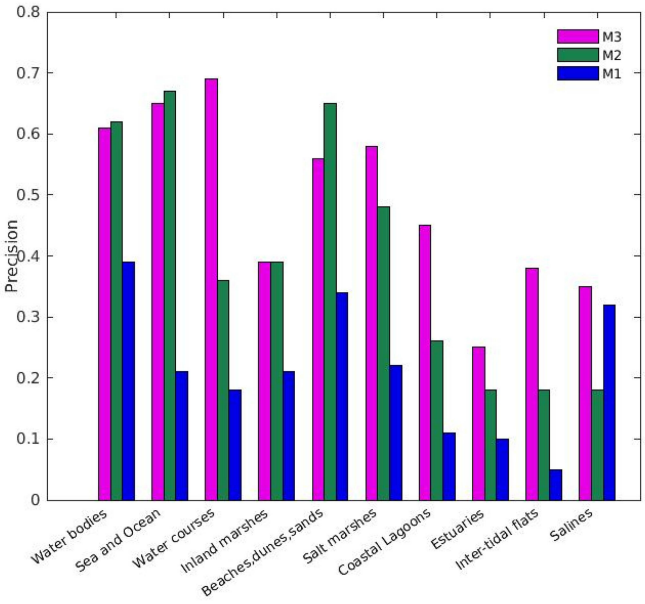

Fig. 8. Class-wise precision of the three band groups for various water-body classes.

in discriminating between coniferous and deciduous forests. Experimentally, we observe the following.

1) Certain vegetation classes seem to be better discriminated in M2 than M1, e.g., coniferous forest, transitional woodland/shrub, broad-leaved forest, etc. Likewise, certain forest classes are also better discriminated in M2 than M1, e.g., nonirrigated arable land, significant natural vegetation, to name a few.
2) The abovementioned observations strongly indicate that the finer-grained agro-forestry classes are more discernible in the feature space of the M2 group than M3. The spectral signature along the six bands helps establish the category of regions, the spatial texture, and geometrical patterns also play an essential role in classifying these classes.
3) It is observed that in M2, most of the vegetated surfaces are retrieved with a precision $\geq 60\%$. This observation also partially confirms that most of these vegetation covers have proper irrigation and drainage facilities in healthy conditions.
4) The relatively lower performance of the agriculture and natural vegetation class could be because of its confusion with the complex cultivation pattern class. During the time of data acquisition, possibly both these regions had cultivated crops. However, their spatial microtextures and geometrical patterns help to quite an extent in making them somewhat discernible.
5) Mixed and coniferous forests show very high accuracy in both M2 and M3. While the primary reason behind this is the significant number of training samples available for these classes, the unique leaf structure of their respective categories also makes them discernible. The coniferous forest predominantly comprises cone-bearing needle-leaved evergreen trees. Mixed forest, on the other hand, is a combination of coniferous and broad-leaved deciduous forests. This property degrades the capacity to discern broad-leaved forests from mixed forests as there is ample confusion between the two. Common factors affect the broad-leaved forest class as it has considerably lower training samples than the former.

Water bodies: For this category, we select ten classes that broadly consist of water. We group them together and thoroughly inspect them in this section. Fig. 8 shows the bar plot of the class-wise precision obtained in each band group. Theoretically, band 2 (categorized under M3) is in the visible blue region and finds several applications in bathymetric mappings,

which involve the study of underwater depth of ocean floors or lake floors [67]. While M3 seems superior in classifying these classes, a few red edge vegetation bands (under M2) are sensitive to moisture content and helps in calculation the moisture index. Experimentally, we observe the following.

