Learning-based SAR-Optical Registration for Navigation: Insights from a Multi-Year, Multi-Season Continental-Scale Dataset

Simon Bertrand, Guillaume Bourmaud, Cornelia Vacar, Lionel Bombrun

Learning-based SAR-Optical Registration for Navigation: Insights from a Multi-Year, Multi-Season Continental-Scale Dataset

Simon Bertrand, Guillaume Bourmaud, Cornelia Vacar and Lionel Bombrun

***Abstract*—This paper addresses the problem of multi-modal image registration between in-flight Synthetic Aperture Radar (SAR) imagery and high-resolution optical reference maps, a key challenge for autonomous navigation in GPS-denied environments. Despite recent advances in deep learning-based registration, robustness under severe sensor degradation and generalization across diverse real-world geographies remain underexplored. To tackle this issue, we introduce a continental-scale dataset over Europe, covering 4.9 million km^2 with multi-temporal and multi-seasonal SAR and optical imagery. Two experimental protocols are proposed: spatial separation and temporal separation. To simulate a terrain-aided navigation scenario, SAR images are corrupted with Gaussian blur and speckle noise. We evaluate four recent deep neural networks and demonstrate the benefits of combining their backbones and loss functions to improve registration accuracy. Detailed experiments on spatial and temporal separation protocols indicate that the right combination of architectural elements can lead to performance gains exceeding 120% in severely degraded conditions. In particular, our findings reveal that the pseudo-Siamese OSMNet network, when trained with a cross-entropy loss, demonstrates the highest robustness.**

***Index Terms*—Remote sensing, image registration, SAR, optical, multimodal fusion, deep learning, big data.**

I. Introduction

A well-known challenge of purely inertial navigation systems is the drift that accumulates over long trajectories due to the integration of measurement errors. To address this issue, strapdown navigation is commonly integrated with Global Positioning System (GPS) measurements [1], [2]. However, in GPS-denied regions, additional sensors must be considered, as suggested in [3] and [4]. Amongst these approaches, an effective solution to mitigating drift is the use of terrain information [5], [6]. In terrain-aided navigation (TAN), electromagnetic sensors are preferred due to their robustness to varying illumination and weather conditions. Moreover, SAR sensors provide high-resolution images that can be compared to geo-referenced maps, enabling precise localization of the carrier [7].

This paper focuses on the image registration process between SAR images acquired during flight and optical reference images, which is crucial for TAN systems. Unlike most existing studies in the literature, the SAR-optical image registration method approach presented here is specifically tailored for navigation applications, while still remaining broadly applicable to remote sensing tasks.

During extended missions, accumulated drift from strapdown navigation can result in significant positional uncertainty for the UAV. To accommodate this uncertainty, the optical reference image must be sized appropriately to ensure complete coverage of the acquired SAR image regardless of the UAV's actual position within the uncertainty bounds. Fig. 1 illustrates this concept and demonstrates the operation of the SAR-optical image alignment system.

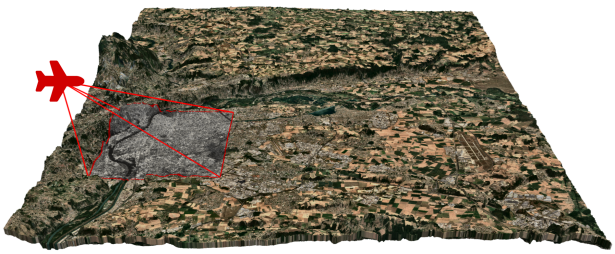


Fig. 1. Example of the SAR-optical image registration principle. The radar-equipped UAV (red shape) is acquiring a SAR image (outlined in red) with the purpose of improving navigation precision. In order to aid navigation, the acquired SAR image must be accurately registered within the georeferenced optical image (background).

Moreover, we aim to assess the robustness of SAR-optical image matching under various challenging conditions. First, we evaluate how SAR image quality from the embedded sensor affects registration performance. Second, we examine the generalization capabilities of matching algorithms by investigating performance on geographically distinct areas not included in the training data (spatial separation) and on temporally mismatched image pairs where the optical reference is outdated or differs seasonally from the SAR acquisition (temporal separation).

To conduct these experiments, we require a dataset of registered SAR-optical image pairs with extensive geographic coverage and multi-temporal acquisitions. Since existing SAR-optical datasets lacked these characteristics, we develop a new comprehensive dataset. This dataset serves as a benchmark for evaluating four state-of-the-art SAR-optical image registration methods and optimizing their performance.

The main contributions of this paper are:

- **Novel dataset creation**: Development of a comprehensive SAR-optical dataset featuring multi-seasonal, multi-annual acquisitions with extensive geographic coverage and substantial data volume, specifically designed for

S. Bertrand is with both CEA CESTA - French Alternative Energies and Atomic Energy Commission, Le Barp, France and Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France (e-mail: simon.bertrand@cea.fr)

G. Bourmaud is with Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France (e-mail: guillaume.bourmaud@u-bordeaux.fr)

C. Vacar is with CEA CESTA - French Alternative Energies and Atomic Energy Commission, Le Barp, France (e-mail: cornelia.vacar@cea.fr)

L. Bombrun is both with Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400 Talence, France and Bordeaux Sciences Agro, F-33175 Gradignan, France (e-mail: lionel.bombrun@u-bordeaux.fr)

navigation-oriented image registration studies, but perfectly adapted for remote sensing applications as well.
- **Robustness analysis**: Systematic evaluation of registration algorithm performance under degraded SAR image quality conditions with varying signal-to-noise ratios.
- **Comprehensive ablation study**: Analysis of different backbone architectures and loss functions to determine their individual contributions to performance and identify optimal model configurations.
- **Geometric constraint analysis**: Investigation of registration performance limitations when extending from pure translation to include rotational transformations of the SAR template.

The following sections are organized as follows. Section II reviews related works and datasets in SAR-optical image registration. Section III introduces our new dataset. Section IV provides a review of the chosen state-of-the-art methods. In Section V, we are presenting the evaluation protocols, the matching performance for various signal-to-noise ratios for the SAR image and conduct a sensitivity study to isolate key architectural components impacting accuracy. Then, we address in Section VI some additional questions on SAR-optical registration beyond the case of translations-only. And finally, Section VII summarizes the main conclusions and outlines directions for future works.

II. Related Work

A. SAR-optical image registration

Registering SAR and optical images is a longstanding challenge due to significant disparities in imaging acquisition geometry, radiometry and noise. Seasonal variations further complicate the registration process. These difficulties are particularly critical in TAN, where accurate cross-modal alignment underpins reliable geolocation and autonomous navigation in GPS-denied environments.

Early methods focused on intensity-based registration using cross mutual information (CMIF) [8], [9], sum of squared differences (SSD) [10], phase congruency [11] or oriented histogram-based similarity metrics [12]. However, their sensitivity to radiometric inconsistencies limited applicability in real-world terrain scenarios. Feature-based methods, such as SIFT and its SAR-adapted variants (e.g., SAR-SIFT [13], OS-SIFT [14], Improved SIFT [15], I-SAR-SIFT [16]), attempted to improve robustness to modality-specific noise. Yet, due to sparse keypoints and inconsistent cross-modal descriptors, these approaches often failed in complex or low-texture terrain. To overcome these limitations, dense registration methods have gained traction. Subpixel correlation [17] and optical flow adaptations [18] offer performance improvements, particularly in high-resolution or topographically variable scenes. More recently, deep learning-based approaches have significantly advanced SAR-optical registration. Beyond traditional feature-learning frameworks, generative paradigms have emerged as powerful tools for mitigating the modality gap. Adversarial approaches, such as GANs [19], [20], [21], have been widely employed to reduce cross-modal discrepancies, while more general generative formulations, including diffusion models [22], show strong potential for enhancing alignment across heterogeneous domains. Other generative strategies, such as the Confucius tri-learning paradigm [23], further exploit multi-view consistency to strengthen representation learning. In parallel, modality fusion techniques have been investigated as an alternative route to overcome multi-modal challenges, as demonstrated in hyperspectral data applications [24], [25]. Complementary to these efforts, certain frameworks [26] advance open set recognition by reasoning over inter-feature relationships rather than relying solely on global feature descriptors. Within convolutional neural networks (CNNs), extracting modality-invariant representations through cross-comparison of SAR and optical imagery has become a standard practice [17]. To better capture structural correspondences under severe radiometric and geometric distortions, advanced CNN designs have introduced co-attention mechanisms [27] and pseudo-Siamese architectures [28]. More recently, in the domain of dense matching, SFCNet [29] incorporated modality-specific Siamese branches tailored for SAR and optical inputs, thereby enhancing cross-modal discrimination and improving registration accuracy.

Furthermore, several methods have specifically targeted high-resolution, multiscale registration. MCGF [30] proposed multiscale convolutional gradient features to improve matching under radiometric distortion. Similarly, OSMNet [31] explored optimized deep architectures for high-resolution SAR-optical matching, emphasizing structural consistency and robustness to environmental variation. Siamese U-Net with FFT correlation [32] leverages both spatial and frequency domain cues for robust alignment. MARU-Net [33] introduces multiscale attention gating and contrastive loss to enhance feature discrimination. Our study focuses on these four dense architectures, aiming to evaluate and optimize their performance and robustness in challenging SAR-optical registration scenarios.

B. Datasets for SAR-optical image registration

A significant number of SAR-optical databases have been proposed recently. The most notable among them is SEN12MS [34], which provides $180,662$ geolocated patches containing the dual-polarized SAR data (VV, VH), corresponding optical imagery and land cover classification maps. A similar large-scale dataset is BigEarthNet V2.0 [35], which includes $549,488$ registered SAR-optical pairs, each covering patches up to 120px in size, acquired by the Sentinel-1 and Sentinel-2 satellites. In addition to these well-established datasets, more recent contributions such as QXS-SAROPT [36] and WHU-OPT-SAR [37] provide multimodal SAR-optical pairs featuring fully polarimetric SAR data (collected by the GaoFen-3 satellite) alongside high-resolution optical images, primarily targeting urban and coastal monitoring applications. SARptical [38] focuses on high-resolution airborne SAR and optical image pairs collected over urban areas, offering multi-temporal data that captures dynamic urban environments but with limited geographic scope. Similarly, So2Sat [39] provides multi-temporal Sentinel-1 SAR and Sentinel-2 optical images over European urban regions,

TABLE I
COMPARISON OF PUBLICLY AVAILABLE DATASETS CONTAINING SAR-OPTICAL IMAGE PAIRS.

