IEEE Transactions on Geoscience and Remote Sensing
2023, Volume 61 Issue 5207111 Pages 11p.
https://doi.org/10.1109/TGRS.2023.3272279
https://archimer.ifremer.fr/doc/00840/95206/

Archimer
https://archimer.ifremer.fr

Guided Unsupervised Learning by Subaperture Decomposition for Ocean SAR Image Retrieval

Ristea Nicolae-Cătălin [1], Anghel Andrei [1], Datcu Mihai [2], Chapron Bertrand [3]

[1] Research Center for Spatial Information (CEOSpaceTech) and the Department of Telecommunications, University Politehnica of Bucharest, Bucharest, Romania
[2] Research Center for Spatial Information (CEOSpaceTech), University Politehnica of Bucharest, Bucharest, Romania
[3] Laboratoire d'Ocanographie Physique et Spatiale (LOPS), Plouzané, Ifremer, France

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.

Keywords : Doppler centroid estimation (DCE), image retrieval, ocean imagery, remote sensing (RS), subapertures decomposition, synthetic aperture radar (SAR), unsupervised learning

Please note that this is an author-produced PDF of an article accepted for publication following peer review. The definitive publisher-authenticated version is available on the publisher Web site.

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 research community over the last decades. Humans had minimal ocean observations before 1978, when Seasat, the first E arth-orbiting satellite designed for remote sensing 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, infrared) being launched to improve our understanding about the ocean.

Nowadays, one of the most used space-borne 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 meters 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], [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 stream 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].

Other major aspect of ocean big data is the costly process of annotating data. Considering the particularities of remote sensing ocean data, only people with expertise can annotate vignettes (e.g., ocean currents direction, waves height, ocean phenomena), making the process time consuming and costly. Inspired from visual data domain, several works leverage unsupervised information to learn deep representations [11], [12], [13], [14]. Nevertheless, even if there is a moderate success on SAR unsupervised 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 [15] 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 approach. Moreover, we tested our approach for a physics guided remote sensing approach (providing as input to the network Doppler centroid images of the vignettes), and we obtained important improvements when using SD beforehand. 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 [15], our current contribution is twofold:

- We are the first who proposed an unsupervised query by example framework for ocean SAR imagery.
- We combined the previous SD algorithm with DCE and obtained superior results for classification and image retrieval.

II. Related Work

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

A. Subaperture Decomposition

The SD algorithm is widely used for SAR imagery [16], [17], [18], [19], [20], [15]. The method was combined with both classical signal processing algorithms [17], [18], [19] and deep learning techniques [16], [15]. In [17] the SD is proposed for the ship detection task, while in [18] 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 subbands. Afterwards, the authors used pretrained DNN for target classification task on the ground [16].

Distinctly, we propose to extend the SD usage from our previous work [15] 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 [21], [22]. Many works have been proposed to improve the Doppler estimation in specific settings [23], [24], [25]. In [23] authors expose an end-to-end Doppler centroid estimation scheme, which resolves the Doppler ambiguity and works on various terrain types, including land, water and ice, while in [24] the authors discuss temporal and phase synchronization for bistatic SAR and the Doppler estimation procedure. Hansen *et al.*[26] presented the processing steps and error corrections needed to retrieve 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, 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 [22], but applied on a sliding two-dimensional 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 [27], transformers have become attractive and powerful choices for SAR related tasks [28], [29], [30]. In [28] a vision transformer (ViT) [27] based 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. In [29] the authors add a gradient profile loss to the classical CNNs and vision transformer based hybrid models for oil spills in SAR imagery. Li *et al.*propose an enhancement Swin transformer detection network, named ESTDNet, to complete the ship detection in SAR images to solve the issues related to the characteristics of strong scattering, multi-scale, 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 down-sampled ocean SAR images, we adopt a generative convolutional transformer with a manageable number of parameters called CyTran [31]. We used it in an unsupervised set-up, 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 data set, which are similar with a query image. Generally, the similarity between the features of the query image and all others images from the data set 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 hand-crafted descriptors [32], [33], [34],