1) It is majorly expected that one can detect water bodies better with the M3 group. This view is supported theoretically by the high reflectance of the water body in the blue wavelength region and near-complete absorption of the other wavelengths. Although most of the classes follow this trend, there are a few classes where M2 performs better than M3.
2) The differences in precision values obtained using M2 in certain water body classes is assumed to be caused due to *coastal algal bloom* [68] and the presence of low to moderately dense chlorophyll content due to *phytoplankton* population [69]. The presence of phytoplankton is responsible for an increase in the reflectance in the red edge region, making the spectral signature of the class different from regular water bodies.
3) An interesting sub-class under the water-bodies is coastal lagoons. The coastal lagoons are essentially transitional zones between land and sea. They are shallow inland water bodies that are intermittently connected to oceans, blocked by land barriers. Therefore, they contain the spectral signatures of both land and water. The precision of this class is observed to be relatively lower than the other classes. This observation is assumed because of its confusion with the individual irrigated lands and water bodies. Besides, there also exists the dilemma of turbidity and *eutrophication* of water, which is a very common ecological phenomenon [70].
4) Estuaries, Intertidal flats, and Salines comprise too few training samples, and hence their performance takes a hit due to the largely imbalanced data classes.

Man-made regions: For this category, we select 11 classes that broadly consist of urban areas. We group them together and carefully inspect them in this section. Fig. 9 shows the bar plot of the class-wise precision captured in each band group. Human-made areas are assumed to be captured well by the M3 group as it has a very high spatial resolution 10 m [71]. Experimentally, we observe the following.

1) Human-made structures can be captured well by high spatial resolution bands [71]. This nature helps the M3 band group of spatial resolution 10 m in providing superior results.
2) In certain regions like the port areas and construction sites, the greenness value is close to 0. Hence, its corresponding reflectance in the red edge bands is also close to 0. Due to this reason, M2, comprising of red edge bands, cannot identify these regions successfully.
3) M1 consists of the coastal aerosol and the water vapor bands. The presence of specific disposed or waste material causes the aerosol bands to help detect dump sites. It is also our assumption that the port areas and construction sites from which the dataset was obtained consisted of specific amounts of fly-ashes and aerosols [72], which lead to a better assessment of these classes in this group.
4) Surprisingly, continuous and discontinuous urban fabrics were also captured well by this band group. This could be because these urban areas were high on particle pollutants [73].

Overall, it is observed that M3 has lesser class-wise precision variation than M2, which has a high sensitivity to the agricultural and forest areas. It is much lower for urban, road, and different human-made targets. We can observe from the plots that the precision mainly falls over nonvegetated areas (in M2). The literature affirms that the bands B5, B6, B7, B8A are sensitive to the *greenness* of vegetation, while B9 and B11 are sensitive to the moisture content, and B12 is excellent at detecting geological features. Hence, in certain human-made classes such as industrial, sports, leisure facilities, road, rail networks, bare rock, mineral extraction sites, construction sites, airports, dump-sites, port areas, and burnt areas, the amount of greenness is little to almost negligible. This is because of their nearly nonexisting vegetation canopy. These areas have intrinsic sharp patterns and are captured well in the M3 group with high spatial resolution. Therefore, while M2 classifies the fine-grained vegetation and forest classes, M3 classifies the overall broad spectrum classes more robustly, such as airports, salt marshes, rail and road networks, bare rock, and coastal lagoons.

There exists a substantial class-wise data imbalance in the dataset. The class-wise precision subplots are arranged to decrease the number of samples present from left to right. It can be noted from Figs. 7–9 that the classes having fewer samples gradually show decreasing results as one could not learn them adequately. Also, another critical contributory factor in the multilabel classification performance is that certain classes appear in combinations throughout the dataset. Hence, the classification performance of at least one of the distinct subclasses of the multilabeled instances often ensures the classification of the other classes in that instance. This phenomenon affects the performance of the model immensely in conventional multilabel classification.

VI. Conclusion

We exploit a simplistic yet efficient representation learning network for multilabel classification that yields superior results to the existing literature. We then group the bands of Sentinel-2 multispectral data based on their spatial resolutions. The effectiveness of any classification or retrieval problem primarily depends on the texture information in the data and the existence of pure end-member within a pixel. With an increase in the spatial resolution, the possibility of detecting a pure end-member in a pixel reduces due to the high chance of nonlinear mixing of several other end-members within the resolution cell.