Reference	Data					Volumetrics			Task
	Source	SAR Mode	Multi Temporality Year / Season	Spatial extent (Mkm^2)	Spatial resolution ($\mathrm{m/px}$)	Image size (px)	#Locations	#Pixels ($\times 10^9$)	
SEN12MS [41]	S1, S2	VV, VH	✗ / ✓	1.2	10	256×256	180,662	71	Registration
BigEarthNet [35]	S1, S2	Single	✗ / ✗	0.8	10	120×120	549,488	31	Classification
QXS-SAROPT [36]	GF3	Single	✗ / ✗	$1.3 \cdot 10^{-3}$	1	256×256	20,000	0.02	Registration Ship detection
WHU-OPT-SAR [37]	GF1, GF3	Single	✗ / ✗	0.05	5	**5 556×3 704**	100	10	Classification
MultiResSAR [42]	S1, GF3 HT1-A, Umbra	Various	✗ / ✗	-	0.16-10	Various	10,850	-	Registration
MultiSenGE [43]	S1, S2	VV, VH	✗ / ✓	0.05	10	256×256	8,157	7	Classification
SSL4EO-S12 [40]	S1, S2	VV, VH	✗ / ✓	1.7	10	264×264	251,079	1,890	Self-supervised learning
SARptical [38]	TerraSAR-X UltraCAM	Single	✗ / ✗	10^{-4}	{1,9}	112×112	10,000	0.5	3-D reconstruction Registration
So2Sat [39]	S1, S2	VV, VH	✗ / ✗	0.04	10	32×32	400,673	7	Classification
MEOW (Ours)	S1, S2 TanDEM-X	VV, VH	✓ / ✓	**4.9**	10	**4 096×4 096**	2,917	**4,013**	Registration (for navigation) Classification

TABLE II
OVERVIEW OF THE DIFFERENT MODALITIES OF THE MEOW DATASET.

Mode	Features				
	Resol.	Format	#Channels	Source	Raw bytes size
Grid	10 $\mathrm{m/px}$	f32	2 (Long, Lat)	GEE	6,264 GB
SAR	10 $\mathrm{m/px}$	f32	2 (VV, VH)	Sentinel 1	6,264 GB
Opt	10 $\mathrm{m/px}$	uint8	3 (R, G, B)	Sentinel 2	2,349 GB
DEM	30 $\mathrm{m/px}$	f16	1 (DEM)	TanDEM-X	98 GB
Cl	10 $\mathrm{m/px}$	uint8	1 (CL)	Sentinel 1 & 2	49 GB

emphasizing urban mapping and classification with good temporal coverage but limited spatial extent. The SSL4EO-S12 dataset [40] significantly advances the quantitative boundaries in the field by offering a high-volume collection of multi-temporal samples that include all channels from the Sentinel-1 & 2 missions. Table I summarizes the key characteristics of these datasets.

Together, these resources span a broad range of spatial resolutions, geographic coverages and temporal scopes. Nevertheless, the lack of a single, unified dataset combining high spatial resolution, extensive temporal coverage (multi-year and multi-season) and diverse geographic representation continues to pose challenges for generalizable deep learning in SAR-optical registration, particularly for analysing the spatial and temporal robustness of the SAR-optical matching performance. To address this gap, we introduce in the following section a custom dataset, named MEOW for Multimodal Earth Observation Warehouse, tailored to the specific needs of SAR-optical image registration for navigation.

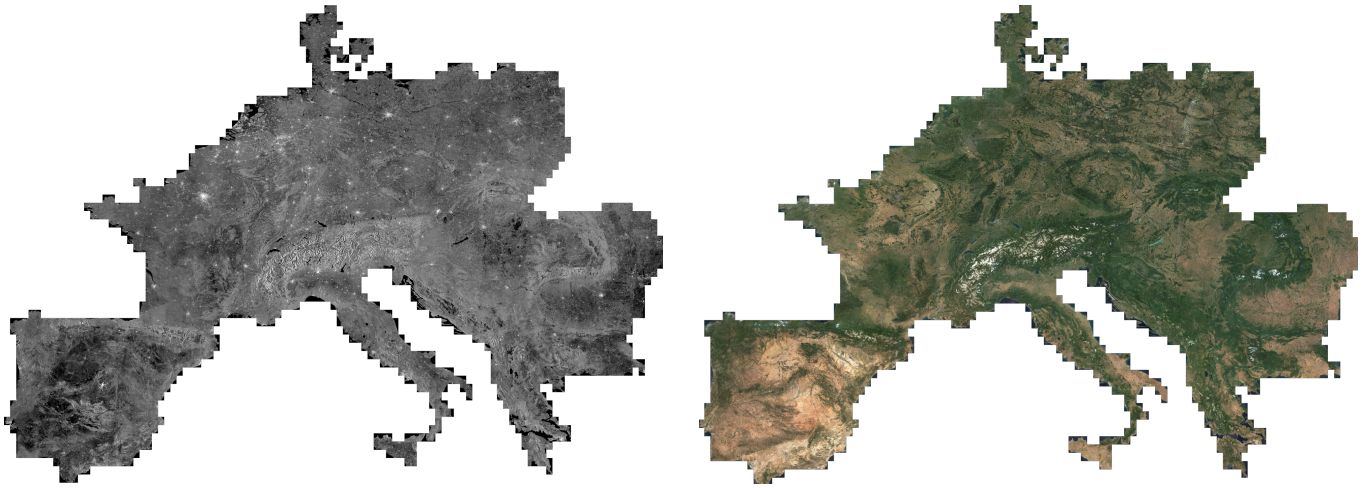
III. MEOW DATASET

A. Dataset description

The proposed MEOW dataset is specifically designed for SAR-optical image registration in the context of navigation. We provide a dataset acquired over Europe, which serves as benchmark for the present study. Table II gives an overview of the main characteristics of this dataset. It is a large-scale dataset in which each image has a dimension of $4\,096 \times 4\,096$ pixels at a spatial resolution of $10\mathrm{m}$. It is designed for dense regional coverage, ensuring seamless tiling with adjacent images and zero overlap. The dataset includes co-registered SAR and optical images. The dual-polarization (VV and VH) SAR images are sourced from the Sentinel-1 sensor, while the optical RGB images are obtained from Sentinel-2. A total of $2,917$ locations are considered across Europe, resulting in a spatial extent of $4.9\mathrm{Mkm}^2$. To assess the influence of temporal (seasonal and inter-annual) variability, SAR and optical images are provided for the four seasons (winter, spring, summer, and autumn) and across four years (2018, 2020, 2022, and 2023). Fig. 2a and 2b illustrate the geographical extent of the MEOW dataset for SAR (VV channel) and optical data acquired during summer 2023, respectively. Fig. 3 shows a time series of optical and SAR image pairs acquired over the Bordeaux region in France. Fig. 3a illustrates the annual variations by visualizing optical and SAR images acquired in summer over 4 different years: 2018, 2020, 2022 and 2023. Fig. 3b illustrates the seasonal variability, with image pairs acquired during the four seasons of 2022. Changes can be noticed especially in the forested area (at the bottom of the image) due to massive wildfires that occurred in summer 2022. Changes can equally be noticed in summer 2023, where areas affected by wildfires have turned into clear-cuts.

To ensure that the proposed dataset can be leveraged by the scientific community for applications beyond SAR-optical image registration, additional modalities are also provided. For instance, a Digital Elevation Model (DEM) derived from TanDEM-X mission is included, with a spatial resolution of $30\mathrm{m}$ [44], [45], [46]. The land cover maps from ESA WorldCover v200 based on Sentinel-1 and Sentinel-2 data are also available [47]. They offer semantic information about the observed scenes on 11 classes including tree cover, grassland, cropland, permanent water bodies, etc. Furthermore, the latitude and longitude coordinates of each pixel are recorded, enabling precise geospatial analysis and integration with other geographic datasets.

As observed in Table I, the proposed MEOW dataset is



(a) SAR images over Europe during summer 2023.

(b) Optical images over Europe during summer 2023.

Fig. 2. Multimodal Earth Observation Warehouse (MEOW) dataset: SAR and optical images acquired over Europe during summer 2023 (WGS84). MEOW provides access to 16 images across different timestamps (4 seasons $\times$ 4 years), all at a resolution of $10\mathrm{m}$ per pixel. Each square block corresponds to a high-resolution $4\,096 \times 4\,096$ pixel image.

the largest one for SAR-optical image registration in terms of spatial extent ($4.9\mathrm{Mkm}^2$), number of pixels ($4,013 \times 10^9$), image size (4096×4096) at 10m/px resolution and multi-temporal diversity (multi-year and multi-season). Our dataset offers several unique advantages compared to the existing ones. First, it covers a continuous spatial region, which facilitates dense learning and detailed spatial analysis. By contrast, similar datasets rely on sparse spatial sampling, making them less practical for navigation applications. Second, each sample covers approximately $1700\mathrm{km}^2$ at 10 m/pixel resolution, to the best of our knowledge, the largest per-sample coverage in existing datasets. This extensive coverage allows flexible resampling to multiple coarser resolutions, not supported by other datasets. Third, our dataset provides multi-year and multi-season coverage, with data collected across four seasons over four different years. This temporal diversity enables studies of algorithm generalization over time and provides valuable insights presented in subsequent sections. Additionally, the dataset comes with a Google Earth Engine generation script, enabling users to create customized datasets tailored to their regions and operational contexts. Finally, it uniquely integrates Digital Elevation Models and land cover classes, facilitating SAR signal simulation and broadening its applicability to multi-modal analysis and land segmentation tasks.

B. MEOW extraction process

Fig. 4 illustrates the extraction process for the creation of the MEOW dataset. For that, a script is created and publicly shared to generate the large scale geospatial dataset using Google Earth Engine (GEE). It consists of four main steps as detailed below.

1) Sources: As previously stated, MEOW is a multimodal dataset with images extracted from the GEE catalog. SAR and optical modalities originate respectively from Sentinel-1 and Sentinel-2 missions by Copernicus. Land cover classes are derived by the European Spatial Agency (ESA) from Sentinel data and DEM is issued from the TanDEM-X mission.