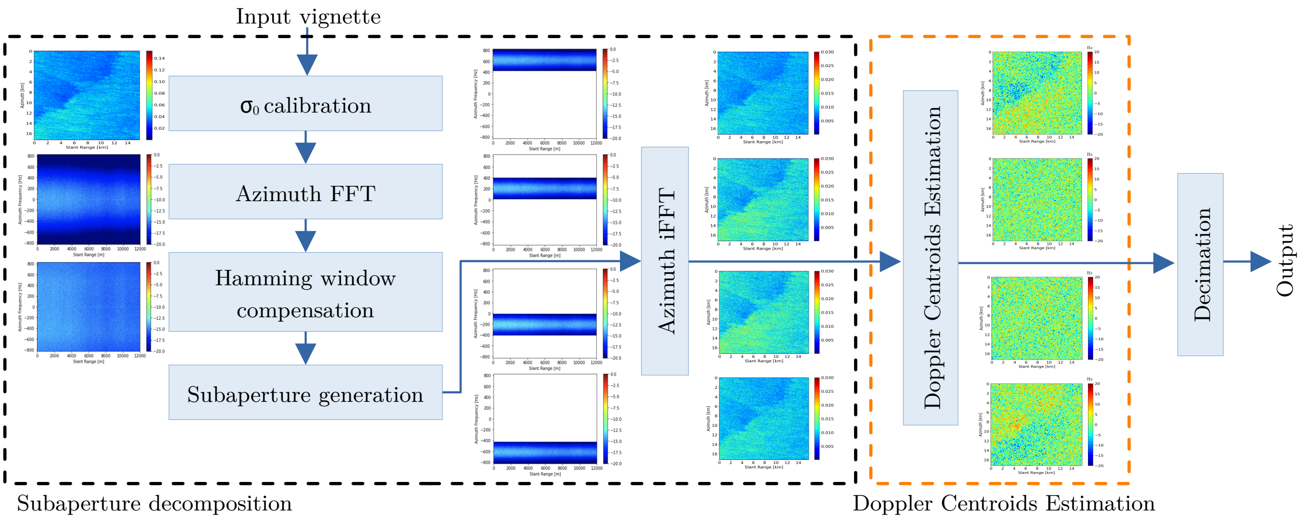

Fig. 1. The preprocessing subaperture decomposition pipeline. An input vignette is processed by a series of blocks, followed by the subaperture generation. Next, each subaperture is processed by the azimuth inverse FFT. The Doppler Centroid Estimation block is an additional algorithm, which can be omitted in accordance with the experiment performed. The output is either the decimated subapertures or the decimated Doppper centroids on the subapertures.

distance metric learning [35], [36], [37], deep learning models [38], [9] and unsupervised learning [11], [12], [13], [14] 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 remote sensing data [1], [5], [4], [39], [40]. In [4] authors 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.*[39] propose an unsupervised domain adaptation model based on convolutional neural networks (CNNs) to learn the domain-invariant feature between SAR images and optical aerial images for SAR image retrieving. In [5] authors propose a plasticity-stability preserving multi-task learning approach to ensure the plasticity and the stability conditions of whole learning procedure independently from 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, in [12] the authors combine the unsupervised feature learning method based on the bag-of-words with k-nearest neighbours 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 remote sensing 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 non-stationary, 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 in the first part of the Fig. 1.

σ_0 calibration. According to [41], the measured normalized radar cross section σ_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, the σ_0 of each input vignette is calibrated by dividing it to a reference factor, constructed by assuming a constant wind of 10 m/s at 45° relative to the antenna look angle.

Azimuth FFT. The output of the σ_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.

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 picture from Fig. 1 left.

Subaperture generation. In the following stage, we filter the processed vignette with 4 shifted Hamming windows (with the same 0.75 coefficient), in order to obtain the corresponding azimuth spectrum for each subaperture.
Azimuth iFFT. Having the azimuth spectrum for each subaperture, we want to transform back the data into time domain by computing an inverse Fast Fourier Transform (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 1 sample in the azimuth axis of X_i. We estimate the Doppler centroids for the i^{th} subaperture as follows:

$$D_i = -PRF \cdot \frac{angle(Z_i)}{2\pi}, \tag{1}$$

where $Z_i = filt(X_i \cdot Y_i^*)$, Y_i^* is the complex conjugate of Y_i, PRF is the pulse repetition frequency, $angle()$ returns the angle of the complex input and $filt()$ is a two dimensional mean filter with $d_1 \times d_2$ kernel size. Each estimated D_i is further decimated. An illustration could be observed in the first two blocks of the Fig. 1.
Decimation. The last stage of the preprocessing pipeline is the decimation. The fine-resolution subapertures or DCE are not necessary for large scale geophysical phenomena, especially since the classes described in [41] have scales of tens to thousands of metres. Therefore to better highlight larger feature patterns, we low-pass-filter each resulted X_i and D_i with a window of 10×10, each filter's coefficient being 0.01. The resulted images are then decimated by $1/10$ yielding a pixel spacing of 50 meters. We highlight that the decimation is performed for both SD and DCE in accordance with the desired output.

C. Deep neural networks for classification

The success of the CNNs in image processing tasks [42] encouraged their introduction in remote sensing applications and SAR imagery [43], [44], [45], [46]. Thus, we followed our previous work [15], proposing a data-centring 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 data set and minimal architectural changes were made: the number of output neurons and the number of input channels.

