Guided Unsupervised Learning by Subaperture Decomposition for Ocean SAR Image Retrieval

Nicolae-Cătălin Ristea, Andrei Anghel, Senior Member, IEEE, Mihai Datcu, Fellow, IEEE, and Bertrand Chapron

Abstract—A spaceborne synthetic aperture radar (SAR) can provide accurate images of the ocean surface roughness day-or-night in nearly all-weather conditions, being a unique asset for many geophysical applications. Considering the huge amount of data daily acquired by satellites, automated techniques for physical features extraction are needed. Even if supervised deep learning methods attain state-of-the-art results, they require a great amount of labeled data, which are difficult and excessively expensive to acquire for ocean SAR imagery. To this end, we use the subaperture decomposition (SD) algorithm to enhance the unsupervised learning retrieval on the ocean surface, empowering ocean researchers to search into large ocean databases. We empirically prove that SD improves the retrieval precision with over 20% for an unsupervised transformer autoencoder network. Moreover, we show that SD brings an important performance boost when Doppler centroid images are used as input data, leading the way to new unsupervised physics-guided retrieval algorithms.

Index Terms—Doppler centroid estimation (DCE), image retrieval, ocean imagery, remote sensing (RS), subapertures decomposition, synthetic aperture radar (SAR), unsupervised learning.

I. INTRODUCTION

EARTH observation (EO) is the integration of information about planet Earth’s physical, chemical, and biological system by remote sensing (RS) technologies provided by earth surveillance techniques, including the collection analysis and presentation of data [1]. Considering that the ocean accounts for about 71% of Earth’s surface, the ocean observation increasingly draws the attention of the research community over the last decades. Humans had minimal ocean observations before 1978, when Seasat, the first Earth-orbiting satellite designed for RS of Earth’s oceans was launched [2]. Although Seasat only operated for about 100 days, the mission acquired more data about the ocean than all previous sensors combined. This event stimulated the fast development of ocean satellite, leading to a growing number of satellites carrying different sensors (e.g., microwave, visible, and infrared) being launched to improve our understanding about the ocean.

Nowadays, one of the most used spaceborne sensors for ocean observation is the synthetic aperture radar (SAR), used by satellite mission Sentinel-1 from 2014, when the WV mode was implemented. The WV modality is available only on the Sentinel-1A/B and is dedicated for retrieving ocean surface properties at global scale [3]. The WV measurements have a spatial resolution of approximately 4 m and a scene footprint of 20 by 20 km. These sensors collect monthly nearly 120 000 WV vignettes of the global ocean surface. Moreover, tens of satellites have also been approved for the next 20 years, conducting to a sharp rise of ocean data. Hence, automated systems designed to interpret, extract, and find features in big data environments are highly needed to exploit all the available information.

An important aspect of ocean big data is that having more data does not guarantee more valuable information extracted. Usually, the key information is sparsely hidden in massive ocean-satellite data. Once with the growing capacity of collecting ocean data, many efforts have been put into developing and validating retrieval algorithms to generate standard time-series global ocean parameters [1], [4], [5], [6], [7], [8]. Currently, one important focus is to develop efficient and intelligent approaches to improve the information extraction capability with powerful deep learning algorithms. Because the physical phenomena, which can occur on the ocean, are diverse, ranging from waves and algal blooms, which are locally generated and their signatures only consist of a tiny percentage of an ocean vignette, to long time-series data (e.g., level of ocean), new data-driven information mining algorithms are required. Moreover, extracting real-time information from high-rate downlink satellite data streams requires high-speed data processing. Deep learning techniques can satisfy all mentioned requirements, proving high efficiency and generalization capacity in image-related tasks [9], [10].

Another major aspect of ocean big data is the costly process of annotating data. Considering the particularities of RS ocean data, only people with expertise can annotate vignettes (e.g., ocean currents direction, waves height, and ocean phenomena), making the process time-consuming and costly. Inspired by

Authorized licensed use limited to: 1558-0644. Downloaded on April 07, 2023 at 17:05:43 UTC from IEEE Xplore. Restrictions apply.
the visual data domain, several works leverage unsupervised information to learn deep representations [1], [11], [12], [13]. Nevertheless, even if there is a moderate success on SAR un-
supervised image retrieval, there are no works, which studies the benefits of unsupervised deep learning (UDL) for ocean SAR image retrieval.

In our paper, we extend the previous work from [14] and address the unsupervised ocean image retrieval task by combining the subaperture decomposition (SD) algorithm with UDL. Using UDL for image retrieval, we exclude the necessity of labeled data, and combining it with the preprocessing algorithm based on SD, we pushed the retrieval accuracy closer to the supervised learning approaches. Moreover, we consid-
ered the case when other than amplitude SAR data are fed into the neural network. Therefore, we tested the capacity of our model to handle physics-guided RS algorithms (providing as input the Doppler centroid images of the vignettes), showing promising retrieval results. The retrieval performance is even further improved when the model is trained with Doppler centroid images estimated on the subapertures, rather than the original SAR image. All the results show that the SD algorithm improves the performance for image retrieval, even if other processing techniques are applied in top of it. Using those processing techniques, we developed an efficient algorithm of query by image, meant to help experts to identify similar phenomena on the ocean surface. Each vignette is described by an embedding vector computed with a pretrained deep neural network (DNN), trained in an unsupervised manner. Moreover, we extended the use case of query by image to a more complex approach of query by physical parameters. More exactly, we estimated the Doppler centroid images of the subaperture single-look complex (SLC) vignettes and used them as inputs of the DNN. In this case, each vignette is described by an embedding vector, taken from a DNN pretrained on Doppler centroid images estimated on subapertures.