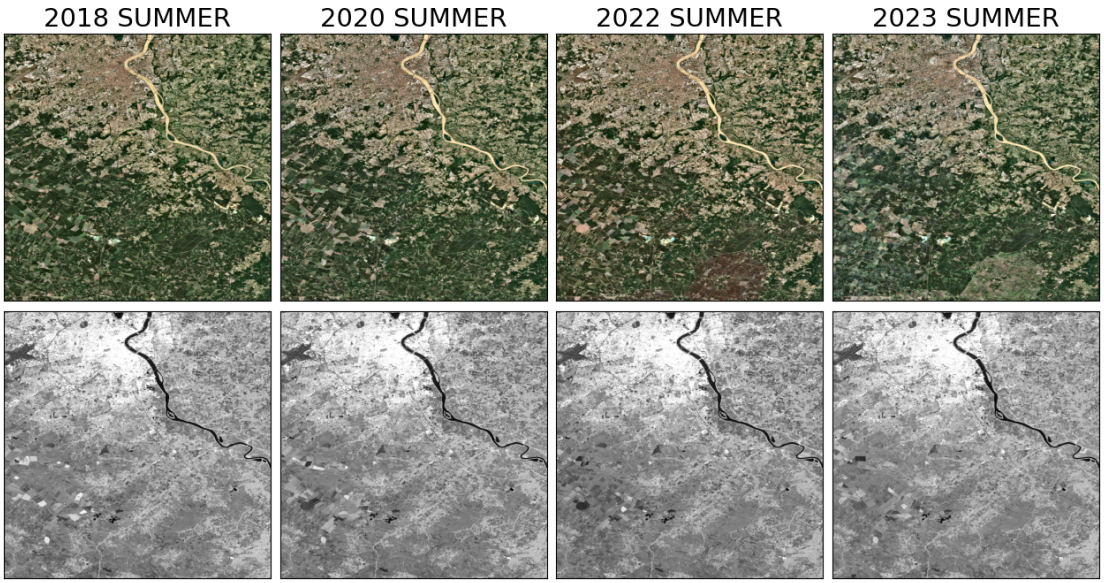
2) Grid generation: A customized and coarse GeoJSON mask is created to outline the boundaries of the area of interest. For the MEOW dataset, this region corresponds to the countries delineated in Fig. 2. The covering grid is generated using the $coveringGrid$ method from GEE. Each cell is defined by 4×4 square cells of $1\,024 \times 1\,024\mathrm{px}$ at $10\mathrm{m/px}$, resulting after recombination in images of $4\,096 \times 4\,096\mathrm{px}$.

3) Cell filtering: Once the grid is generated, cells are filtered according to specific criteria. First, images should be available for the considered cell. Second, cells containing a large amount of sea are discarded since for the considered TAN application only regions located on land surfaces are of interest. To this end, the mask of sea regions issued from Hybrid Coordinate Ocean Model (HYCOM) is employed.

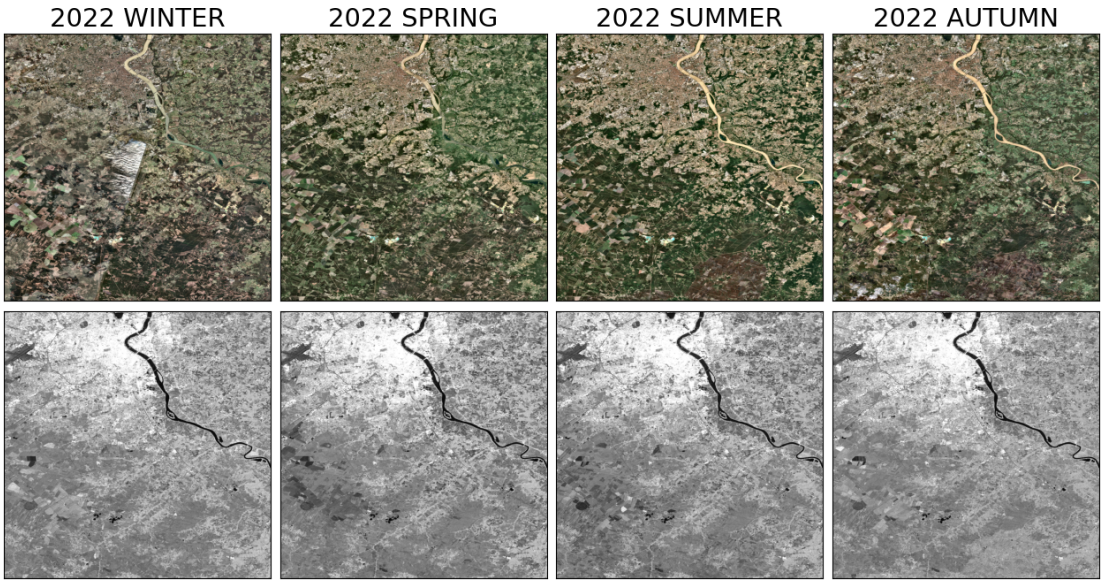
4) Data Processing: Sentinel data undergo Copernicus Level-2 processing, including orthorectification and sub-pixel registration. In addition, SAR images in VV and VH polarizations are clamped between -25 and $5\mathrm{dB}$. Optical data are rescaled from $[0 - 10,000]$ to $[0 - 2,700]$, gamma-corrected ($\gamma = 1.3$) and loosely mapped to 8-bit integers ($2,700 \rightarrow 255$). The DEM, originally at a spatial resolution of $30\mathrm{m/px}$, is interpolated to match the $10\mathrm{m/px}$ resolution.

Images from the same year and same season are averaged using a median filter. This effectively reduces the speckle noise for SAR images. Note that for optical data, images with more than 40% cloud cover are discarded from this process.

The MEOW dataset is publicly available for academic research at: $http : //github.com/Simon - Bertrand/SAROPTReg - Insights$. The dataset can be customized for any region by specifying the corresponding GeoJSON file, and image dimensions can be adjusted by configuring the script parameters accordingly. In addition, several Python objects are provided, including the data loader, class descriptions, color maps for visualization.



(a) Optical and SAR images acquired during the summer of 2018, 2020, 2022 and 2023. The area affected by the fire is visible on the optical image from the summer 2022 then in 2023, on the bottom of the image.



(b) Optical and SAR images from the different seasons of 2022. The inter-season variability is obvious on both optical and SAR imagery, significant in the areas affected by the fire, but also in other areas due to crop and soil humidity differences.

Fig. 3. Focus on a region in Gironde, France. Annual variations are illustrated by visualizing optical and SAR images acquired in summer over 4 different years, while the seasonal variability is illustrated on the year 2022. This area has been subject to massive wildfires in the summer of 2022, inducing landscape changes visible on both optical and SAR imagery.

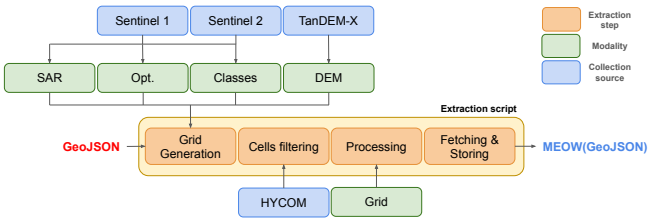


Fig. 4. Overview of the extraction pipeline of the MEOW dataset.

Now that the proposed MEOW dataset is introduced, the next section presents an application on SAR-optical image registration.

IV. SAR-optical image registration

A. General principle of SAR-optical image registration

Fig. 5 illustrates the general principle of learning-based approaches [30], [31], [32], [33] for 2D translational registration between a reference optical image $\mathtt{I}_{\mathrm{O}}$ of size H_{O}xW_{O}x3 and a SAR patch $\mathtt{I}_{\mathrm{S}}$ of size H_{S}xW_{S}x1. In practice, the SAR patch is smaller than the optical image, and the objective is to estimate the 2D translation that aligns the SAR patch with the corresponding region in the optical image. First, descriptors $\mathtt{D}_{\mathrm{O}}$

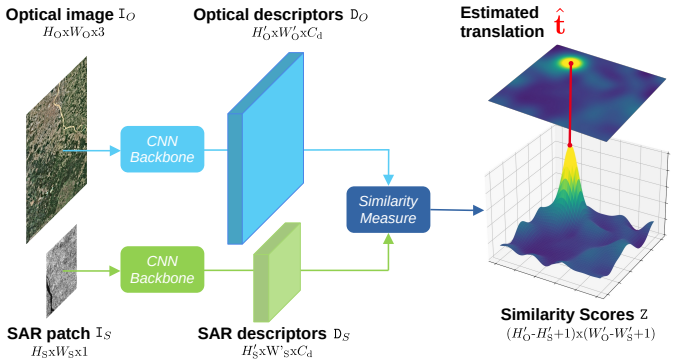


Fig. 5. Principle of learning-based SAR-optical image registration. Descriptors are extracted from both optical and SAR images using a CNN backbone with either Siamese or pseudo-Siamese architecture. These descriptors are compared using a similarity measure. The estimated translation is selected as the one yielding the maximum similarity score between the optical and SAR descriptors.

of size H'_OxW'_OxC_d are extracted from the optical image using a CNN backbone. Similarly, descriptors $\mathtt{D}_S$ of size H'_SxW'_SxC_d are extracted from the SAR patch. A similarity measure, such as a cross-correlation, is used to compute a 2D heatmap $\mathtt{Z}$ of size (H'_O-H'_S+1)x(W'_O-W'_S+1) of similarity scores between the SAR descriptors and the optical descriptors. The location of the maximum value in this heatmap $\mathtt{Z}$ indicates the estimated 2D translation $\hat{\mathbf{t}}$ that best aligns the SAR patch with the optical image:

$$\hat{\mathbf{t}} = \arg\max_{u,v} \mathtt{Z}(u,v). \tag{1}$$

When the same CNN backbone with shared weights is used to compute descriptors for both the optical image and the SAR patch, the registration pipeline is referred to as Siamese. If the backbones are identical but the weights are not shared, it is referred to as pseudo-Siamese.

To evaluate the registration accuracy, a supervised approach is considered. For that, the dataset is first split into two parts: a training set and a testing set. SAR-Optical registration models are trained on the training set and evaluated on the testing set. To assess the registration accuracy of a given model, the Correct Matching Rate (CMR) [29] is evaluated on the testing set composed of N optical/SAR pairs [1]. For a given tolerance threshold r, the CMR is computed as follows:

$$\text{CMR}(r) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{(\|\hat{\mathbf{t}}_n - \mathbf{t}_n^{\text{GT}}\|_2 \leq r)} \tag{2}$$

where $\mathbb{1}$ is the indicator function and $\|\cdot\|_2$ is the ℓ_2 norm. The CMR represents the proportion of registration attempts for which the estimated translation $\hat{\mathbf{t}}$ lies within a radius r of the ground truth translation $\mathbf{t}^{\text{GT}}$. Its value ranges between 0 and 1, with values closer to 1 indicating a better model.

In the following, four recent state-of-the-art deep neural networks (MCGF [30], OSMNet [31], FFT U-Net [32] and MARU-Net [33]) are considered and evaluated on the proposed MEOW dataset. Each of these models are presented in the next subsection.

[1]Given that the standard deviation of the estimated CMR is small (approximately 10^{-3}) relative to the variability across models, confidence intervals for the CMR are omitted in the results presented in the paper.

B. Comparison of state-of-the-art network backbones

Table III presents a comparison of the backbones of the four studied SAR-optical image registration models (MCGF [30], OSMNet [31], FFT U-Net [32] and MARU-Net [33]) based on their architecture type (Siamese or pseudo-Siamese), the number of parameters, the computational time and the memory footprint.