D. 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 64×64 pixels, due to memory overflow (measured on a Nvidia GeForce RTX 3090 GPU with 24GB of VRAM).

The downsampling block starts with a convolutional layer formed of 32 filters with a spatial support of 7×7, which are applied using a padding of 3 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 3×3 and are applied at a stride of 2, using a padding of 1. Each layer is followed by batch-norm [49] and Rectified Linear Units (ReLU) [50]. 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 [31]. More precisely, the input tensor is interpreted as a set of overlapping visual tokens. The sequence of tokens is projected onto a set of weight matrices implemented as depthwise separable convolution operations. The convolutional projection is formed of three nearly identical projection blocks, with separate parameters. The output query, keys and values are passed to a multi-head attention layer, with the goal of capturing the interaction among all tokens by encoding each entity in terms of the global contextual information. Next, the output passes through a batch-norm and a pointwise convolution. Lastly, the result of the convolutional transformer block is processed by the upsampling 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 auto-encoding. Specifically, we want to exactly reproduce the input data, by optimizing the following loss function between the input X and output $\hat{X}$.

$$\mathcal{L} = (X - \hat{X})^2 \tag{2}$$

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 with a red arrow.

E. 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 the Algorithm 1.

We consider as requested input the database, the query image and some hyper-parameters. In the first stage, we train in an unsupervised fashion the CyTran auto-encoder 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 $\hat{f}$ the pretrained model that computes the descriptive

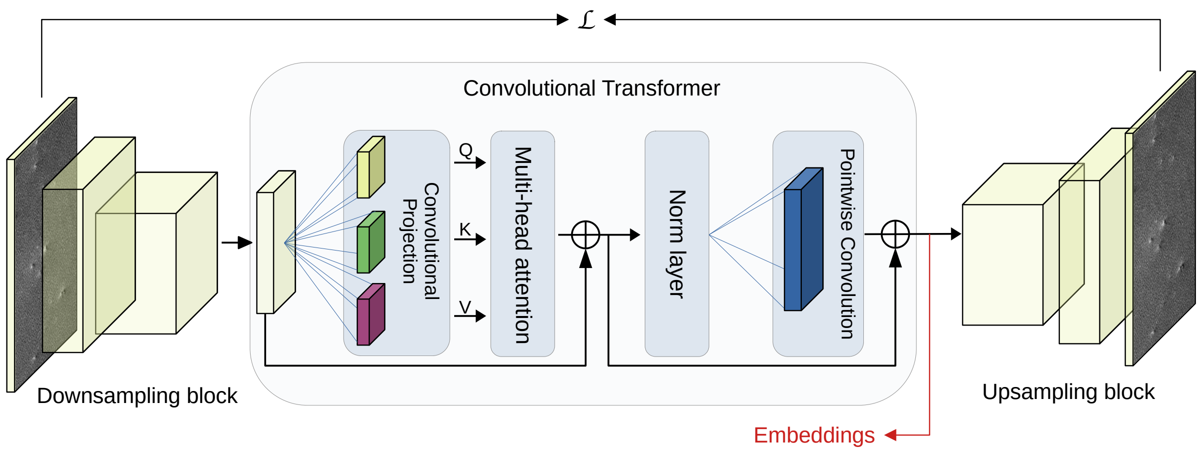

Fig. 2. CyTran generative architecture. The model is formed of a downsampling block comprising convolutional layers, a convolution transformer block comprising a multi-head self-attention mechanism, and an upsampling block comprising transposed convolutions. By $\mathcal{L}$ we denote the mean square 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_{max} - the number of images returned; η - learning rate; $\mathcal{L}$ - loss function; d - cosine similarity.

Notations: f - the CyTran model; $\hat{f}$ - embedding function from the CyTran model; θ - the weights of the model; *sort* - a function that jointly sorts the input set; $\mathcal{N}(0, \Sigma)$ - the normal distribution of mean 0 and standard deviation Σ; $\mathcal{U}$ - uniform distribution