In summary, with respect to our previous work [14], our current contribution is twofold.

1) We are the first who proposed an unsupervised query by example framework for ocean SAR imagery.

2) We pushed the results of unsupervised ocean image retrieval close to the supervised methods, reducing the needs of labeled data.

3) We combined the previous SD algorithm with Doppler centroid estimation (DCE) and obtained superior results for classification (when we compare DCE against SD with DCE) and image retrieval. By employing DCE, we lead the way to more complex retrieval algorithms, which can find similar physical phenomena on the ocean.

II. RELATED WORK

This section makes a state-of-the-art analysis relative to the proposed methodology and covers the following topics: SD in SAR imagery, DCE methods, transformer models, and image retrieval techniques.

A. Subaperture Decomposition

The SD algorithm is widely used for SAR imagery [14], [15], [16], [17], [18], [19]. The method was combined with both classical signal processing algorithms [16], [17], [18] and deep learning techniques [14], [15]. In [16], the SD is proposed for the ship detection task, while in [17], it is used for target characterization. Moreover, the SD algorithm was used to translate a single channel SAR image into three channels image alike representation, by decomposing it into three sub-bands. Afterward, Wang et al. [15] used pretrained DNN for target classification task on the ground.

Distinctly, we propose to extend the SD usage from our previous work [14] by combining the SD with unsupervised learning to improve the ocean image retrieval. To the best of our knowledge, we are the first who use SD in an unsupervised deep retrieval algorithm. Moreover, we use the SD to improve the classification and unsupervised retrieval for the DCE algorithm.

B. Doppler Centroids Estimation

For decades, Doppler centroids are used in processing SAR data [20], [21]. Many works have been proposed to improve the Doppler estimation in specific settings [22], [23], [24]. Wong and Cumming [22] expose an end-to-end DCE scheme, which resolves the Doppler ambiguity and works on various terrain types, including land, water, and ice, while López-Dekker et al. [23] discuss temporal and phase synchronization for bistatic SAR and the Doppler estimation procedure. Hansen et al. [25] presented the processing steps and error corrections needed to retrieve the estimates of sea surface range Doppler velocities from ENVISAT advanced SAR wide swath medium-resolution image products. They addressed the retrieval accuracy based on examination of the corrected Doppler shift measurements.

Mainly, the DCE approach was used in various SAR processing chains from focusing algorithms to parameters estimation. Differently, we combine the SD and a dc estimation algorithm (the one proposed in [21], but applied on a sliding 2-D window) to obtain physics-based representations for ocean SAR vignettes. Those representations are used for classification and unsupervised physics-guided image retrieval, empirically proving that the SD significantly improves the performance when is used as a preprocessing stage before DCE. This leads the way to new unsupervised physics-guided retrieval algorithms.

C. Transformer Models

Due to the recent progress of attention mechanisms [26], transformers have become attractive and powerful choices for SAR-related tasks [27], [28], [29]. In [27], a vision transformer (ViT)-based [26] representation learning framework is proposed, which use self-attention to replace convolution, which shifts the focus from the information in local neighborhoods to the long-range interactions between each pixel. Basit et al. [28] add a gradient profile loss to the classical convolutional neural networks (CNNs) and ViT-based hybrid models for oil spills in SAR imagery. Li et al. propose an enhancement Swin transformer detection network (ESTDNet) to complete the ship detection in SAR images to solve the issues related to the...
characteristics of strong scattering, multiscale, and complex backgrounds of ship objects in SAR images.

In our work, we aim to benefit from the modeling power of transformers while being able to process reasonable downsampled ocean SAR images, and we adopt a generative convolutional transformer with a manageable number of parameters called CyTran [30]. We used it in an unsupervised setup, showing that, using SD as a preprocessing stage, we improve the SAR image descriptors, leading to an important precision boost for image retrieval.

D. Image Retrieval

The content-based image retrieval aims to find images from a large-scale dataset, which are similar with a query image. In general, the similarity between the features of the query image and all other images from the dataset is used to rank the images for retrieval. Thus, the performance of any image retrieval algorithm depends on the similarity computation between samples. Ideally, the similarity score between two images should be discriminative, robust, and efficient. Various methods based on handcrafted descriptors [31], [32], [33], distance metric learning [34], [35], [36], deep learning models [9], [37], and unsupervised learning [1], [11], [12], [13] have been proposed for the image retrieval task. However, the deep learning has emerged as a dominating alternative of hand-designed feature engineering, the features being learned automatically from data.

More closely to our task, there have been several works for content-based image retrieval from RS data [1], [4], [5], [11], [38]. Espinoza-Molina and Datcu [4] propose a classical approach for EO image retrieval based on enriched metadata, semantic annotations, and image content. The solution generates an EO data model by using automatic feature extraction, processing the EO product metadata and defining semantics, which later is used to answer complex queries. Ye et al. [11] propose an unsupervised domain adaptation model based on CNNs to learn the domain-invariant feature between SAR images and optical aerial images for SAR image retrieving. Sumbul and Demir [5] propose a plasticity–stability preserving multitask learning approach to ensure the plasticity and the stability conditions of whole learning procedure independently of the number and type of tasks. This is achieved by defining two novel loss functions, the plasticity preserving loss and the stability preserving loss. They reported superior results compared with state-of-the-art methods for content-based image retrieval.