MCGF [30] is a lightweight pseudo-Siamese architecture with only 58k parameters, a memory footprint of 120 MB, and a forward pass time of 19.3 ms, making it highly efficient. It begins by extracting multi-oriented gradient features that capture structural patterns within each image. These features are then processed by a shallow pseudo-Siamese network operating at multiple scales, resulting in multiscale convolutional gradient features.

OSMNet [31] is a pseudo-Siamese architecture with the largest memory footprint (579 MB), but only 653k parameters, making it less prone to overfitting. Its forward pass time is 22% slower than MCGF, but comparable to FFT U-Net and MARU-Net. OSMNet effectively captures both high-level semantic information and low-level fine-grained details, which are crucial for accurate feature matching.

FFT U-Net [32] is a Siamese architecture with 1.9M parameters and a large memory footprint (478 MB). It is based on the classical U-Net [48] with shared weights, and is designed to capture global context while preserving fine spatial details and precise positional information.

MARU-Net [33] is a Siamese, multiscale, attention-gated residual U-Net [49]. Its memory footprint remains limited (287 MB), but with 8.9M parameters, it is by far the largest model.

C. Comparison of state-of-the-art network losses

Table III provides a comparative overview of the loss functions used by the four models: MCGF [30], OSMNet [31], FFT U-Net [32], and MARU-Net [33]. The comparison is based on the similarity measure, computation time, type of loss, and the positive mask, *i.e.*, the neighborhood expected to activate around the ground-truth pixel in the heatmap $\mathtt{Z}$.

MCGF [30] relies on a simple Cross-Correlation (CC) similarity measure, normalized by the size of the SAR descriptors, and uses a contrastive loss with a 5-pixel cross-shaped positive mask.

OSMNet [31] is the only model that does not rely on a form of cross-correlation. Instead, it uses a Sum of Squared Differences (SSD) measure, normalized by the size of the SAR descriptors. It also introduces a Triplet loss, which proves suboptimal, as both our experimental results and theoretical analysis (see Appendix A) demonstrate.

FFT U-Net [32] uses a Zero-Normalized Cross-Correlation (ZNCC) similarity measure with a temperature factor (γ), combined with a standard Cross-Entropy (CE) loss and a Dirac positive mask.

MARU-Net [33] employs the same similarity measure as FFT

TABLE III
COMPARISON OF BACKBONE ARCHITECTURES AND LOSS FUNCTIONS.

Model	Backbone				Loss function			
	Type	# Param.	Forward pass (ms)	Memory (MB)	Sim. scores (Z)	Sim. time (ms)	Loss type	Positive mask
MCGF [30]	Pseudo-Siam.	58k	19.3	120	$\frac{\mathrm{CC}(\mathrm{D_O},\mathrm{D_S})}{H'_S W'_S}$	3.7	Contrastive	
OSMNet [31]	Pseudo-Siam.	653k	23.5	579	$1\text{-}\frac{\mathrm{SSD}(\mathrm{D_O},\mathrm{D_S})}{H'_s W'_S}$	3.0	Triplet	
FFT U-Net [32]	Siamese	1.9M	23.3	478	$\frac{\mathrm{ZNCC}(\mathrm{D_O},\mathrm{D_S})}{\gamma}$	4.0	Cross-Entropy	
MARU-Net [33]	Siamese	8.9M	23.3	287	$\frac{\mathrm{ZNCC}(\mathrm{D_O},\mathrm{D_S})}{\gamma}$	3.9	Binary Cross-Entropy	
							Gauss. weighted contrastive	

Memory footprint and computational time are measured on a NVIDIA RTX 3070 Ti GPU, in FP32, using optical images of size 512×512 and SAR patches of size 128×128.

U-Net but replaces the loss function with a linear combination of a pixelwise binary CE loss with a Dirac positive mask and a Gaussian-weighted contrastive loss with a large positive mask of 13 active pixels.

Let us highlight, that each similarity measure is implemented using Fast Fourier Transforms (FFT), making the similarity scores computation time negligible compared to the backbone.

Having detailed these four state-of-the-art neural network backbones and their associated loss functions, the next section focuses on evaluating their accuracy on the MEOW dataset. Through a series of experiments, we analyze and compare their effectiveness in SAR-optical image registration under various conditions. We also demonstrate that the performance of these models can be significantly improved by changing the loss function and/or adopting a pseudo-Siamese backbone instead of a Siamese architecture.

V. EXPERIMENTS

Evaluating deep learning models for SAR-optical image registration in real-world navigation scenarios requires addressing key challenges related to generalization, robustness, and architectural design choices. Unlike conventional remote sensing applications, terrain-aided navigation systems must operate reliably despite sensor degradation, geographic variability, and temporal misalignment between reference maps and acquired images.

Our experimental framework systematically investigates these aspects by first examining spatial and temporal generalization, testing how models trained on Eastern European regions perform in Western Europe and how models trained on historical data generalize to future acquisitions. Robustness is then assessed by applying controlled speckle noise and Gaussian blur to SAR images, simulating real-world sensor quality variations to establish operational limits for accurate registration. Lastly, we conduct a comprehensive architectural sensitivity analysis by combining various backbone networks, similarity measures, and loss functions across 36 hybrid models to isolate the contribution of each component. As a final experiment, we also evaluate the models' robustness to image rotations, which are common in navigation contexts.

Using the large and diverse MEOW dataset, this experimental design advances both the theoretical understanding of cross-modal registration and practical guidance for developing robust navigation systems capable of functioning in GPS-denied environments.

A. Evaluation protocols

The MEOW dataset consists of SAR-optical image pairs acquired over the European continent for each season (winter, spring, summer, and autumn) across the years 2018, 2020, 2022, and 2023 (see Section III). As a result, the dataset includes multiple observations of the same geographical areas under varying temporal and seasonal conditions. These characteristics enable us to experimentally investigate two key questions:

1) **Spatial generalization -** If a model is trained on data from a geographical area A, how well does it perform when deployed in a different region B, geographically separated from A?
2) **Temporal generalization -** If a model is trained on data up to a certain year Y, how well does it perform on data acquired several years after Y?

To address these questions, we define two evaluation protocols based on spatial and temporal separation, as illustrated in Fig. 6 and Fig. 7.

The first protocol involves spatial separation between training and test sets. SAR-optical registration models are trained on pairs acquired in Eastern Europe and tested on pairs from Western Europe. Examples of such SAR-optical pairs are shown in Fig. 7a. Note that a given region (e.g., the area around Berlin) appears in the training set but not in the test set. To marginalize the effects of years and seasons, pairs from different seasons and years are included in both sets.

The second protocol focuses on temporal separation. Here, the models are trained on SAR–optical pairs acquired in 2018. The test data include SAR–optical pairs from 2018 to 2023, excluding pairs in which both timestamps are from 2018, as such pairs were used for training. Unlike the first protocol,

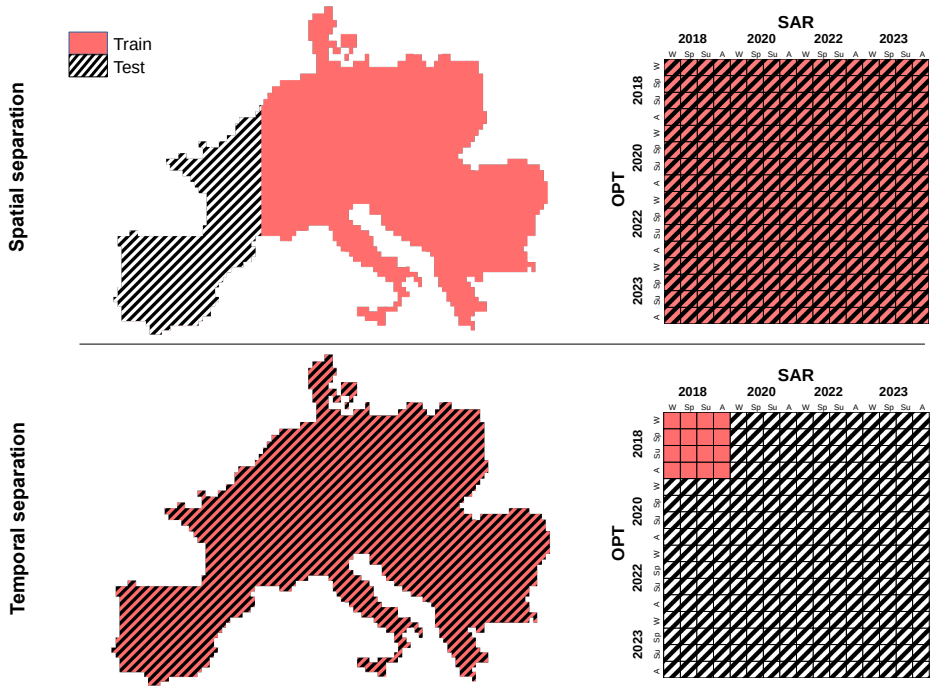


Fig. 6. Illustration of the spatial separation (top) and temporal separation (bottom) protocols. In the spatial separation protocol, the models are trained on SAR-optical pairs issued from the area on the map highlighted in red (Western Europe) and tested on the remaining regions (see top left image). In this protocol, all possible two-element temporal combinations are considered (see top right image). Conversely, the temporal separation protocol uses the entire spatial extent of the MEOW dataset for both training and testing (see bottom left image), but splits the data based on time: training data consists solely of acquisitions from 2018, while test data includes the other two-element temporal combinations that exclude the training pairs (see bottom right image).

there is no spatial separation between training and test sets. As illustrated in Fig. 7b, a given region (e.g., the area around Amsterdam) appears in both sets but at different times.

B. Implementations details

The four models are implemented in PyTorch. Each model is trained on a single Nvidia A100 GPU with early stopping. We use Adam optimizer, a batch size of 16 and a constant learning rate of 5×10^{-4}. We consider optical images of size 512×512 pixels and SAR patches of size 128×128 pixels.In the following subsections, SAR-optical image registration models are first evaluated on the spatial separation protocol.