From these grouped bands, we demonstrate the identifiability of each of the land-cover classes in all these band groups. We further interpret the observations from the abovementioned framework and study the band properties of this multispectral RS data. The BigEarthNet dataset was created by exploiting the labels from the Corine land-use/land-cover classes. This results in the presence of mixed classes in the patches. Our experiments have supported this observation and are discussed in detail by exploiting agriculture and forestry domain knowledge. In addition, it can also be seen from the abovementioned results

that we have successfully addressed the training data imbalance part, and even the lower number of instance classes perform decently well.

Furthermore, we introduced a novel modified cross-triplet loss for multilabeled data for metric learning and established its efficacy over the conventional triplet loss. The standard triplets do not consider the possibility of having a common subset of classes between the positive and negative examples of a multilabeled anchor image and, therefore, can spread apart the intraclass distances. In the proposed loss, we consider this fact and observe a distinct improvement in the overall results.

We thoroughly demonstrate our classification and retrieval results on the large-scale benchmark BigEarthNet data. We show that the proposed framework outperforms the current literature in all the evaluation metrics to validate our claim. In the future, we would like to extend this study to investigate the effect of a weighted grouping of bands and a learnable band selection process. We also plan to perform a few case studies using these specific band groups from forest-fire and snow/cloud detection problems and study their efficacy.

ACKNOWLEDGMENT

Ushasi Chaudhuri would like to thank the Prof. M. Crucianu and Prof. M. Ferecatu for their guidance in improving the quality of this article.

REFERENCES

[1] N. Li *et al.*, "Multiparameter optimization for mineral mapping using hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1348–1357, Apr. 2018.

[2] Y. Zhang, W.-K. Huen, and P. O. Ang, "The hyper-spectral characteristics of coral species and habitats in Hong Kong," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 746–753, Apr. 2013.

[3] A. S. Sahadevan, P. Shrivastava, B. S. Das, and M. Sarathjith, "Discrete wavelet transform approach for the estimation of crop residue mass from spectral reflectance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2490–2495, Jun. 2014.

[4] G. Zhang, P. Ghamisi, and X. X. Zhu, "Fusion of heterogeneous earth observation data for the classification of local climate zones," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7623–7642, Oct. 2019.

[5] F. Van der Meer, H. Van der Werff, and F. Van Ruitenbeek, "Potential of ESA's Sentinel-2 for geological applications," *Remote Sens. Environ.*, vol. 148, pp. 124–133, 2014.

[6] J. Wu, "Introduction to convolutional neural networks," Nat. Key Lab. for Novel Softw. Technol., Nanjing Univ., 2017, vol. 5, pp. 978–973.

[7] H. H. Aghdam and E. J. Heravi, *Guide to Convolutional Neural Networks*, vol. 10. New York, NY, USA: Springer, 2017, pp. 978–973.

[8] J. Delegido, J. Verrelst, L. Alonso, and J. Moreno, "Evaluation of sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content," *Sensors*, vol. 11, no. 7, pp. 7063–7081, 2011.

[9] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for VHR remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4775–4784, Aug. 2017.

[10] U. Chaudhuri, S. Chaudhuri, and S. Chaudhuri, "Gucnet: A guided clustering-based network for improved classification," in *Proc. IEEE 25th Int. Conf. Pattern Recognit.*, 2021, pp. 7335–7342.

[11] G.-S. Xia *et al.*, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[12] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land-cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 549–553, Apr. 2017.

[13] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

[14] U. Chaudhuri, B. Banerjee, A. Bhattacharya, and M. Datcu, "Cmir-Net: A deep learning based model for cross-modal retrieval in remote sensing," *Pattern Recognit. Lett.*, vol. 131, pp. 456–462, 2020.

[15] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "Sen12ms—A curated dataset of georeferenced multi-spectral sentinel-1/2 imagery for deep learning and data fusion," 2019, *arXiv:1906.07789*.

[16] 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.*, to be published, doi: 10.1109/LGRS.2021.3056392.

[17] G. Sumbul and B. Demİr, "A deep multi-attention driven approach for multi-label remote sensing image classification," *IEEE Access*, vol. 8, pp. 95 934–95 946, 2020.

[18] U. Chaudhuri, B. Banerjee, A. Bhattacharya, and M. Datcu, "Crossatnet—A novel cross-attention based framework for sketch-based image retrieval," *Image Vis. Comput.*, vol. 104, 2020, Art. no. 104003.