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

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

```
Stage 1: Unsupervised pre-training of auto-encoding model
for $i \leftarrow 1$ to $n$ do
    $t \leftarrow 0$
    while converge criterion not met do
        $X^{(t)} \leftarrow$ mini-batch $\sim \mathcal{U}(DB)$
        $\theta_i^{(t+1)} = \theta_i^{(t)} - \eta^{(t)} \nabla \mathcal{L}\left(\theta_i^{(t)}, X^{(t)}\right)$
        $t \leftarrow t + 1$
Stage 2: Processing the database
$\hat{DB}$ = empty
for $X \leftarrow DB$ do
    $\hat{X} = \hat{f}(X)$
    $\hat{DB} \leftarrow (X, \hat{X})$
Stage 3: SAR image retrieval
$D$ = empty
$\hat{Q} = \hat{f}(Q)$
for $(X, \hat{X}) \leftarrow \hat{DB}$ do
    $m \leftarrow d(\hat{Q}, \hat{X})$
    $D \leftarrow (X, m)$
$D \leftarrow sort(D)$ with respect to $m$
$U \leftarrow D[1 : N_{max}]$
```

embeddings for each vignette. Next, we process each SAR image from the database, associating in $\hat{DB}$ a pair formed by the original image X and the corresponding embedding $\hat{X}$ vector. At the end of the stage two, 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. Lastly, in the Stage 3 we perform the actual image retrieval. We compute the embedding vector $\hat{Q}$ of the query image Q by $\hat{Q} = \hat{f}(Q)$. Afterwards, we calculate the cosine distance between $\hat{Q}$ and each embedding vector from $\hat{DB}$, as follows:

$$d(\hat{Q}, \hat{X}) = \frac{\hat{Q} \cdot \hat{X}}{||\hat{Q}|| \cdot ||\hat{X}||}, \tag{3}$$

where $||x||$ stands from the L_2-norm of vector x.

All the distances m associated with the vignettes from DB are cashed in D. In the last two steps from Stage 3, we sort D in accordance with the distance m and take the most N_{max} similar examples. In this manner, we can obtain an arbitrary N_{max} number of the most similar examples, by only computing the distance between the query embedding vector and sorting the result.

We emphasis that our algorithm is general, does not require labels and is not constrained for any specific input data. To demonstrate the generality of our method, we considered as input two distinct distribution data, the SAR subapertures and the Doppler centroids estimated from subapertures. Therefore, we perform a content based image retrieval from both raw data and physics aware representations. The algorithm feed with the latter data type could build a more complex search engine, capable to find phenomena based on specific physical features (e.g., ocean currents with a certain speed).

IV. Experimental setup

Data set. TenGeoP-SARwv data set contains over 37,000 ocean vignettes with 10 geophysical phenomena. Following [15], we used the raw vignettes from the TenGeoP-SARwv data set, with the assigned labels, and randomly split the data

TABLE I
Retrieval results on TenGeoPSAR-wv 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		WS		MCC		RC		BS		SI		Ic		LWA		AF		OF		Overall	
	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50
S-Vig	100	99.8	99.8	99.7	100	99.6	99.4	98.5	99.4	98.9	99.8	99.7	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	99.2	98.8	98.2	96.5	99.8	99.7	99.4	98.9	99.4	98.1	98.2	96.7	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

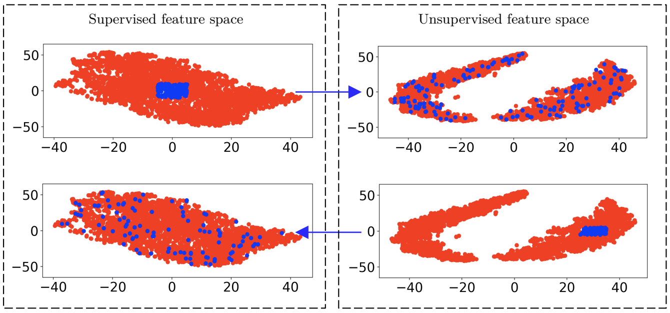

Fig. 3. Embedding space comparison for the biological slicks class between supervised and unsupervised trainings. The figures are horizontally corespondent, indicating the samples annalogy between feature spaces. The dimensional reduction is computed with T-SNE.

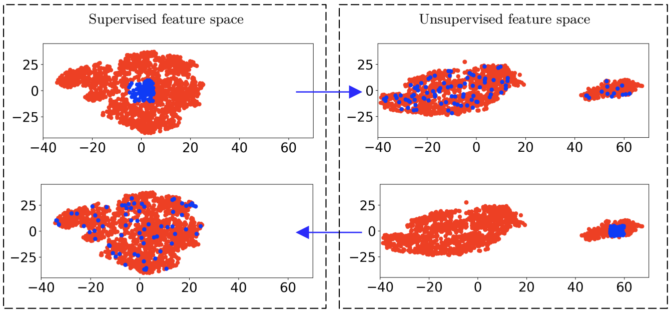

Fig. 4. Embedding space comparison for the low wind area class between supervised and unsupervised trainings. The figures are horizontally corespondent, indicating the samples annalogy between feature spaces. The dimensional reduction is computed with T-SNE.

in training (70%), validation (15%) and test (15%). Moreover, for Doppler based experiments we processed the raw vignettes in accordance with the full data pipeline described in Fig. 1. Further, for brevity we will use the following abbreviations for data set classes: POW - Pure Ocean Waves, WS - Wind Streaks, MCC - Micro Convective Cells, RC - Rain Cells, BS - Biological Slicks, SI - Sea Ice, Ic - Iceberg, LWA - Low Wind Area, AF - Atmospheric Front, OF - Oceanic Front.