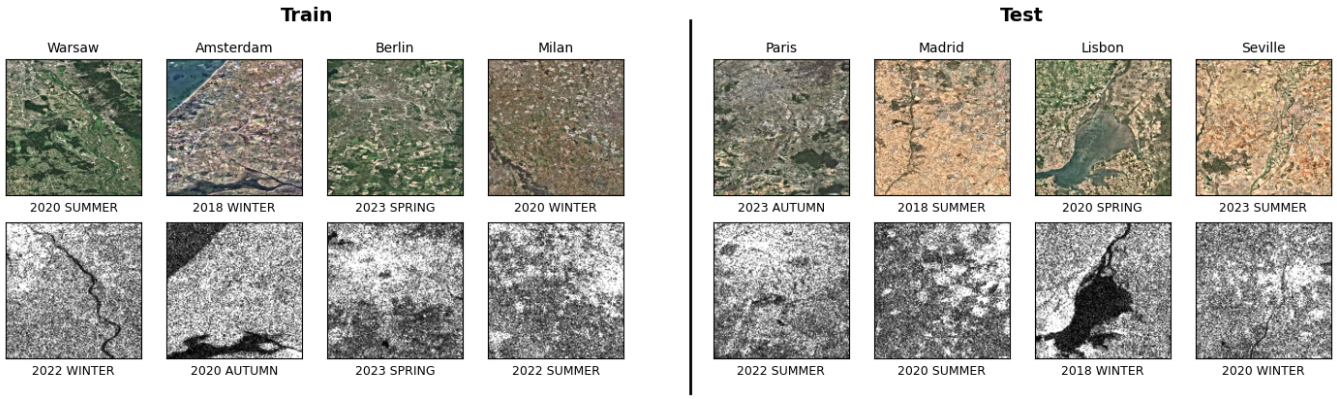
C. Spatial separation experiments

This experiment aims to answer the following question: *If a model is trained on data from Eastern Europe, how well does it perform when deployed in Western Europe?* To investigate this, we evaluate the robustness of various models under controlled degradation of the SAR patches. Two types of degradation are applied: Gaussian blur with standard deviation σ, and synthetic speckle noise characterized by the number of looks L following the fully developed speckle model [50], [51]. The Gaussian blur models spatial imprecision introduced during SAR focusing (see [52], Chapter 3), which primarily arises from uncertainties in key reconstruction parameters, such as carrier speed, look angle and acceleration. The defocusing is inherent in TAN applications, as these parameters are precisely the variables TAN systems seek to estimate during image processing. This effect is implemented as a convolution with a Gaussian kernel. The degradation model adopted in this work, combining Gaussian blur and speckle noise, provides a simplified approximation of real SAR degradation. In practice, more complex effects may arise, including geometric distortions, layover, shadowing, atmospheric influences, or sensor-specific artifacts, which are not modeled here. This approach also assumes fully developed speckle, which can be restrictive since speckle properties vary with surface type. We acknowledge these limitations and clarify that the present study is intended as a controlled validation under approximated degradations.

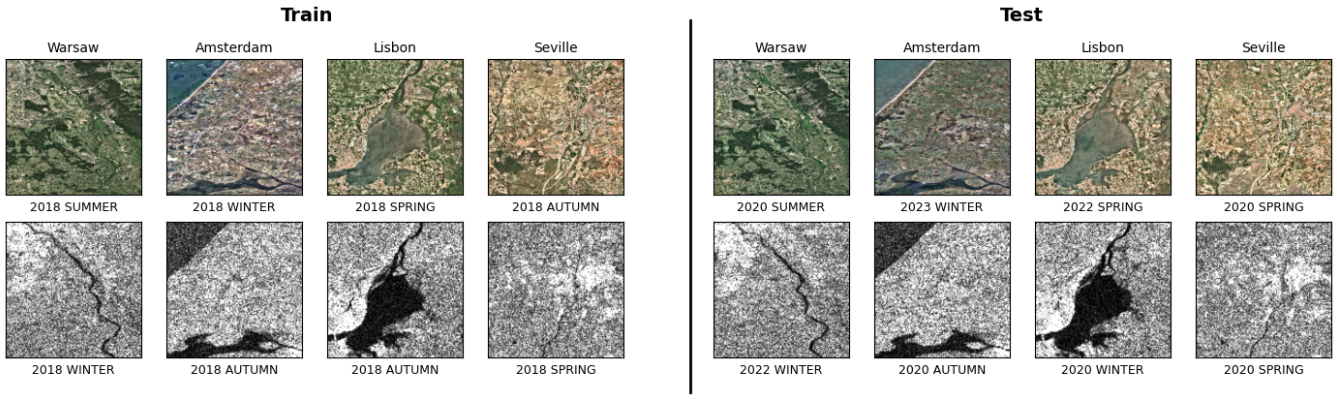
We consider the following nine degradation configurations:

- Fixed $\sigma = 0.5$ with varying $L = \{5, 8, 15, 25, 35, 75\}$.
- Fixed $L = 35$ with varying $\sigma = \{0.25, 0.5, 1, 2\}$.

1) Results on the original models: Fig. 8 reports the CMR(0) performance as a function of σ and L for four models: MCGF [30], OSMNet [31], FFT U-Net [32], and MARU-Net [33]. In low-degradation settings (top-right of the degradation cross), all models achieve nearly 100% CMR(0), indicating strong performance in registering high-quality radar-optical image pairs. However, performance degrades significantly with increasing blur or noise. Notably, MCGF, despite being the smallest model, systematically outperforms the other models. These trends are confirmed in Table IV, which shows CMR results at multiple pixel thresholds. This suggests a performance bottleneck in the larger OSMNet, FFT U-Net,



(a) Spatial separation



(b) Temporal separation

Fig. 7. Examples of optical and SAR image pairs used for the (a) spatial separation and (b) temporal separation protocols.

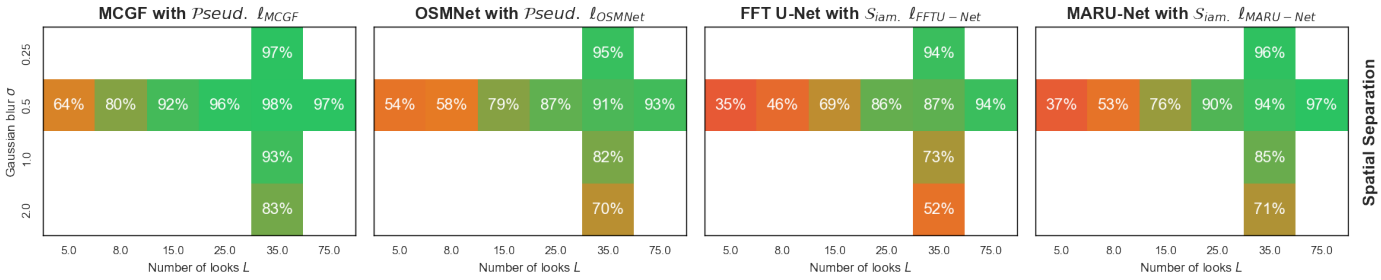


Fig. 8. Evolution of CMR(0) as a function of the standard deviation σ of the gaussian blur and the number of looks L for the four original models on the spatial separation protocol. First column: MCGF [30], second column: OSMNet [31], third column: FFT U-Net [32] and last column: MARU-Net [33].

TABLE IV
EVOLUTION OF CMR(0), CMR(1) AND CMR(2) FOR ORIGINAL MODELS ON SPATIAL SEPARATION PROTOCOL WITH $L = 8$ AND $\sigma = 0.5$. THE BEST RESULT IS IN BOLD, AND THE SECOND-BEST IS UNDERLINED.

Model name	CMR(0)	CMR(1)	CMR(2)
MCGF [30]	**0.798**	**0.852**	**0.859**
OSMNet [31]	<u>0.582</u>	<u>0.793</u>	<u>0.832</u>
FFT U-Net [32]	0.465	0.539	0.562
MARU-Net [33]	0.531	0.609	0.625

and MARU-Net architectures. In the next section, we analyze this bottleneck to improve their robustness under real-world conditions.

2) Results on hybrid models: In this section, we investigate the bottleneck identified earlier by introducing a series of hybrid models. These models are constructed by combining three key components from the original architectures (see Table III):

1) the model backbone (MCGF, OSMNet, FFT U-Net, or MARU-Net),
2) either with a pseudo-Siamese ($\mathcal{P}seud.$) architecture or a Siamese ($\mathcal{S}iam.$) architecture, and
3) the original loss function and similarity score associated with each model (ℓ_{MCGF}, ℓ_{OSMNet}, $\ell_{FFTU\text{-}Net}$, and $\ell_{MARU\text{-}Net}$).

For $\ell_{FFTU\text{-}Net}$ and $\ell_{MARU\text{-}Net}$, we additionally optimize the temperature factor γ. We also introduce a fifth loss

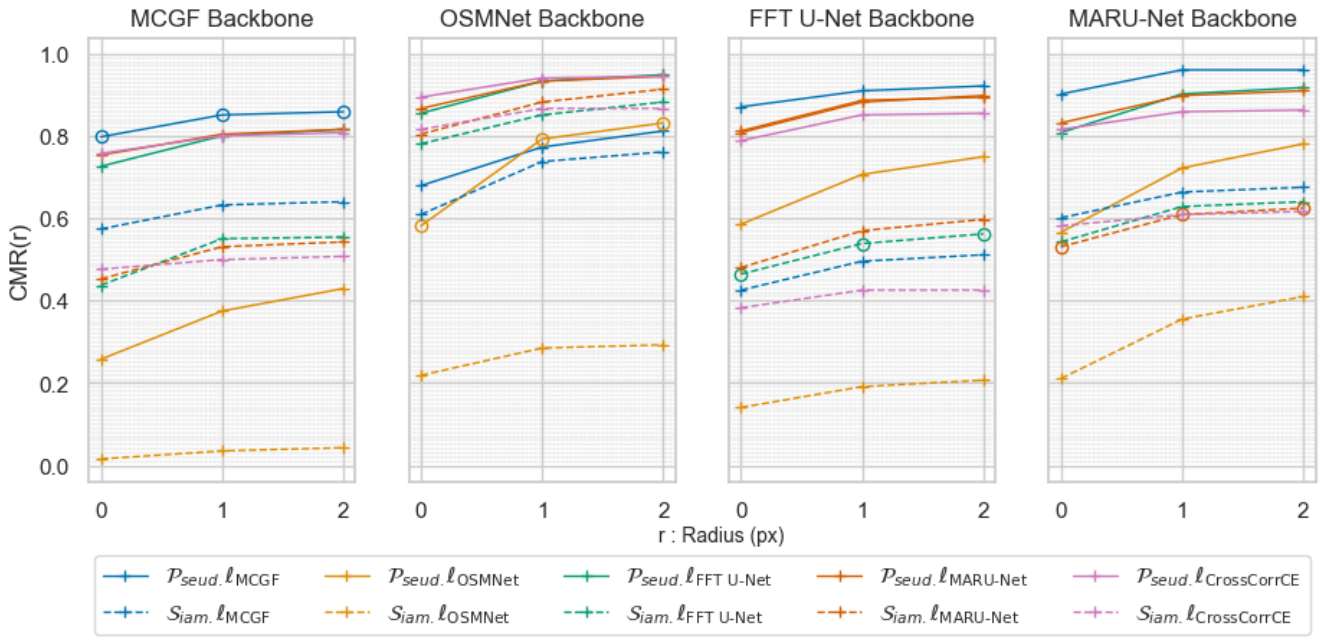


Fig. 9. Evolution of CMR(0), CMR(1) and CMR(2) for hybrid models on the spatial separation protocol with $L = 8$ and $\sigma = 0.5$. Circle markers represent the original model. Dashed lines correspond to Siamese networks, while solid lines denote pseudo-Siamese architectures.

function, denoted $\ell_{CrossCorrCE}$, which combines a standard cross-entropy loss with a cross-correlation (CC). Each hybrid model is trained from scratch, as described in Section V-B.