Regarding the unsupervised image retrieval, Tang et al. [12] combine the unsupervised feature learning method based on the bag-of-words with $k$-nearest neighbors algorithm for text to image SAR image retrieval. Ye et al. [11] propose an unsupervised domain adaptation model based on CNN to learn the domain-invariant feature between SAR images and optical aerial images for SAR image retrieving.

Distinct from all mentioned methods, we exploit the SD RS algorithms to improve the performance of unsupervised image retrieval algorithm, by enriching the ocean SAR image descriptive embeddings. Moreover, we are the first, which perform physics-guided unsupervised image retrieval based on DCE, opening the frontiers for a new research area.

III. Method

A. Subaperture Decomposition

A classical SAR system acquires the backscatter returned from irradiated targets in different positions and different azimuth angles along the radar trajectory. The real antenna aperture is replaced by the synthetic aperture to obtain high azimuth resolution. Considering that the ocean surface is highly nonstationary, observing it from different angles might bring additional information about the illuminated area. Thus, we decompose the vignette into subapertures, each one corresponding to the image formed using only a part of the total azimuth angle. Decomposing the vignette, we can mimic different observation angles of the same scene, gathering more information. The SD algorithm is visually described in the first part of Fig. 1.
1) \( \sigma_0 \) Calibration: According to [39], the measured normalized radar cross section \( \sigma_0 \) by SAR over the ocean is highly dependent on the local ocean surface wind and viewing angles (incidence and azimuth) of the radar. Therefore, \( \sigma_0 \) of each input vignette is calibrated by dividing it by a reference factor, constructed by assuming a constant wind of 10 m/s at 45° relative to the antenna look angle.

2) Azimuth FFT: The output of the \( \sigma_0 \) calibration block is fed into the azimuth fast Fourier transform (FFT) block, where we perform the FFT along the azimuth axis to obtain the vignette’s spectrum. The number of FFT points is equal to the number of points in the azimuth direction.

3) Hamming Window Compensation: The spectrum composed by the azimuth FFT block is compensated with a Hamming window, having a coefficient of 0.75, in order to obtain a flat azimuth spectrum. The result is shown in the second and third pictures from Fig. 1 (left).

4) Subaperture Generation: In the following stage, we filter the processed vignette with four shifted Hamming windows (with the same 0.75 coefficient), in order to obtain the corresponding azimuth spectrum for each subaperture.

5) Azimuth iFFT: Having the azimuth spectrum for each subaperture, we want to transform back the data into time domain by computing an inverse FFT (iFFT), with the same parameters from the azimuth FFT block. The time domain subapertures are forward processed by the DCE pipeline.

B. Doppler Centroid Estimation

Let \( X_i \in \mathbb{R}^{m \times n} \) be the \( i \)th subaperture for a vignette, where \( m, n \in \mathbb{N} \) are the azimuth and range dimensions. Let \( Y_i \in \mathbb{R}^{m \times n} \) be the delayed version with one sample in the azimuth axis of \( X_i \). We estimate the Doppler centroids for the \( i \)th subaperture as follows:

\[
D_i = -\text{PRF} \cdot \frac{\text{angle}(Z_i)}{2\pi}
\]

where \( Z_i = \text{filt}(X_i \cdot Y_i^*) \), \( Y_i^* \) is the complex conjugate of \( Y_i \), PRF is the pulse repetition frequency, \( \text{angle}() \) returns the angle of the complex input, and \( \text{filt}() \) is a 2-D mean filter with \( d_1 \times d_2 \) kernel size. Each estimated \( D_i \) is further decimated to obtain a smaller dimensional image in terms of width and height. An illustration could be observed in the orange box from Fig. 1. If the DCE is employed in the pipeline, each Doppler image is estimated on a subaperture and further decimated.

Decimation: The last stage of the preprocessing pipeline is the decimation. The fine-resolution subapertures or DCE is not necessary for large-scale geophysical phenomena, especially since the classes described in [39] have scales of tens to thousands of meters. Therefore, to better highlight larger feature patterns, we low-pass-filter each resulted \( X_i \) and \( D_i \) with a window of \( 10 \times 10 \), each filter’s coefficient being 0.01. The resulted images are then decimated by 1/10 yielding a pixel spacing of 50 m. We highlight that the decimation is performed for both SD and DCE in accordance with the desired output.

C. Unsupervised Neural Network

In our work, we used the CyTran generative architecture formed of a convolutional downsampling block, a convolutional transformer block, and a deconvolutional upsampling block, as illustrated in Fig. 2. We underline that, without the convolutional downsampling block and the replacement of dense layers with convolutional layers inside the transformer block, the transformer would not be able to learn to generate images larger than \( 64 \times 64 \) pixels, due to memory overflow (measured on a Nvidia GeForce RTX 3090 GPU with 24 GB of VRAM).

The downsampling block starts with a convolutional layer formed of 32 filters with a spatial support of \( 7 \times 7 \), which are applied using a padding of three pixels to preserve the spatial dimension, while enriching the number of feature maps to 32. Next, we apply three convolutional layers composed of 32, 64, and 128 filters, respectively. All convolutional filters have a spatial support of \( 3 \times 3 \) and are applied at a stride of 2, using a padding of 1. Each layer is followed by batch norm [40] and rectified linear units (ReLUs) [41]. The downsampling block is followed by the convolutional transformer block, which provides an output tensor of the same size as the input tensor. The convolutional transformer block is inspired by the block proposed in [30].