Hyper-parameters tuning. For the classification experiment, we tuned the hyper-parameters similar to [15]. Regarding the CyTran model, we used the same network hyper-parameters as proposed in [31], 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.

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.

TABLE II
Accuracy results for a ResNet18 model on the TenGeoP-SARwv test set. We denote by "Subaperture (1)" that the input is only the first subaperture, while for "Subapertures" all four are considered. The significantly better results (level 0.01) than corresponding baselines, according to a paired McNemar's test, are marked with †.

Vignette	98.0
Subaperture (1)	94.0
Subapertures	98.9†
DCE Vignette	78.6
DCE Subapertures (1)	75.3
DCE Subapertures	93.3†

V. Experimental Results

Classification results. We extend the results from [15] in Table II, where we report the classification accuracy obtained for the ResNet18 model on TenGeoP-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

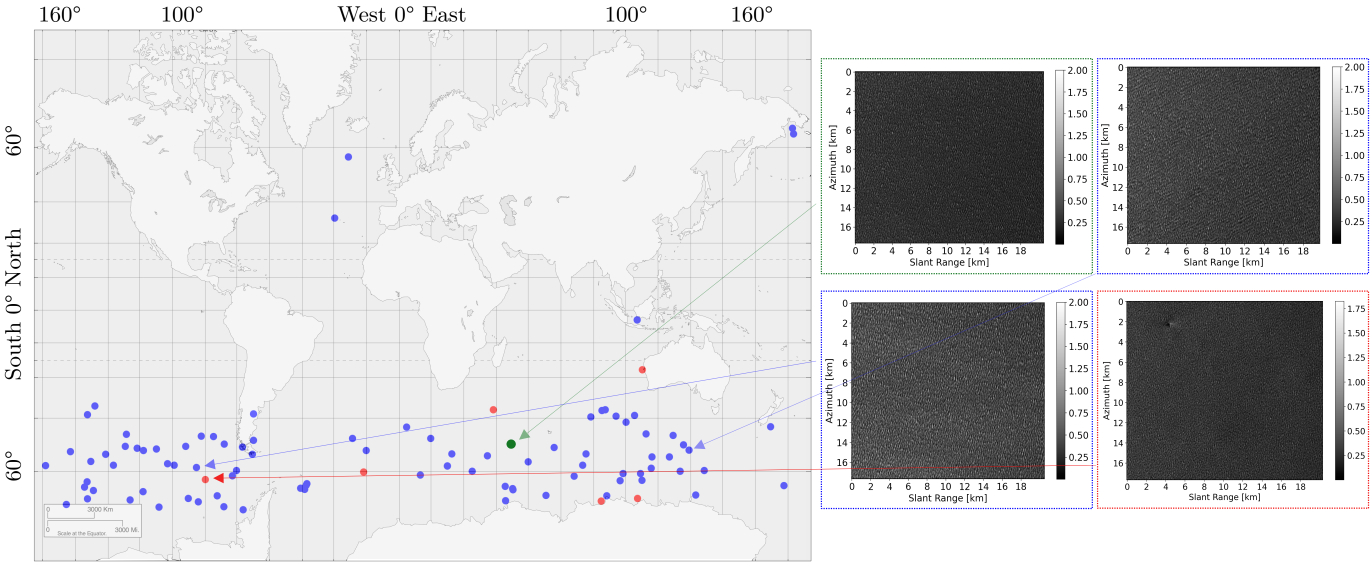

Fig. 5. Retrieval results based on embeddings from CyTran model trained on all subapertures from the original vignette. We present the most similar $N_{max} = 100$ samples with localisation information. In green is represented the query image, in blue the images found from the same class and in red the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