Quantitative results for the hybrid models under the degradation setting of $L = 8$ and $\sigma = 0.5$ - Fig. 9 presents CMR(0), CMR(1), and CMR(2) results for the hybrid models under the degradation setting of $L = 8$ and $\sigma = 0.5$. Results are grouped by backbone: MCGF, OSMNet, FFT U-Net, and MARU-Net (from left to right). Pseudo-Siamese architectures (solid lines) consistently outperform their Siamese counterparts (dashed lines), suggesting that asymmetric feature extraction for SAR and optical images, as implemented in pseudo-Siamese networks, provides enhanced robustness to SAR degradation. Interestingly, the original models (shown with circular markers) do not always yield the best performance. For all backbones except MCGF, the best-performing configurations use a loss function and similarity score different from the original one. MCGF remains the only case where the original configuration is optimal. This demonstrates that the bottleneck identified in the previous section has been fully resolved: MCGF is now substantially outperformed by hybrid configurations employing deeper backbones such as OSMNet and MARU-Net.

Table V summarizes the best-performing hybrid models. Across all backbones, pseudo-Siamese architectures consistently yield the highest accuracy. The pseudo-Siamese design leverages the multimodality of Sentinel-2 (optical) and Sentinel-1 (radar) data. For MCGF, FFT U-Net, and MARU-Net backbones, the most effective loss/similarity combination is ℓ_{MCGF}, *i.e.*, a contrastive loss with cross-correlation. For the OSMNet backbone, the best results are obtained using $\ell_{CrossCorrCE}$, *i.e.*, a standard cross-entropy loss combined with cross-correlation.

TABLE V
SUMMARY OF THE BEST HYBRID MODELS OBTAINED FOR EACH BACKBONE ON THE SPATIAL SEPARATION PROTOCOL WITH $L = 8$ AND $\sigma = 0.5$. THE BEST RESULT IS IN BOLD, AND THE SECOND-BEST IS UNDERLINED.

Backbone	Siamese?	Loss	CMR(0)	CMR(1)	CMR(2)
MCGF	$\mathcal{P}seud.$	ℓ_{MCGF}	0.798	0.852	0.859
OSMNet	$\mathcal{P}seud.$	$\ell_{CrossCorrCE}$	<u>0.895</u>	<u>0.941</u>	<u>0.945</u>
FFT U-Net	$\mathcal{P}seud.$	ℓ_{MCGF}	0.871	0.910	0.922
MARU-Net	$\mathcal{P}seud.$	ℓ_{MCGF}	**0.902**	**0.961**	**0.961**

Notably, ℓ_{OSMNet} significantly degrades performance for all backbones. A theoretical analysis of this loss function, highlighting potential issues, is provided in Appendix A.

Quantitative results for the best hybrid models across all degradation configurations - We addressed the bottleneck identified in Section V-C1 by searching for optimal hybrid models under a specific degradation condition ($L = 8$, $\sigma = 0.5$). We now retrain these optimal hybrid models across all degradation configurations. Fig. 10 presents the evolution of CMR(0) and its relative error, comparing hybrid models to their original counterparts. As shown, the proposed hybrid models significantly enhance registration accuracy, particularly under the most challenging degradation conditions. Notably, the pseudo-Siamese OSMNet model trained with the $\ell_{\text{CrossCorrCE}}$ loss consistently ranks among the top-performing models across all noise levels, demonstrating a strong spatial generalization ability. Therefore, this model will be used exclusively in the temporal separation experiment (see Section V-D).

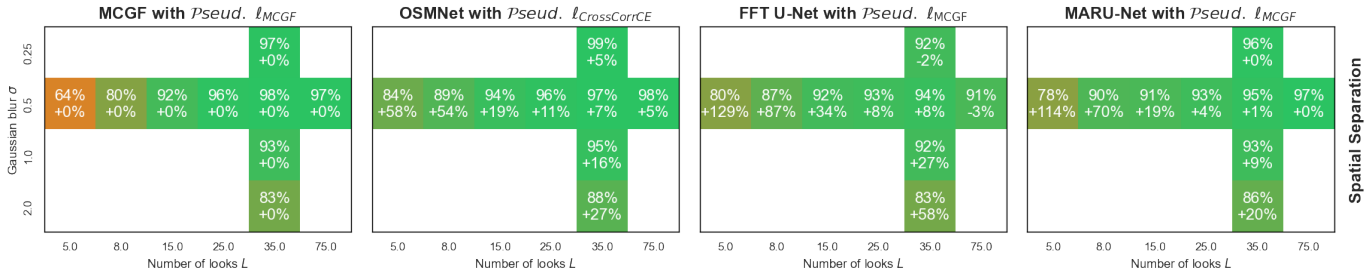


Fig. 10. CMR(0) of progressive degradation applied to the best-performing models from the ablation study (case $L = 8$, $\sigma = 0.5$): generalized performance and relative error compared to their original versions. From left to right: Original MCGF, OSMNet with $\mathcal{P}seud.\ \ell_{CrossCorrCE}$, FFT U-Net with $\mathcal{P}seud.\ \ell_{MCGF}$ and MARU-Net with $\mathcal{P}seud.\ \ell_{MCGF}$.

3) Spatial analysis: To further evaluate the spatial generalization ability of the best-performing hybrid models for each backbone, we conduct a spatial analysis. Fig. 11 presents maps of the CMR(0) metric under degraded conditions ($L = 8$, $\sigma = 0.5$). The first column shows the CMR(0) results for the original models, while the second column displays the results obtained with the best hybrid models. As shown, the best hybrid models significantly improve registration accuracy compared to their original counterparts.

Notably, the pseudo-Siamese OSMNet model trained with the $\ell_{\text{CrossCorrCE}}$ loss exhibit strong spatial generalization, as there is no apparent distinction between training and testing regions in terms of CMR(0). Its registration accuracy is slightly reduced in:

- mountainous regions such as the Alps (part of the training set), likely due to SAR-specific artifacts such as layover and shadowing, as well as the presence of snow in some optical images,
- in arid regions, such as the south of Italy (part of the training set) and Spain (part of the testing set).

These findings clearly confirm the added value of using hybrid models, and their strong spatial generalization capability.

D. Temporal separation experiments

This experiment addresses the following question: *If a model is trained on data from 2018, how well does it generalize to data acquired several years later?* To investigate these aspects, we evaluate the best-performing hybrid model identified in the previous section: the pseudo-Siamese OSMNet model trained with the $\ell_{CrossCorrCE}$ loss.

The model is tested under degraded conditions ($L = 8$, $\sigma = 0.5$), using the temporal separation protocol illustrated in Fig. 7b. Training is performed from scratch on SAR-optical image pairs from 2018, as described in Sec. V-B.

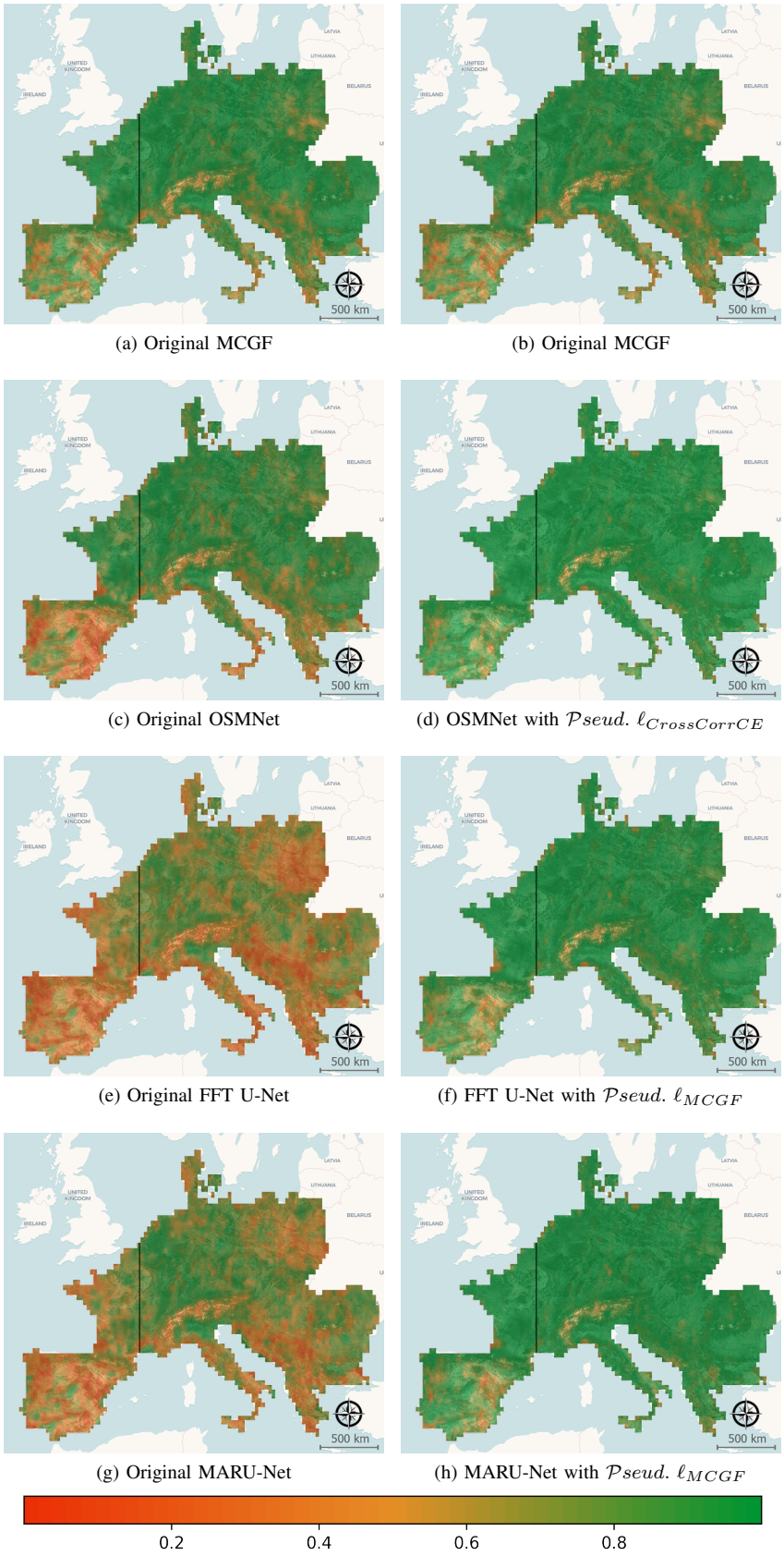
We assess generalization by plotting CMR(0) for each season and year of the optical images against each season and year of the SAR patches. The results are presented in Fig. 12, and our key findings are as follows:

- All performances remain high (>91%).
- For optical images from 2018, performance does not degrade when tested with SAR patches from later years, indicating strong temporal generalization.
- All plots exhibit a pseudo-periodic pattern with a period of one year, reflecting seasonal variations in SAR data.
- Optical images acquired in summer yield better performance than those from other seasons. In particular, winter 2018 optical images result in lower performance due to cloud and snow coverage.