Let \( X \in \mathbb{R}^{c \times m \times n} \), where \( c \) is the number of channels and \( m \) and \( n \) are the width and height, be the input tensor for the transformer block. The spatial dimensions of the visual tokens are determined by the receptive field of the filters in the convolutional projection layer, illustrated in Fig. 2. Considering that the convolutional projection is formed of three nearly identical projection blocks, with separate parameters, let \( W_0, W_K, \) and \( W_V \) denote the learnable parameters of the three projection layers. The query, key, and value embeddings are computed as follows:

\[
Q = \text{conv_projection}(X, W_0)
\]

\[
K = \text{conv_projection}(X, W_K)
\]

\[
V = \text{conv_projection}(X, W_V)
\]

where \( Q \in \mathbb{R}^{n_q \times d_k}, K \in \mathbb{R}^{n_k \times d_k}, \) and \( V \in \mathbb{R}^{n_v \times d_k} \). For the subsequent operation involving matrix multiplications, we need \( q_k = d_k = n_q \). Due to the equal number of filters in the pointwise convolution in all three blocks, \( q_k = d_k = d_v \). The output query, keys, and values are passed to a multihead attention layer, with the goal of capturing the interaction among all tokens by encoding each entity in terms of the global contextual information. Formally, the multihead attention layer is described as follows:

\[
U = \text{softmax} \left( \frac{Q \cdot K^T}{\sqrt{d_k}} \right) \cdot V
\]

where \( K^T \) is the transpose of \( K \) and \( U \) is the output of the block, which is further fed into the batch-norm layer and the final pointwise convolution. Finally, the result of the
convolutional transformer block is processed by the upsample block, being designed to revert the transformation of the downsampling block.

We use CyTran architecture in an unsupervised manner, aiming for the identity function by performing input autoencoding. Specifically, we want to exactly reproduce the input data, by optimizing the following loss function between the input $X$ and the output and with red arrow is illustrated the place where the descriptive embeddings are taken.

$$\mathcal{L} = (X - \hat{X})^2.$$  \hfill (4)

Finally, after the unsupervised training procedure, we use the CyTran model to encode into embeddings the input data for image retrieval. The embeddings are taken after the convolutional transformer block, as depicted in Fig. 2 (red arrow).

D. Content-Based Image Retrieval

Considering a very large database with ocean SAR images, we propose an unsupervised algorithm, which can find similar vignettes, serving researchers as a tool to study physical phenomena on the ocean surface. We formally describe the steps in Algorithm 1.

We consider as requested input the database, the query image, and some hyperparameters. In the first stage, we train in an unsupervised fashion the CyTran autoencoder model denoted by $f$. We optimize the model, such that we obtain a close reconstruction of the input $X$. In the next stage, we remove the upsampling block from the CyTran model, and we define by $f_e$ the pretrained model that computes the descriptive embeddings for each vignette. Next, we process each SAR image from the database, associating in DB$_e$ a pair formed by the original image $X$ and the corresponding embedding $X_e$ vector. At the end of Stage 2, we will have an associated embedding for each vignette. We highlight that, Stage 1 and Stage 2, must be performed only once and do not introduce any time overhead in the retrieval stage. Furthermore, for brevity, we will use the following abbreviations for dataset classes: POWs—pure ocean waves, WSs—wind streaks, MCCs—micro convective cells, RCs—rain cells, BSs—biological slicks, SI—sea ice, Ic—iceberg, LWA—low wind area, AF—atmospheric front, and OF—oceanic front.

IV. EXPERIMENTAL SETUP

A. Dataset

TenGeo SARwv dataset contains over 37,000 ocean vignettes with ten geophysical phenomena. Following [14], we used the amplitude vignettes from the TenGeo SARwv dataset, with the assigned labels, and randomly split the data in training (70%), validation (15%), and test (15%). Moreover, for Doppler-based experiments, we processed the amplitude vignettes in accordance with the full data pipeline described in Fig. 1. Furthermore, we use CyTran generative architecture. The model is formed of a downsampling block comprising convolutional layers, a convolution transformer block comprising a multhead self-attention mechanism, and an upsampling block comprising transposed convolutions. By $\mathcal{L}$, we denote the mean square error loss function between the input and the output and with red arrow is illustrated the place where the descriptive embeddings are taken.
Algorithm 1 Physics-Guided Content-Based SAR Image Retrieval

Input: $DB$ - database with SAR images; $(X)$ - samples from $DB$; $Q$ - query SAR image; $N_{\text{max}}$ - the number of images returned; $\eta$ - learning rate; $\mathcal{L}$ - loss function; $d$ - cosine similarity.

Notations: $f$ - the CyTran model; $f_e$ - embedding function from the CyTran model; $\theta$ - the weights of the model; $\mathcal{S}$ - a function that jointly sorts the input set; $\mathcal{N}(0, \Sigma)$ - the normal distribution of mean 0 and standard deviation $\Sigma$; $\mathcal{U}$ - uniform distribution; $DB_e$ - database with image-embeddings pairs, $X_e$ - the embedding of the $X$ image, $Q_e$ - the embedding of the query image $Q$.

Initialization: $\theta^{(0)} \sim \mathcal{N}(0, \Sigma)$

Output: $U$ - a set of $N_{\text{max}}$ elements from $DB$ similar to the query SAR image $Q$.