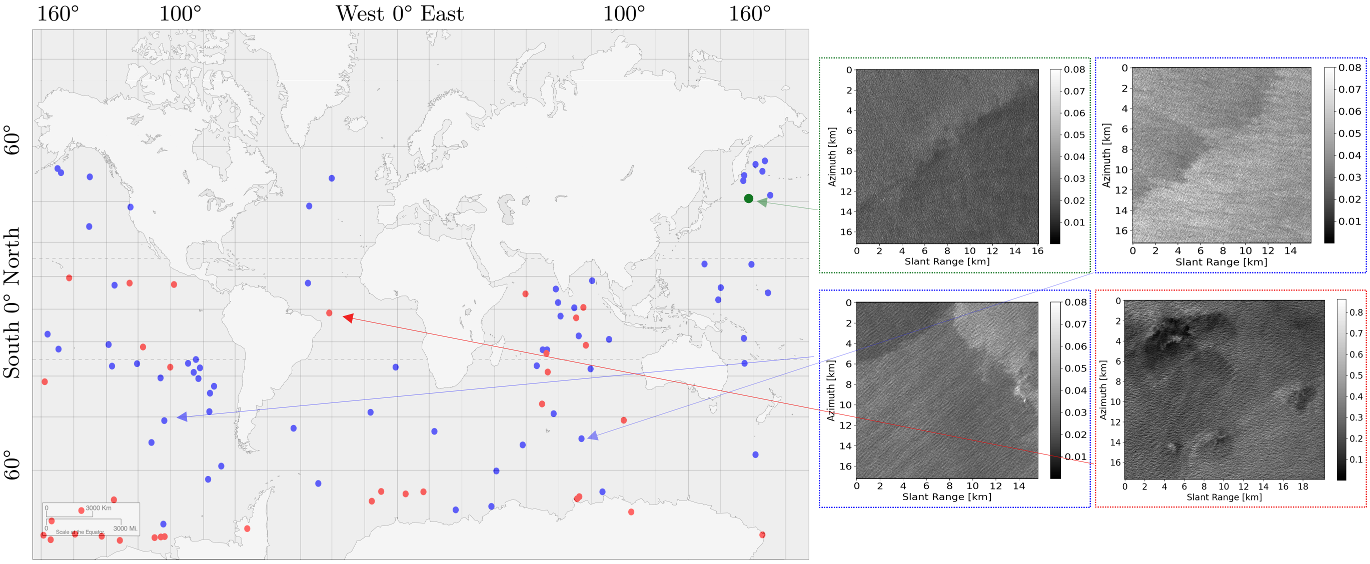

Fig. 6. Retrieval results based on embeddings from CyTran model trained on all subapertures from the original vignette. We present the most similar $N_{max} = 100$ samples with localisation information. In green is represented the query image, in blue the images found from the same class and in red the images found from wrong classes. In the right side, we show the original vignette for some samples: green and blue (Atmospheric Front), red (Micro Convective Cells).

14.7%. This highlights that the SD algorithm applied on the ocean vignettes helps the training process for both raw SAR data and DCE.

Unsupervised training results. We trained the CyTran [31] auto-encoder model on the TenGeoP-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 auto-encoder, 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 both the supervised trained model and CyTran. For a more accurate comparison, we did the visualization class by class and included the results for biological slicks, in Fig. 3, and low wind area, 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 TenGeoP-SARwv we could find distinguishable phenomena. Moreover, the distance metric is not preserved between feature spaces, emphasized by the blue points from both Fig. 3 and Fig. 4, which are close in one feature space and randomly spread into the other.

Retrieval results. On the one hand, in Table I we reported the retrieval performance on embeddings provided by CyTran network, trained on original vignette and subapertures. We

TABLE III
RETRIEVAL RESULTS ON TENGEOPSAR-WV 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 DCE ON THE ORIGINAL VIGNETTE (VIG) AND DCE ON ALL SUBAPERTURES (SUBAP). BY P@m WE DENOTE THE PRECISION SCORE FOR THE MOST SIMILAR m SAMPLES.

Method	POW		WS		MCC		RC		BS		SI		Ic		LWA		AF		OF		Overall	
	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50	P@5	P@50
S-Dop vig	88.6	86.3	70.2	62.2	67.2	55.1	87.6	83.6	89.2	87.5	94.6	92.6	67.6	54.4	95.2	94.9	65.2	54.2	43.6	27.2	76.9	69.8
S-Dop Subap	96.2	94.64	94.2	92.0	97.0	96.4	96.2	95.5	98.4	97.0	97.6	97.0	81.6	73.3	98.2	97.1	91.8	89.8	62.6	47.8	91.3	88.0
U-Dop vig	84.2	76.3	58.0	42.8	43.2	27.3	43.4	23.6	61.2	46.1	47.0	26.7	32.6	17.2	79.0	73.5	35.8	16.9	27.4	9.6	51.1	36.0
U-Dop Subap	90.4	83.5	74.4	63.1	59.4	47.0	55.4	40.3	75.0	64.8	61.4	42.9	55.6	36.2	85.2	69.2	58.8	36.8	52.2	36.0	66.7	52.0

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 data set, the algorithm has a huge impact in 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 III 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.