[19] 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.

[20] U. Chaudhuri, B. Banerjee, A. Bhattacharya, and M. Datcu, "Attention-driven graph convolution network for remote sensing image retrieval," *IEEE Geosci. Remote Sens. Lett.*, to be published, doi: 10.1109/LGRS.2021.3105448.

[21] J. Zhao, M. Datcu, Z. Zhang, H. Xiong, and W. Yu, "Contrastive-regulated CNN in the complex domain: A method to learn physical scattering signatures from flexible polsar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10 116–10 135, Dec. 2019.

[22] Z. Huang, M. Datcu, Z. Pan, and B. Lei, "Deep Sar-Net: Learning objects from signals," *ISPRS J. Photogrammetry Remote Sens.*, vol. 161, pp. 179–193, 2020.

[23] G. Sumbul, M. Charfuelan, B. Demir, and V. Markl, "Bigearthnet: A large-scale benchmark archive for remote sensing image understanding," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5901–5904.

[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[25] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[26] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[27] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, "Hyperspectral image classification with attention-aided CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2281–2293, Mar. 2021.

[28] P. Duan, P. Ghamisi, X. Kang, B. Rasti, S. Li, and R. Gloaguen, "Fusion of dual spatial information for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7726–7738, Sep. 2021.

[29] R. Bahmanyar, D. Espinoza-Molina, and M. Datcu, "Multisensor earth observation image classification based on a multimodal latent dirichlet allocation model," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 459–463, Mar. 2018.

[30] S. Cui, G. Schwarz, and M. Datcu, "Remote sensing image classification: No features, no clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 11, pp. 5158–5170, Nov. 2015.

[31] R. Tănase, R. Bahmanyar, G. Schwarz, and M. Datcu, "Discovery of semantic relationships in Polsar images using latent Dirichlet allocation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 237–241, Feb. 2017.

[32] Z. Xu, K. Guan, N. Casler, B. Peng, and S. Wang, "A 3D convolutional neural network method for land cover classification using lidar and multi-temporal landsat imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 144, pp. 423–434, 2018.

[33] R. Fan, R. Feng, L. Wang, J. Yan, and X. Zhang, "Semi-MCNN: A semisupervised multi-CNN ensemble learning method for urban land cover classification using submeter HRRS images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4973–4987, 2020.

[34] W. Li, G. Wu, F. Zhang, and Q. Du, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.

[35] S.-W. Chen and C.-S. Tao, "Polsar image classification using polarimetric-feature-driven deep convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 627–631, Apr. 2018.