Additionally, we will investigate whether temporal generalization performance remains uniform across the entire geographic extent of our dataset. Based on the preceding analysis, we determined that the optimal registration performance is achieved by using a reference optical image acquired in summer. Moreover, no clear trend indicates that the optical reference image must be from the same year as the SAR image. Therefore, we fixed the optical reference image to summer 2018. Fig. 13 illustrates the local variation in registration performance between 2018 and 2023. Overall, performance remains stable with localized variations, particularly in mountainous areas that also proved challenging in the spatial separation protocol. The area highlighted by the blue map marker in Fig. 13 shows decreased performance in 2023 compared to 2018, corresponding to the region shown in Fig. 3, where extensive wildfires altered the landscape between 2018 and 2023.

VI. Perspective: Beyond 2D Translations

A last experiment is conducted to assess the impact of SAR rotation on model performance under the degraded setting characterized by $L = 8$ and $\sigma = 0.5$. Fig. 14 shows the impact of SAR image rotation on the CMR performance of the pseudo-Siamese OSMNet model with $\ell_{CrossCorrCE}$ loss. The x-axis indicates increasing rotation ranges applied to the SAR images, while the y-axis shows the corresponding CMR values. It clearly shows that the model is highly sensitive to SAR rotation. This insight is critical for real-world applications. Registration performance may degrade severely in scenarios where SAR and optical images are not well aligned in orientation. Future work will aim to develop methods that enhance this rotational tolerance alongside the extraction of scale-invariant features, thereby enabling robust registration of rigid transformations between SAR and optical images.



(a) Original MCGF (b) Original MCGF

(c) Original OSMNet (d) OSMNet with $\mathcal{P}seud.\ \ell_{CrossCorrCE}$

(e) Original FFT U-Net (f) FFT U-Net with $\mathcal{P}seud.\ \ell_{MCGF}$

(g) Original MARU-Net (h) MARU-Net with $\mathcal{P}seud.\ \ell_{MCGF}$

Fig. 11. Spatial analysis of CMR(0) metric for $L = 8$ and $\sigma = 0.5$. First column: original models and second column: best hybrid models obtained after the ablation study for each backbone.The vertical black line marks the boundary between the training set (right) and the test set (left).

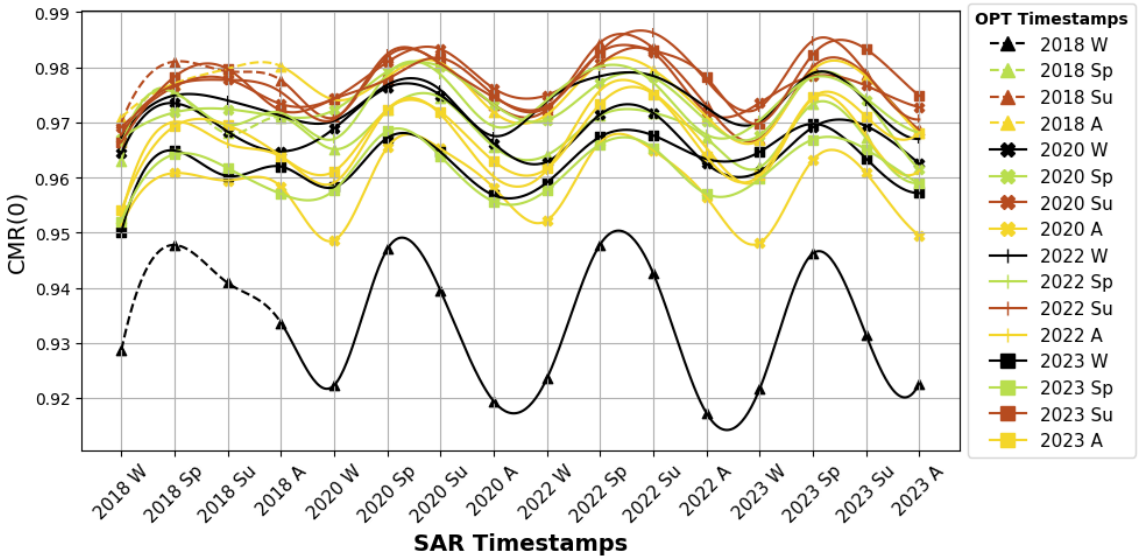


Fig. 12. CMR(0) of the best performing OSMNet with CrossCorrCE over temporal separation strategy in the case $L = 8$ and $\sigma = 0.5$ depending on the timestamps of the SAR-optical pairing. Dashed lines correspond to the training set.

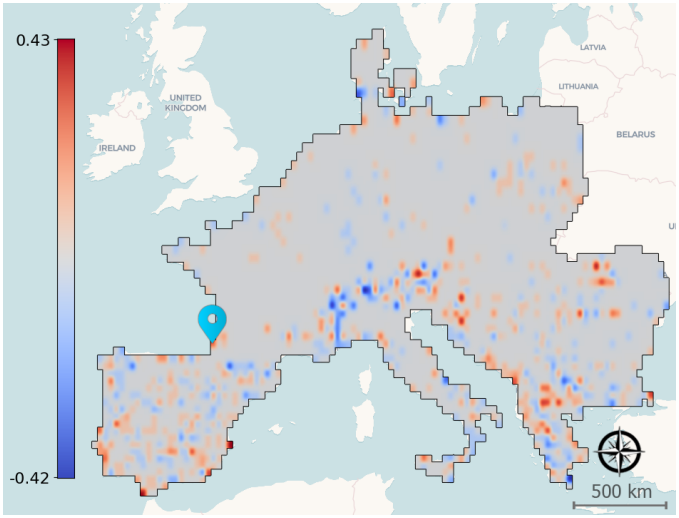


Fig. 13. Local trends in temporal generalization over the entire dataset. We illustrate the difference in registration performance (CMR(0) metric) for SAR images acquired in 2018 and 2023, respectively. The optical images are from summer 2018.

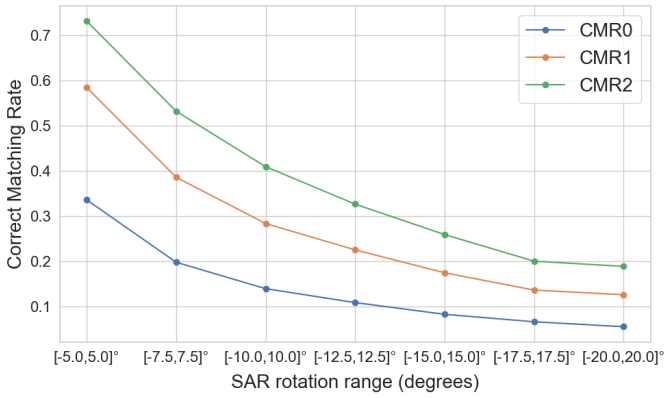


Fig. 14. Impact of SAR rotation on registration accuracy using Pseudo OSMNet with CrossCorrCE loss.

VII. Conclusion

This study addresses the critical challenge of multi-modal SAR-optical image registration for autonomous navigation in GPS-degraded environments. Our work reveals significant insights into the limitations of current architectures and proposes robust solutions tailored to real-world operational conditions.

The main contributions include the creation of the MEOW dataset, an extensive resource covering 4.9 million km^2 with multi-temporal and multi-seasonal acquisitions. Our systematic evaluation of four state-of-the-art architectures under controlled degradation demonstrates the superiority of pseudo-siamese networks over traditional siamese architectures, especially under severe SAR image degradation. An in-depth ablation study reveals that the optimal combination of architectural components can improve performance by over 120% in the most degraded scenarios. Spatial and temporal separation experiments confirm the robustness of the proposed hybrid models across diverse European geographies and temporal conditions, establishing their potential for real-world terrain navigation applications. These results set new benchmarks for deep learning-based cross-modal alignment and provide a solid foundation for the development of robust autonomous navigation systems in constrained operational environments.

Future work will focus on extending the registration framework to handle non-translational transformations, including scale and rotation changes. We also aim to address the limitations of the current study, such as the assumption of fully developed speckle and the simplified degradation models, to improve robustness in more realistic SAR-optical scenarios.

Acknowledgment

The authors would like to thank Yannick Chevalier from the CEA for his useful advice on SAR image processing.

APPENDIX A
ANALYSIS OF THE OSMNET LOSS FUNCTION

The OSMNet loss function [31] is defined as

$$\ell_{OSMNet} = \log\left[1 + \sum_{z_n \in N_s} \sum_{z_p \in P_s} \exp\left(\frac{\xi_n(z_n) + \xi_p(z_p)}{\gamma}\right)\right], \tag{3}$$

where N_s and P_s denote the sets of negative and positive samples, respectively. The masks indicating positive (red) and negative (white) samples are shown in Table III. The parameter γ is a temperature scaling factor. The terms ξ_n and ξ_p are given by

$$\xi_n(z_n) = \omega_n(z_n - \Delta_n), \tag{4}$$

$$\xi_p(z_p) = -\omega_p(z_p - \Delta_p), \tag{5}$$

with $\Delta_p = 1 - m$ and $\Delta_n = m$ where m is the margin. The self-adaptive weights are defined as $\omega_n = [z_n - Y_n]_+$ and $\omega_p = [Y_p - z_p]_+$, where $[\cdot]_+$ denotes the ReLU function and the targets are set to $Y_n = -m$ and $Y_p = 1 + m$ [31]. Substituting these yields

$$\xi_n(z_n) = [z_n + m]_+ (z_n - m), \tag{6}$$

$$\xi_p(z_p) = -[1 + m - z_p]_+ (z_p - 1 + m). \tag{7}$$

As summarized in Table III, OSMNet uses a similarity measure based on the normalized sum of squared differences (SSD):

$$Z = 1 - \hat{\text{SSD}} = 1 - \frac{\text{SSD}}{H'_{\text{S}} W'_{\text{S}}}, \tag{8}$$

In theory, the loss function should encourage minimizing SSD for positive samples and maximizing it for negative samples. Accordingly, the similarity scores are defined as $z_p = 1 - \hat{\text{SSD}}_p$ and $z_n = 1 - \hat{\text{SSD}}_n$. Substituting gives

$$\xi_n(\hat{\text{SSD}}_n) = \left[1 - \hat{\text{SSD}}_n + m\right]_+ (1 - \hat{\text{SSD}}_n - m), \tag{9}$$

$$\xi_p(\hat{\text{SSD}}_p) = -\left[m + \hat{\text{SSD}}_p\right]_+ (m - \hat{\text{SSD}}_p). \tag{10}$$

Since $m > 0$ and $\hat{\text{SSD}}_p \geq 0$, (10) simplifies to

$$\xi_p(\hat{\text{SSD}}_p) = -m^2 + \hat{\text{SSD}}_p^2. \tag{11}$$

Thus, minimizing the OSMNet loss in (3) drives $\hat{\text{SSD}}_p$ toward zero. In the ideal case, where the SAR feature and its corresponding positive optical features are identical, then $\hat{\text{SSD}}_p = 0$.