Additionally, we showed the retrieval results for the unsupervised embeddings trained on subapertures for two query images: in Fig. 5 for pure ocean waves class and in Fig. 6 for atmospheric front 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 from Fig. 5 and Fig. 6), which can indicate the presence of two phenomena in the same location or other intrinsic similarities.

VI. CONCLUSION

In this work, we extended the previous approach from [15] by using the SD algorithm for unsupervised feature learning with transformer networks. The unsupervised features were used for a 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 in retrieval performance when more physics based algorithms, as DCE, are used for ocean retrieval. This experiment allow us to build more complex searching engines, which could find similar physical parameters, instead of similar structures (e.g., ocean currents speed). Summing up, we used a data-centring approach to improve the performance classification and retrieval algorithms, in both supervised and unsupervised settings.

ACKNOWLEDGMENT

This work was supported by IFREMER and by the grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P4-ID-PCE-2020-2120, within PNCDI III.

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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 8, pp. 3876–3891, 2015.

[2] R. H. Stewart, "Seasat: results of the mission," *Bulletin of the American Meteorological Society*, vol. 69, no. 12, pp. 1441–1447, 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," *Geophysical Research Letters*, 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 Transactions on Geoscience and Remote Sensing*, vol. 51, no. 11, pp. 5145–5159, 2013.

[5] G. Sumbul and B. Demir, "Plasticity-stability preserving multi-task learning for remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, 2022.

[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 Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 4743–4750, 2011.

[7] G. Zheng, X. Li, L. Zhou, J. Yang, L. Ren, P. Chen, H. Zhang, and X. Lou, "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 Transactions on Geoscience and Remote Sensing*, vol. 56, no. 9, pp. 5244–5260, 2018.

[8] X. Li, B. Liu, G. Zheng, Y. Ren, S. Zhang, Y. Liu, L. Gao, Y. Liu, B. Zhang, and F. Wang, "Deep-learning-based information mining from ocean remote-sensing imagery," *National Science Review*, vol. 7, no. 10, pp. 1584–1605, 2020.

[9] S. R. Dubey, "A decade survey of content based image retrieval using deep learning," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[10] B. Neyshabur, S. Bhojanapalli, D. McAllester, and N. Srebro, "Exploring generalization in deep learning," *Advances in neural information processing systems*, vol. 30, 2017.

[11] F. Ye, W. Luo, M. Dong, H. He, and W. Min, "Sar image retrieval based on unsupervised domain adaptation and clustering," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1482–1486, 2019.

[12] X. Tang, X. Zhang, F. Liu, and L. Jiao, "Unsupervised deep feature learning for remote sensing image retrieval," *Remote Sensing*, vol. 10, no. 8, p. 1243, 2018.

[13] 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 8, pp. 3876–3891, 2015.

[14] Y. Liu, L. Ding, C. Chen, and Y. Liu, "Similarity-based unsupervised deep transfer learning for remote sensing image retrieval," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 11, pp. 7872–7889, 2020.

[15] D. Ristea, Anghel, "Guided deep learning by subaperture decomposition: ocean patterns from sar imagery," in *Proceedings of IGARSS*. IEEE, 2021, pp. 4067–4070.

[16] Z. Wang, X. Fu, and K. Xia, "Target classification for single-channel sar images based on transfer learning with subaperture decomposition," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2020.

[17] C. Brekke, S. N. Anfinsen, and Y. Larsen, "Subband extraction strategies in ship detection with the subaperture cross-correlation magnitude," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 4, pp. 786–790, 2013.

[18] J. Singh, M. Soccorsi, and M. Datcu, "Sar complex image analysis: A gauss markov and a multiple sub-aperture based target characterization," in *Proceedings of IGARSS*, 2010, pp. 1585–1588.

[19] A. Focsa, M. Datcu, S. Toma, A. Anghel, and R. Cacoveanu, "Opportunistic bistatic sar image classification using sub-aperture decomposition," in *Proceedings of COMM*, 2020, pp. 203–207.

[20] Z. Lu, G. Hua-dong, and H. Chun-ming, "Sar ocean stationary targets detection," *Remote Sensing Technology and Application*, vol. 22, no. 3, pp. 321–325, 2011.

[21] R. Raney, "Doppler properties of radars in circular orbits," *International Journal of Remote Sensing*, vol. 7, no. 9, pp. 1153–1162, 1986.

[22] S. N. Madsen, "Estimating the doppler centroid of sar data," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 25, no. 2, pp. 134–140, 1989.

[23] F. Wong and I. G. Cumming, "A combined sar doppler centroid estimation scheme based upon signal phase," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 34, no. 3, pp. 696–707, 1996.