[36] S. Ji, C. Zhang, A. Xu, Y. Shi, and Y. Duan, "3D convolutional neural networks for crop classification with multi-temporal remote sensing images," *Remote Sens.*, vol. 10, no. 1, p. 75, 2018.
[37] J. Jiang, X. Feng, F. Liu, Y. Xu, and H. Huang, "Multi-spectral RGB-NIR image classification using double-channel CNN," *IEEE Access*, vol. 7, pp. 20 607–20 613, 2019.
[38] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, pp. 430–443, 2019.
[39] J. Lee, D. Han, M. Shin, J. Im, J. Lee, and L. J. Quackenbush, "Different spectral domain transformation for land cover classification using convolutional neural networks with multi-temporal satellite imagery," *Remote Sens.*, vol. 12, no. 7, p. 1097, 2020.
[40] M. Ferecatu, M. Crucianu, and N. Boujemaa, "Retrieval of difficult image classes using SVD-based relevance feedback," in *Proc. 6th ACM SIGMM Int. Workshop Multimedia Inf. Retrieval*, 2004, pp. 23–30.
[41] B. Demir and L. Bruzzone, "Hashing-based scalable remote sensing image search and retrieval in large archives," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 892–904, Feb. 2016.
[42] U. Chaudhuri, B. Banerjee, and A. Bhattacharya, "Siamese graph convolutional network for content based remote sensing image retrieval," *Comput. Vis. Image Understanding*, vol. 184, pp. 22–30, 2019.
[43] N. Khan, U. Chaudhuri, B. Banerjee, and S. Chaudhuri, "Graph convolutional network for multi-label VHR remote sensing scene recognition," *Neurocomputing*, vol. 357, pp. 36–46, 2019.
[44] G.-S. Xia, X.-Y. Tong, F. Hu, Y. Zhong, M. Datcu, and L. Zhang, "Exploiting deep features for remote sensing image retrieval: A systematic investigation," 2017, *arXiv:1707.073 21*.
[45] U. Chaudhuri, B. Banerjee, A. Bhattacharya, and M. Datcu, "A simplified framework for zero-shot cross-modal sketch data retrieval," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 182–183.
[46] D. Pflugmacher, A. Rabe, M. Peters, and P. Hostert, "Mapping pan-european land cover using landsat spectral-temporal metrics and the european LUCAS survey," *Remote Sens. Environ.*, vol. 221, pp. 583–595, 2019.
[47] E. E. Berman, D. K. Bolton, N. C. Coops, Z. K. Mityok, G. B. Stenhouse, and R. D. Moore, "Daily estimates of landsat fractional snow cover driven by modis and dynamic time-warping," *Remote Sens. Environ.*, vol. 216, pp. 635–646, 2018.
[48] C. Robinson *et al.*, "Large scale high-resolution land cover mapping with multi-resolution data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12726–12735.
[49] D. Chen*, D. Stow, and P. Gong, "Examining the effect of spatial resolution and texture window size on classification accuracy: An urban environment case," *Int. J. Remote Sens.*, vol. 25, no. 11, pp. 2177–2192, 2004.
[50] J. R. Irons *et al.*, "The effects of spatial resolution on the classification of thematic mapper data," *Int. J. Remote Sens.*, vol. 6, no. 8, pp. 1385–1403, 1985.
[51] T. Dutta, A. Singh, and S. Biswas, "Adaptive margin diversity regularizer for handling data imbalance in zero-shot sbir," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 349–364.
[52] J. Zhang *et al.*, "Generative domain-migration hashing for sketch-to-image retrieval," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 297–314.
[53] R. Barandela, E. Rangel, J. S. Sánchez, and F. J. Ferri, "Restricted decontamination for the imbalanced training sample problem," in *Iberoamerican Congress on Pattern Recognition*. Berlin, Germany: Springer, 2003, pp. 424–431.
[54] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
[55] Y. Xian, T. Lorenz, B. Schiele, and Z. Akata, "Feature generating networks for zero-shot learning," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2018, pp. 5542–5551.
[56] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.
[57] G. Sumbul and B. Demir, "A novel multi-attention driven system for multi-label remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5726–5729.
[58] G. Sumbul *et al.*, "BigEarthNet dataset with a new class-nomenclature for remote sensing image understanding," 2021, *arXiv:2001.06372*.
[59] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2016, pp. 2818–2826.
[60] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, 2016, pp. 770–778.
[61] P. Ulmas and I. Liiv, "Segmentation of satellite imagery using u-net models for land cover classification," 2020, *arXiv:2003.02899*.
[62] S. Vincenzi *et al.*, "The color out of space: learning self-supervised representations for earth observation imagery," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, 2020, pp. 3034–3041.
[63] H. Yessou, G. Sumbul, and B. Demir, "A comparative study of deep learning loss functions for multi-label remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2020, pp. 1349–1352.
[64] C. Greenwell, S. Workman, and N. Jacobs, "Implicit land use mapping using social media imagery," in *Proc. IEEE Appl. Imagery Pattern Recognit. Workshop*, 2019, pp. 1–4.
[65] I. Kakogeorgiou and K. Karantzalos, "Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing," 2021, *arXiv:2104.01375*.
[66] M. Zotov and J. Gamper, "Conditional denoising of remote sensing imagery using cycle-consistent deep generative models," 2019, *arXiv:1910.14567*.
[67] I. Caballero and R. P. Stumpf, "Retrieval of nearshore bathymetry from Sentinel-2A and 2B satellites in South Florida coastal waters," *Estuarine, Coastal Shelf Sci.*, vol. 226, 2019, Art. no. 106277.
[68] V. Klemas, "Remote sensing of algal blooms: an overview with case studies," *J. Coastal Res.*, vol. 28, no. 1 A, pp. 34–43, 2012.
[69] D. Blondeau-Patissier, J. F. Gower, A. G. Dekker, S. R. Phinn, and V. E. Brando, "A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans," *Prog. Oceanogr.*, vol. 123, pp. 123–144, 2014.
[70] M.-T. Sebastiá-Frasquet, J. A. Aguilar-Maldonado, E. Santamaría-Del-Ángel, and J. Estornell, "Sentinel 2 analysis of turbidity patterns in a coastal lagoon," *Remote Sens.*, vol. 11, no. 24, p. 2926, 2019.
[71] X. Yang, *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment*. Hoboken, NJ, USA: Wiley, 2011.
[72] S. Jaremenko, K. Garmonov, and R. Sheps, "Research of air pollution by dust aerosols during construction," in *Proc. IOP Conf. Series, Mater. Sci. Eng., Int. Conf. Construction, Architecture Techno sphere Saf.*, 2017, vol. 262, no. 1, p. 012189.
[73] S. Janhäll, "Review on urban vegetation and particle air pollution-deposition and dispersion," *Atmospheric Environ.*, vol. 105, pp. 130–137, 2015.