Stage 1: Unsupervised pre-training of autoencoding model

1: for $i \leftarrow 1$ to $n$ do
2:   $t \leftarrow 0$
3:   while converge criterion not met do
4:     $X^{(t)} \leftarrow$ mini-batch $\sim \mathcal{U}(DB)$
5:     $\theta_i^{(t+1)} = \theta_i^{(t)} - \eta \nabla\mathcal{L}(\theta_i^{(t)}, X^{(t)})$
6:   $t \leftarrow t + 1$

Stage 2: Processing the database

7: $DB_e = \emptyset$
8: for $X \leftarrow DB$ do
9:   $X_e = f_e(X)$
10: $DB_e \leftarrow (X, X_e)$

Stage 3: SAR image retrieval

11: $D = \emptyset$
12: $Q_e = f_e(Q)$
13: for $(X, X_e) \leftarrow DB_e$ do
14:   $m \leftarrow d(Q_e, X_e)$
15:   $D \leftarrow (X, m)$
16: $D \leftarrow \text{sort}(D)$ with respect to $m$
17: $U \leftarrow D[1 : N_{\text{max}}]$

B. Hyperparameters Tuning

For the classification experiment, we tuned the hyperparameters similar to [14]. Regarding the CyTran model, we used the same network hyperparameters as proposed in [30], only adjusting the input and output number of channels, in accordance with the input type. We trained the model for 100 epochs using Adam optimizer and a mini-batch size of 16. Regarding DCE, we used $d_1 = d_2 = 32$ for the mean filter.

C. Evaluation Metrics

We reported the accuracy for the classification task and performed McNemar statistic tests to show the statistical significance of our results. Regarding the retrieval task, considering that we target big data streams, we reported the precision for 5 ($P@5$) and 50 ($P@50$) examples. Each score was averaged for 100 queries; more precisely, we computed $P@5$ and $P@50$ for 100 query samples and averaged the results.

D. DNNs for Classification

The success of the CNNs in image processing tasks [42] encouraged their introduction in RS applications and SAR imagery [43], [44], [45], [46]. Thus, we followed our previous work [14], proposing a data-centering approach, rather than a novel model architecture. We focused our attention on the preprocessing stage and employed two well-known architectures, ResNet18 [47] and InceptionV3 [48], for the ocean SAR image classification task. The networks were pretrained on the ImageNet dataset, and minimal architectural changes were...
TABLE II

RETRIEVAL RESULTS ON TENGEO SARwV TEST SET CONSIDERING THE EMBEDDINGS FROM RESNET18 (S—SUPERVISED TRAINING) AND CYTRAN (U—UNSUPERVISED TRAINING) MODELS. WE REPORTED RESULTS WHEN WE CONSIDER AS INPUT DATA THE ORIGINAL VIGNETTE (VIG) AND ALL SUBAPERTURES (SUBAP). BY P@m, WE DENOTE THE PRECISION SCORE FOR THE MOST SIMILAR m SAMPLES

| Method | POW P@5 | POW P@50 | WS P@5 | WS P@50 | MCC P@5 | MCC P@50 | RC P@5 | RC P@50 | BS P@5 | BS P@50 | SI P@5 | SI P@50 | Ic P@5 | Ic P@50 | LWA P@5 | LWA P@50 | AF P@5 | OF P@5 | Overall |
|--------|---------|---------|--------|---------|---------|---------|--------|---------|--------|---------|--------|---------|--------|---------|---------|---------|--------|--------|---------|
| S-Vig  | 100     | 99.8    | 99.8   | 99.7    | 100     | 99.6    | 99.4   | 98.5    | 99.4   | 98.9    | 99.8   | 98.9    | 97.2   | 96.4    | 97.2   | 96.9    | 97.4   | 94.8   | 91.6    | 89.4    | 98.1    | 97.4    |
| S-Subap| 99.6    | 99.8    | 99.2   | 99.5    | 100     | 99.8    | 99.8   | 99.5    | 99.8   | 99.4    | 99.4   | 99.8    | 99.8   | 98.1    | 98.8   | 94.4    | 97.2   | 89.6    | 98.9    | 97.2    |
| U-Vig  | 76.8    | 64.0    | 46.0   | 32.1    | 37.6    | 22.1    | 46.6   | 28.8    | 49.0   | 33.3    | 38.4   | 22.7    | 30.0   | 11.8    | 89.8   | 84.5    | 33.8   | 17.2    | 26.6    | 9.3     | 47.4    | 32.6    |
| U-Subap| 89.8    | 83.2    | 82.2   | 70.9    | 64.2    | 51.6    | 57.0   | 38.9    | 78.6   | 66.9    | 76.0   | 54.2    | 57.0   | 38.0    | 91.0   | 82.1    | 63.8   | 47.2    | 66.8    | 39.6    | 72.6    | 57.3    |

Fig. 5. Retrieval results based on embeddings from the CyTran model trained on all subapertures from the original vignette. We present the most similar \(N_{\text{max}} = 100\) samples with localization information. The query image is represented in green, the images found from the same class are in blue, and the images found from wrong classes are in red. In the right side, we show the original vignette for some samples: green and blue (POWs) and red (Ic).

made: the number of output neurons and the number of input channels.

V. EXPERIMENTAL RESULTS

A. Classification Results

We extend the results from [14] in Table I, where we report the classification accuracy obtained for the ResNet18 model on TenGeo SARwv test set, considering multiple inputs data types. When we consider as training input all the subapertures computed on the vignette, we observe a performance boost of 0.9%, with respect to the model trained on the original vignette. But, when we feed only the first subaperture, an accuracy drop of 4% occurs. Similarly, when we train the model on DCE on subapertures against DEC on original vignette, we observe a drastically improvement of 14.7%. This highlights that the SD algorithm applied on the ocean vignettes helps the training process for both amplitude SAR data and DCE.

In addition, we tested the resilience of our network to \(\sigma_0\) calibration errors, which could easily appear in the preprocessing pipeline. More precisely, the pretrained ResNet18 model, trained on amplitude SAR data, was tested on the same test set, but without any \(\sigma_0\) calibration. Even if the data were not calibrated, the model accuracy drop was only 4.1%, attaining an accuracy of 93.9. This highlights that the model is robust with respect to \(\sigma_0\) calibration errors, being suitable to be deployed in real environments.