[24] P. López-Dekker, J. J. Mallorqui, P. Serra-Morales, and J. Sanz-Marcos, "Phase synchronization and doppler centroid estimation in fixed receiver bistatic sar systems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3459–3471, 2008.

[25] S. Zhang, Y. Gao, M. Xing, R. Guo, J. Chen, and Y. Liu, "Ground moving target indication for the geosynchronous-low earth orbit bistatic multichannel sar system," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 5072–5090, 2021.

[26] M. W. Hansen, F. Collard, K.-F. Dagestad, J. A. Johannessen, P. Fabry, and B. Chapron, "Retrieval of sea surface range velocities from envisat asar doppler centroid measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3582–3592, 2011.

[27] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of NIPS*, 2017, pp. 5998–6008.

[28] H. Dong, L. Zhang, and B. Zou, "Exploring vision transformers for polarimetric sar image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2021.

[29] 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 Sensing*, vol. 14, no. 9, p. 2085, 2022.

[30] 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 Sensing*, vol. 14, no. 13, p. 3186, 2022.

[31] N.-C. Ristea, A.-I. Miron, O. Savencu, M.-I. Georgescu, N. Verga, F. S. Khan, and R. T. Ionescu, "Cytran: Cycle-consistent transformers for non-contrast to contrast ct translation," *arXiv preprint arXiv:2110.06400*, 2021.

[32] C. Leng, H. Zhang, B. Li, G. Cai, Z. Pei, and L. He, "Local feature descriptor for image matching: A survey," *IEEE Access*, vol. 7, pp. 6424–6434, 2018.

[33] L. Koteswara Rao, P. Rohini, and L. Pratap Reddy, "Local color oppugnant quantized extrema patterns for image retrieval," *Multidimensional Systems and Signal Processing*, vol. 30, no. 3, pp. 1413–1435, 2019.

[34] A. K. Bedi and R. K. Sunkaria, "Local tetra-directional pattern–a new texture descriptor for content-based image retrieval," *Pattern Recognition and Image Analysis*, vol. 30, no. 4, pp. 578–592, 2020.

[35] A. Bellet, A. Habrard, and M. Sebban, "A survey on metric learning for feature vectors and structured data," *arXiv preprint arXiv:1306.6709*, 2013.

[36] J. Wang, T. Zhang, N. Sebe, H. T. Shen *et al.*, "A survey on learning to hash," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 769–790, 2017.

[37] M. Kaya and H. Ş. Bilge, "Deep metric learning: A survey," *Symmetry*, vol. 11, no. 9, p. 1066, 2019.

[38] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proceedings of ACMMM*, 2014, pp. 157–166.

[39] F. Ye, W. Luo, M. Dong, H. He, and W. Min, "Sar image retrieval based on unsupervised domain adaptation and clustering," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1482–1486, 2019.

[40] W. Xiong, Z. Xiong, Y. Zhang, Y. Cui, and X. Gu, "A deep cross-modality hashing network for sar and optical remote sensing images retrieval," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 5284–5296, 2020.

[41] C. Wang, A. Mouche, P. Tandeo, J. E. Stopa, N. Longepe, G. Erhard, R. C. Foster, D. Vandemark, and B. Chapron, "A labelled ocean sar imagery dataset of ten geophysical phenomena from sentinel-1 wave mode," *Geoscience Data Journal*, vol. 6, no. 2, pp. 105–115, 2019.

[42] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85–112, 2020.

[43] C. Wang, P. Tandeo, A. Mouche, J. E. Stopa, V. Gressani, N. Longepe, D. Vandemark, R. C. Foster, and B. Chapron, "Classification of the global sentinel-1 sar vignettes for ocean surface process studies," *Remote Sensing of Environment*, vol. 234, p. 111457, 2019.

[44] B. Quach, Y. Glaser, J. E. Stopa, A. A. Mouche, and P. Sadowski, "Deep learning for predicting significant wave height from synthetic aperture radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 1859–1867, 2020.

[45] A. Colin, C. Peureux, R. Husson, N. Longépé, R. Rauzy, R. Fablet, P. Tandeo, S. Saoudi, A. Mouche, and G. Dibarboure, "Segmentation of sentinel-1 sar images over the ocean, preliminary methods and assessments," in *Proceedings of IGARSS*. IEEE, 2021, pp. 4067–4070.

[46] I. De Gelis, A. Colin, and N. Longépé, "Prediction of categorized sea ice concentration from sentinel-1 sar images based on a fully convolutional network," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of CVPR*, 2016, pp. 770–778.

[48] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of CVPR*, 2016, pp. 2818–2826.

[49] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of ICML*. PMLR, 2015, pp. 448–456.

[50] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of ICML*, 2010.