Ushasi Chaudhuri (Student Member, IEEE) received the bachelor's degree in electronics and telecommunication engineering from Mumbai University, Mumbai, India, and the master's degree in visual computing from the Indian Institute of Technology (IIT) Kharagpur, Kharagpur, India, with her research thesis on the analysis of document images based on rough set reducts, in 2018. She is currently working toward the Ph.D. degree visual computing and machine intelligence in remote sensing with the Centre of Studies in Resources Engineering (CSRE), IIT Bombay, Mumbai, India.

Her current research is focused on content-based retrieval of remote sensing (RS) data using graph convolution networks. She is also working on cross-modal retrieval of RS and multimedia data with label scarcity.

Subhadip Dey (Student Member, IEEE) received the B.Tech. degree in agricultural engineering from Bidhan Chandra Krishi Viswavidyalaya, Haringhata, India, in 2016, and the M.Tech. degree in aquacultural engineering from the Department of Agricultural and Food Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India, 2018. He is currently working toward the Ph.D. degree in agricultural applications in remote sensing with the Microwave Remote Sensing Lab, Centre of Studies in Resources Engineering (CSRE), Indian Institute of Technology Bombay, Mumbai, India.

His current research interests are land cover classification and agricultural crop mapping and monitoring using synthetic aperture radar data.

Mihai Datcu (Fellow, IEEE) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica Bucharest (UPB), Bucharest, Romania, in 1978 and 1986, respectively, and the habilitation a diriger des recherches degree in computer science from the University Louis Pasteur, Strasbourg, France, in 1999.

From 1992 to 2002, he had a longer Invited Professor assignment with the Swiss Federal Institute of Technology, ETH Zurich, Zürich, Switzerland. Since 1981, he has been a Professor with the Department of Applied Electronics and Information Engineering, Faculty of Electronics, Telecommunications, and Information Technology, UPB, working on image processing and electronic speckle interferometry. Since 1993, he has also been a Scientist with German Aerospace Center, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen, Germany. He is developing algorithms for model-based information retrieval from high-complexity signals and methods for scene understanding from very high resolution synthetic aperture radar (SAR) and interferometric SAR data. He is currently a Senior Scientist and Image Analysis Research Group Leader with Remote Sensing Technology Institute, DLR. Since 2011, he has been leading the Immersive Visual Information Mining Research Lab, Munich Aerospace Faculty, and has been the Director of the Research Center for Spatial Information, UPB. Since 2001, he has been initiating and leading the Competence Centre on Information Extraction and Image Understanding for Earth Observation, ParisTech, Paris Institute of Technology, Telecom Paris, Paris, France, a collaboration of DLR with the French Space Agency (CNES). He has been a Professor with the DLR-CNES Chair, ParisTech, Paris Institute of Technology, Telecom Paris. He initiated the European frame of projects for Image Information Mining (IIM) and is involved in research programs for information extraction, data mining and knowledge discovery, and data understanding with the European Space Agency (ESA), NASA, and in a variety of national and European projects. He and his team have developed and are currently developing the operational IIM processor in the Payload Ground Segment systems for the German missions TerraSAR-X, TanDEM-X, and the ESA Sentinel-1 and -2. He has authored more than 450 scientific publications, which include 80 journal papers, and a book on number theory. He is involved in research related to information theoretical aspects and semantic representations in advanced communication systems. His research interests include Bayesian inference, information and complexity theory, stochastic processes, model-based scene understanding, image information mining, for applications in information retrieval, and understanding of high-resolution SAR, and optical observations.