B. Unsupervised Training Results

We trained the CyTran [30] autoencoder model on the TenGeo SARwv training set and choose the best model with respect to the reconstruction loss on the evaluation set. We highlight that multiple models were tried (e.g., ResNet autoencoder and U-Net), but they did not converge to optimal reconstruction results; therefore, we excluded them.

We visualized with T-SNE the embedding feature space when we considered as input all the subapertures on the original vignette. For a more accurate comparison, we did the visualization class by class and included the results for BSs, in Fig. 3, and LWA, in Fig. 4. We note that the feature space is distinct (for CyTran embeddings, we clearly see two cloud points for both figures), suggesting that into the same annotated class from TenGeo SARwv, we could find distinguishable phenomena. Moreover,
the distance metric is not preserved between feature spaces, emphasized by the blue points from both Figs. 3 and 4, which are close in one feature space and randomly spread into the other.

C. Retrieval Results

On the one hand, in Table II, we reported the retrieval performance on embeddings provided by CyTran network, trained on original vignette and subapertures. We compared the retrieval results against the embeddings computed by ResNet18 model trained in a supervised fashion. When we compare the supervised embeddings on original vignette (S-Vig) and subapertures (S-Subap), the results are comparable, with overall differences smaller than 1% for both $P@5$ and $P@50$. But, the SD algorithm offers a consistent precision boost when we refer to the retrieval results with unsupervised embeddings. We observe that the unsupervised embeddings on subapertures raise the $P@5$ and $P@50$ for each and every class, with an overall improvement of 25.2% for $P@5$ and 24.7% for $P@50$. Even if the SD does not bring a precision boost for supervised embeddings, most probably because of the saturated accuracy on the dataset, the algorithm has a huge impact on unsupervised scenarios, reducing the retrieval performance gap between supervised and unsupervised approaches.

On the other hand, we did a more physics-based experiment, considering as input data the DCE, which are directly correlated with physic phenomena (e.g., ocean currents). In Table IV, we reported the retrieval performance on embeddings provided by CyTran network, trained on DCE on original vignette and subapertures. We compared the retrieval results against the embeddings computed by ResNet18 model trained in a supervised fashion. As we would expect from the classification experiment, the retrieval performance is considerable improved when the supervised embeddings based on subapertures are used. The same trend is observed for the unsupervised embeddings. More precisely, the $P@5$ for U-Dop Subap is with 15.6% higher than U-Dop Vig, and the $P@50$ is with 16.0% higher. Thus, SD algorithm has a major positive impact on the retrieval task, when DCE data are used, leading the way to more complex search engines.

We illustrated the retrieval results for the unsupervised embeddings trained on subapertures for two query images: in Fig. 5 for POWs class and in Fig. 6 for AF class. For both figures, we observe that the most similar images found are randomly spread in the geographical area where the phenomena could appear, indicating that the unsupervised model does not overfit with respect to the geographical area. Moreover, structural similarities were observed for the images found with wrong label (the red points in Figs. 5 and 6), which can indicate the presence of two phenomena in the same location or other intrinsic similarities.

In addition, we compared our unsupervised CyTran model with two state-of-the-art retrieval algorithms applied in computer vision [49] and RS [50], including the results in Table III. We observe that both CyTran models, trained on amplitude vignettes and subapertures, surpass the method proposed in [49], but DCSH [50] attains superior retrieval results compared with our method based on CyTran trained on amplitude vignettes. Nevertheless, the CyTran model trained on subapertures surpasses DCSH [50] with more than 7% for...
both $P@5$ and $P@50$. This highlights that our algorithm leverages the advantages of powerful transformer networks and meaningful physics-based algorithms.

VI. ABLATION STUDY

We study the importance of SD algorithm for the classification task in Table I, considering as training data the amplitude SAR and the DCE images. We observe that introducing SD improves both scenarios, especially when the network was trained with DCE data. Nevertheless, we observe that if we combine SD and DCE (DCE Subapertures from Table I), we obtain lower performance compared to SD only. This highlights that classifying the scenes based on physical parameters is considerably more difficult. Regarding the retrieval task, the same pattern is observed. Adding the SD preprocessing step improves the retrieval results for all unsupervised experiments, but when we compare Tables II and IV, we observe that combining SD and DCE attains inferior results to SD only.

VII. CONCLUSION

In this work, we extended the previous approach from [14] by using the SD algorithm for unsupervised feature learning with transformer networks. The unsupervised features were used for an SAR retrieval algorithm on the ocean surface, showing important improvements in performance when the SD was used as a pretraining stage for the models. Moreover, we showed that the SD method has a huge impact on retrieval performance when more physics-based algorithms, as DCE, are used for ocean retrieval. This experiment allows us to build more complex searching engines, which could find similar physical parameters, instead of similar structures (e.g., ocean currents speed). In summary, we used a data-centering approach to improve the performance classification and retrieval algorithms, in both supervised and unsupervised settings.

REFERENCES

[1] L. Jiao, X. Tang, B. Hou, and S. Wang, “SAR images retrieval based on semantic classification and region-based similarity measure for earth observation,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 8, pp. 3876–3891, Sep. 2015.

[2] R. H. Stewart, “Seasat: Results of the mission,” Bull. Amer. Meteorological Soc., vol. 69, no. 12, pp. 1441–1447, Dec. 1988.