Dr. Datcu is a Member of the ESA's ϕ lab and Big Data from Space (BiDS).He was the recipient of the IEEE Geoscience and Remote Sensing Society Prize Best Paper Award in 2006, the National Order of Merit with the rank of Knight, for outstanding international research results, awarded by the President of Romania, in 2008, and the Romanian Academy Prize Traian Vuia, for the development of the SAADI image analysis system and his activity in image processing, in 1987. He was a Guest Editor for a Special Issue on IIM of the IEEE and other journals.

Biplab Banerjee received the M.E. degree in computer science and engineering from Jadavpur University, Kolkata, India, in 2010, and the Ph.D. degree in image analysis from the Indian Institute of Technology Bombay (IIT Bombay), Mumbai, India, in 2015.

He was a Postdoctoral Researcher with the University of Caen Basse-Normandy, Caen, France, and the Istituto Italiano di Technologia Genova, Genoa, Italy. He then was an Assistant Professor with the Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Roorkee, India, between 2016 and 2018. He holds a visiting Professor tenures with Technische Universität München, Munich, Germany, Ghent University, Ghent, Belgium, Kyungpook National University, Daegu, South Korea, to name a few. He is currently an Assistant Professor (since 2018) of Machine Learning and Visual Computing with the Centre of Studies in Resources Engineering (CSRE), and the Center of Machine Intelligence and Data Science (MInDS), IIT Bombay. He is closely associated with the Vision and Image Processing Group at the Department of Electrical Engineering, IIT Bombay. Besides, he is currently an Engineering Advisor of AI for AWL Inc., Tokyo, Japan. His research topics include zero-shot learning, meta learning, multitask learning, domain adaptation and transfer learning, multimodal analysis of remote sensing data, deep reinforcement learning, etc.

Dr. Banerjee was the recipient of the Excellence in Ph.D. Thesis Award for his Ph.D. thesis from the IIT Bombay

Avik Bhattacharya (Senior Member, IEEE) received the integrated M.Sc. degree in mathematics from the Indian Institute of Technology, Kharagpur, India, in 2000 and the Ph.D. degree in remote sensing image processing and analysis from Télécom ParisTech, Paris, France, and the Ariana Research Group, Institut National de Recherche en Informatique et en Automatique (INRIA), Sophia Antipolis, Nice, France, in 2007.

He is currently an Associate Professor at the Centre of Studies in Resources Engineering (CSRE), Indian Institute of Technology Bombay (IITB), Mumbai, India. Before joining IITB, he was a Canadian Government Research Fellow with the Canadian Centre for Remote Sensing (CCRS), Ottawa, ON, Canada. His current research interests include SAR polarimetry, statistical analysis of polarimetric SAR images, applications of Radar Remote Sensing in Agriculture, Cryosphere, Urban and Planetary studies.

Dr. Bhattacharya is the Editor-in-Chief of IEEE Geoscience and Remote Sensing Letters. He was an Associate Editor of IEEE GRSL. He has been the Guest Editor of the special issue on Applied Earth Observations and Remote Sensing in India in the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS), 2017. He was one of the guest editors of the special stream on Advanced Statistical Techniques in SAR Image Processing and Analysis in IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, 2018. He is the Founding Chairperson of the IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY (GRSS) Chapter of the Bombay Section. He is currently leading the Microwave Remote Sensing Lab at CSRE, IITB. He received the Natural Sciences and Engineering Research Council of Canada visiting scientist fellowship at the Canadian national laboratories from 2008 to 2011.