[3] J. E. Stopa, F. Ardhuin, R. Husson, H. Jiang, B. Chapron, and F. Collard, “Swell dissipation from 10 years of Envisat advanced synthetic aperture radar in wave mode,” Geophys. Res. Lett., vol. 43, no. 7, pp. 3423–3430, 2016.

[4] D. Espinoza-Molina and M. Datcu, “Earth-observation image retrieval based on content, semantics, and metadata,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 11, pp. 5145–5159, Nov. 2013.

[5] G. Sumbul and B. Demir, “Plasticity-stability preserving multi-task learning for remote sensing image retrieval,” IEEE Trans. Geosci. Remote Sens., vol. 60, 2022, Art. no. 5620116.

[6] X. Yang, X. Li, W. G. Pichel, and Z. Li, “Comparison of ocean surface winds from ENVISAT ASAR, Metop ASCAT scatterometer, buoy measurements, and NOGAPS model,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 12, pp. 4743–4750, Dec. 2011.

[7] G. Zheng et al., “Development of a gray-level co-occurrence matrix-based texture orientation estimation method and its application in sea surface wind direction retrieval from SAR imagery,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 9, pp. 5244–5260, Sep. 2018.

[8] X. Li et al., “Deep-learning-based information mining from ocean remote-sensing imagery,” Nat. Sci. Rev., vol. 7, no. 10, pp. 1584–1605, Oct. 2020.

[9] S. R. Dubey, “A decade survey of content based image retrieval using deep learning,” IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 5, pp. 2687–2704, May 2021.

[10] B. Neyshabur, S. Bhojanapalli, D. Macllester, and N. Srbro, “Exploring generalization in deep learning,” in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 1–10.

[11] F. Ye, W. Luo, M. Dong, H. He, and W. Min, “SAR image retrieval based on unsupervised domain adaptation and clustering,” IEEE Geosci. Remote Sens. Lett., vol. 16, no. 9, pp. 1482–1486, Sep. 2019.

[12] 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.

[13] 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.

[14] N.-C. Ristea, A. Anghel, M. Datcu, and B. Chapron, “Guided deep learning by subaperture decomposition: Ocean patterns from SAR imagery,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2022, pp. 4067–4070.

[15] Z. Wang, X. Fu, and K. Xia, “Target classification for single-channel SAR images based on transfer learning with subaperture decomposition,” IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1–5, 2022.

[16] C. Brekke, S. N. Anfinsen, and Y. Larsen, “Subband extraction strategies in ship detection with the subaperture cross-correlation magnitude,” IEEE Trans. Geosci. Remote Sens. Lett., vol. 10, no. 4, pp. 786–790, Jul. 2013.

[17] J. Singh, M. Soccorsi, and M. Datcu, “SAR complex image analysis: A Gauss Markov and a multiple sub-aperture based target characterization,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., Jul. 2010, pp. 1585–1588.
[18] A. Focsa, M. Datcu, S.-A. Toma, A. Anghel, and R. Cacoveanu, “Opportunististic bistatic SAR image classification using sub-aperture decomposition,” in Proc. 13th Int. Conf. Commun. (COMM), Jun. 2020, pp. 203–207.

[19] Z. Lu, G. Hua-dong, and H. Chun-ming, “SAR ocean stationary targets detection,” Remote Sens. Technol. Appl., vol. 22, no. 3, pp. 321–325, Dec. 2020.

[20] R. K. Raney, “Doppler properties of radars in circular orbits,” Int. J. Remote Sens., vol. 7, no. 9, pp. 1153–1162, Sep. 1986.

[21] S. N. Madsen, “Estimating the Doppler centroid of SAR data,” IEEE Trans. Aerosp. Electron. Syst., vol. 25, no. 2, pp. 134–140, Mar. 1989.

[22] F. Wong and I. G. Cumming, “A combined SAR Doppler centroid estimation scheme based upon signal phase,” IEEE Trans. Geosci. Remote Sens., vol. 34, no. 3, pp. 696–707, May 1996.

[23] M. W. Hansen, F. Collard, K. Dagestad, J. A. Johannessen, P. Fabry, and B. Chapron, “Retrieval of sea surface range velocities from Envisat ASAR Doppler centroid measurements,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 10, pp. 3582–3592, Oct. 2011.

[24] A. Basit, M. A. Siddique, M. K. Bhatti, and M. S. Sarfraz, “Comparison of CNNs and vision transformers-based hybrid models using gradient profile loss for classification of oil spills in SAR images,” Remote Sens., vol. 14, no. 9, p. 2085, Apr. 2022.

[25] K. Li, M. Zhang, M. Xu, R. Tang, L. Wang, and H. Wang, “Ship detection in SAR images based on feature enhancement Swin transformer and adjacent feature fusion,” Remote Sens., vol. 14, no. 13, p. 3186, Jul. 2022.

[26] N.-C. Ristea et al., “CyTran: A cycle-consistent transformer with multi-level consistency for non-contrast to contrast CT translation,” 2021, arXiv:2110.06400.

[27] C. Yang, T. Zhang, K. Song, N. Sehe, and H. T. Shen, “A survey on learning to hash,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 769–790, May 2018.

[28] M. Kaya and H. C. S. Bilge, “Deep metric learning: A survey,” Pattern Recognit. Image Anal., vol. 29, no. 4, pp. 578–592, Oct. 2020.

[29] A. Bellek, A. Habrard, and M. Sebban, “A survey on metric learning for feature vectors and structured data,” 2013, arXiv:1306.6709.

[30] A. K. Bedi and R. K. Sunkaria, “Local tetra-directional pattern—A new texture descriptor for content-based image retrieval,” 2013, arXiv:1306.6709.

[31] J. Wang, T. Zhang, J. Song, N. Sehe, and H. T. Shen, “A survey on learning to hash,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 769–790, May 2018.

[32] M. Kaya and H. C. S. Bilge, “Deep metric learning: A survey,” Pattern Recognit. Image Anal., vol. 29, no. 4, pp. 578–592, Oct. 2020.

[33] A. Bellek, A. Habrard, and M. Sebban, “A survey on metric learning for feature vectors and structured data,” 2013, arXiv:1306.6709.

[34] J. Wang, T. Zhang, J. Song, N. Sehe, and H. T. Shen, “A survey on learning to hash,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 769–790, May 2018.

[35] M. Kaya and H. C. S. Bilge, “Deep metric learning: A survey,” Pattern Recognit. Image Anal., vol. 29, no. 4, pp. 578–592, Oct. 2020.

[36] A. Bellek, A. Habrard, and M. Sebban, “A survey on metric learning for feature vectors and structured data,” 2013, arXiv:1306.6709.

[37] A. K. Bedi and R. K. Sunkaria, “Local tetra-directional pattern—A new texture descriptor for content-based image retrieval,” 2013, arXiv:1306.6709.

[38] J. Wang, T. Zhang, J. Song, N. Sehe, and H. T. Shen, “A survey on learning to hash,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 769–790, May 2018.

[39] M. Kaya and H. C. S. Bilge, “Deep metric learning: A survey,” Pattern Recognit. Image Anal., vol. 29, no. 4, pp. 578–592, Oct. 2020.

[40] A. Bellek, A. Habrard, and M. Sebban, “A survey on metric learning for feature vectors and structured data,” 2013, arXiv:1306.6709.
Mihai Datcu (Fellow, IEEE) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica of Bucharest (UPB), Bucharest, Romania, in 1978 and 1986, respectively, and the Habilitation a Diriger des Recherches degree in computer science from University Louis Pasteur, Strasbourg, France, in 1999. Since 1981, he has been a Professor with the Department of Applied Electronics and Information Engineering, Faculty of Electronics, Telecommunications and Information Technology, UPB. Since 1993, he has been a Scientist with the German Aerospace Center (DLR), Wellling, Germany. He has held a Visiting Professor appointments with the University of Oviedo, Oviedo, Spain; University Louis Pasteur; International Space University, Strasbourg; the University of Siegen, Siegen, Germany; the University of Innsbruck, Innsbruck, Austria; the University of Alcala, Alcala, Spain; the University Tor Vergata, Rome, Italy; the University of Trento, Trento, Italy; Unicamp, Campinas, Brazil; the China Academy of Science, Shenyang, China; the Universidad Pontificia de Salamanca, campus de Madrid, Spain; the University of Camerino, Camerino, Italy; and the Swiss Center for Scientific Computing, Manno, Switzerland. From 1992 to 2002, he had an Invited Professor Assignment with the Swiss Federal Institute of Technology (ETH Zürich), Zürich, Switzerland. Since 2001, he has been initiating and leading the Competence Center on Information Extraction and Image Understanding for Earth Observation, Paris Institute of Technology (ParisTech), Paris, France, a collaboration of DLR with the French Space Agency (CNES). He has been a Professor holder of the DLRCNES Chair with ParisTech. He has initiated the European frame of projects for image information mining and is involved in research programs for information extraction, data mining and knowledge discovery, and data science with the European Space Agency (ESA), National Aeronautics and Space Administration (NASA), and in a variety of national and European projects. He is the Director of the Research Center for Spatial Information, UPB. He is a Senior Scientist and the Data Intelligence and Knowledge Discovery Research Group Leader with the Remote Sensing Technology Institute, DLR, and delegate in the German Aerospace Center - Office National d’études et de Recherches Aérospatiales (DLR-ONERA) Joint Virtual Center for Artificial Intelligence (AI) in Aerospace. He is a member of the ESA Working Group Big Data from Space and a Visiting Professor with the ESA’s-Lab. His research interests include explainable and physics-aware artificial intelligence, smart radar sensors design, and quantum machine learning with applications in Earth observation.

Dr. Datcu was a recipient of the Romanian Academy Prize Traian Vuia for the development of the system for automated analysis of digital images analysis system and his activity in image processing in 1987, 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 Chaire d’excellence internationale Blaise Pascal 2017 for international recognition in the field of data science in Earth observation. In 2022, he received the IEEE Geoscience and Remote Sensing Society (GRSS) David Landgrebe Award in recognition of outstanding contributions to Earth observation analysis using innovative concepts for big data analysis, image mining, machine learning, smart sensors, and quantum resources. He is the Representative of Romania in the Earth Observation Program Board. He has served as a co-organizer for international conferences and workshops and as a guest editor for a special issue on AI and big data of the IEEE and other journals.

Bertrand Chapron was born in Paris, France, in 1962. He received the B.Eng. degree from the Institut National Polytechnique de Grenoble, Grenoble, France, in 1984, and the Ph.D. degree in fluid mechanics from the University of Aix-Marseille II, Marseille, France, in 1988. He spent three years as a Post-Doctoral Research Associate at the National Aeronautics and Space Administration (NASA)/Goddard Space Flight Center (GSFC)/Wallops Flight Facility, Wallops Island, VA, USA. He has experience in applied mathematics, physical oceanography, electromagnetic waves theory, and its application to ocean remote sensing. He is currently responsible for the Oceanography from Space Laboratory, Ifremer, Plouzané, France.