However, (9) can be rewritten as a truncated quadratic loss:

$$\xi_n(\hat{\text{SSD}}_n) = \begin{cases} (1 - \hat{\text{SSD}}_n)^2 - m^2 & \text{if } \hat{\text{SSD}}_n < 1 + m \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Therefore, the minimum is achieved at $\hat{\text{SSD}}_n = 1$. This result is surprising because the margin m does not act as an explicit separation between the SAR feature and negative optical features; rather, it sets the truncation threshold of the quadratic loss. As a consequence, minimizing the OSMNet loss does not encourage the SAR feature to be pushed far from the negatives: it is sufficient for $\hat{\text{SSD}}_n$ to reach 1. Among the losses compared in Table III, only OSMNet lacks this repulsive effect, which likely contributes to its weaker registration performance, as demonstrated by the results in Fig. 9.

REFERENCES

[1] G. Zhang and L.-T. Hsu, "Intelligent GNSS/INS integrated navigation system for a commercial UAV flight control system," *Aerospace Science and Technology*, vol. 80, pp. 368–380, 2018.

[2] C. Chi, X. Zhan, S. Wang, and Y. Zhai, "Enabling robust and accurate navigation for UAVs using real-time GNSS precise point positioning and IMU integration," *The Aeronautical Journal*, pp. 1–22, 10 2020.

[3] D. A. Grejner-Brzezinska, C. K. Toth, H. Sun, X. Wang, and C. Rizos, "A robust solution to high-accuracy geolocation: Quadruple integration of GPS, IMU, pseudolite, and terrestrial laser scanning," *IEEE TIM*, vol. 60, no. 11, pp. 3694–3708, 2011.

[4] N. Gyagenda, J. V. Hatilima, H. Roth, and V. Zhmud, "A review of GNSS-independent UAV navigation techniques," *Robotics and Autonomous Systems*, vol. 152, p. 104069, 2022.

[5] C. V. Angelino, V. Baraniello, and L. Cicala, "High altitude UAV navigation using IMU, GPS and camera," *FUSION 2013*, pp. 647–654, 1 2013.

[6] F. Yao, C. Lan, L. Wang, H. Wan, T. Gao, and Z. Wei, "GNSS-denied geolocalization of UAVs using terrain-weighted constraint optimization," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 135, p. 104277, 2024.

[7] Z. Sjanic and F. Gustafsson, "Navigation and SAR focusing with map aiding," *IEEE TAES*, vol. 51, no. 3, pp. 1652–1663, 2015.

[8] J. Öfverstedt, J. Lindblad, and N. Sladoje, "Fast computation of mutual information in the frequency domain with applications to global multimodal image alignment," *Pattern Recognit. Lett.*, vol. 159, pp. 196–203, 2022.

[9] S. Suri and P. Reinartz, "Mutual-information-based registration of TerraSAR-X and Ikonos imagery in urban areas," *IEEE TGRS*, vol. 48, no. 2, pp. 939–949, 2010.

[10] H. Zhang, L. Lei, W. Ni, T. Tang, J. Wu, D. Xiang, and G. Kuang, "Optical and SAR image matching using pixelwise deep dense features," *IEEE GRSL*, vol. 19, pp. 1–5, 2022.

[11] Y. Xiang, R. Tao, F. Wang, H. You, and B. Han, "Automatic registration of optical and SAR images via improved phase congruency model," *IEEE J-STARS*, vol. 13, pp. 5847–5861, 2020.

[12] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE TGRS*, vol. 55, no. 5, pp. 2941–2958, 2017.

[13] F. Dellinger, J. Delon, Y. Gousseau, J. Michel, and F. Tupin, "SAR-SIFT: A SIFT-like algorithm for SAR images," *IEEE TGRS*, vol. 53, no. 1, pp. 453–466, 2015.

[14] Y. Xiang, R. Tao, L. Wan, and F. Wang, "OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE TGRS*, vol. 56, no. 6, pp. 3078–3090, 2018.

[15] B. Fan, C. Huo, C. Pan, and Q. Kong, "Registration of optical and SAR satellite images by exploring the spatial relationship of the improved SIFT," *IEEE GRSL*, vol. 10, no. 4, pp. 657–661, 2013.

[16] S. Paul and U. C. Pati, "SAR image registration using an improved SAR-SIFT algorithm and Delaunay-triangulation-based local matching," *IEEE J-STARS*, vol. 12, no. 8, pp. 2958–2966, 2017.

[17] N. Merkle, W. Luo, S. Auer, R. Müller, and R. Urtasun, "Exploiting deep matching and SAR data for the geo-localization accuracy improvement of optical satellite images," *Remote Sensing*, vol. 9, no. 6, p. 586, 2017.

[18] X. Jiang, J. Ma, G. Xiao, Z. Shao, and X. Guo, "A review of multimodal image matching: Methods and applications," *Information Fusion*, vol. 73, pp. 22–71, 2021.

[19] X. Yang, Z. Wang, J. Zhao, and D. Yang, "Fg-gan: A fine-grained generative adversarial network for unsupervised sar-to-optical image translation," *IEEE TGRS*, vol. 60, pp. 1–11, 2022.

[20] N. Merkle, S. Auer, R. Müller, and P. Reinartz, "Exploring the potential of conditional adversarial networks for optical and sar image matching," *IEEE J-STARS*, vol. 11, no. 6, pp. 1811–1820, 2018.

[21] F. Luo, L. Hong, L. Duan, Y. Wu, and H. Gong, "Robust registration of SAR and optical images based on deep learning and improved harris algorithm," *Scientific Reports*, vol. 12, p. 6794, 2022.

[22] Z. Guo, J. Liu, Q. Cai, Z. Zhang, and S. Mei, "Learning sar-to-optical image translation via diffusion models with color memory," *IEEE J-STARS*, vol. 17, pp. 14 454–14 470, 2024.

[23] P. Ren, Z. Han, Z. Yu, and B. Zhang, "Confucius tri-learning: A paradigm of learning from both good examples and bad examples," *Pattern Recognition*, vol. 163, p. 111481, 2025.

[24] X. Tao, B. Koirala, A. Plaza, and P. Scheunders, "A new dual-feature fusion network for enhanced hyperspectral unmixing," *IEEE TGRS*, vol. 62, pp. 1–13, 2024.

[25] Y. Xu, W. Hao, F. Zhou, C. Luo, X. Sun, S. Rahardja, and P. Ren, "MambaHSISR: Mamba hyperspectral image super-resolution," *IEEE TGRS*, vol. 63, pp. 1–16, 2025.

[26] Y. Fu, Z. Liu, and J. Lyu, "Reason and discovery: A new paradigm for open set recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 7, pp. 5586–5599, 2025.

[27] S. Leprince, S. Barbot, F. Ayoub, and J. P. Avouac, "Automatic and precise orthorectification, coregistration, and subpixel correlation of satellite images, application to ground deformation measurements," *IEEE TGRS*, vol. 45, no. 6, pp. 1529–1558, 2007.

[28] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 148–164, 2018.

[29] H. Zhang, W. Ni, W. Yan, D. Xiang, J. Wu, X. Yang, and H. Bian, "Registration of multimodal remote sensing image based on deep fully convolutional neural network," *IEEE J-STARS*, vol. 12, no. 8, pp. 3028–3042, 2019.

[30] L. Zhou, Y. Ye, T. Tang, K. Nan, and Y. Qin, "Robust matching for SAR and optical images using multiscale convolutional gradient features," *IEEE GRSL*, vol. 19, pp. 1–5, 2022.

[31] H. Zhang, L. Lei, W. Ni, T. Tang, J. Wu, D. Xiang, and G. Kuang, "Explore better network framework for high resolution optical and SAR image matching," *IEEE TGRS*, vol. 60, pp. 1–18, 2022.

[32] Y. Fang, J. Hu, C. Du, Z. Liu, and L. Zhang, "SAR-optical image matching by integrating siamese U-Net with FFT correlation," *IEEE GRSL*, vol. 19, pp. 1–5, 2022.

[33] M. Gazzea, O. Sommervold, and R. Arghandeh, "MARU-Net: Multiscale attention gated residual U-Net with contrastive loss for SAR-optical image matching," *IEEE J-STARS*, vol. 16, pp. 4891–4899, 2023.

[34] M. Schmitt, L. H. Hughes, C. Qiu, and X. X. Zhu, "SEN12MS – a curated dataset of georeferenced multi-spectral Sentinel-1/2 imagery for deep learning and data fusion," *ISPRS Ann. Photogramm. Remote Sens.*, vol. IV-2/W7, pp. 153–160, 2019.

[35] K. N. Clasen, L. Hackel, T. Burgert, G. Sumbul, B. Demir, and V. Markl, "reBEN: Refined BigEarthNet dataset for remote sensing image analysis," 2024. [Online]. Available: https://arxiv.org/abs/2407.03653

[36] M. Huang, Y. Xu, L. Qian, W. Shi, Y. Zhang, W. Bao, N. Wang, X. Liu, and X. Xiang, "The QXS-SAROPT dataset for deep learning in SAR-optical data fusion," 2021.

[37] X. Li, G. Zhang, H. Cui, S. Hou, S. Wang, X. Li, Y. Chen, Z. Li, and L. Zhang, "MCANet: A joint semantic segmentation framework of optical and SAR images for land use classification," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 106, p. 102638, 2022.

[38] M. Kampffmeyer, D. Eriksson, and X. X. Zhu, "SARptical: A dataset for SAR-optical image matching in dense urban areas," in *IGARSS 2019*. IEEE, 2019, pp. 4189–4192.

[39] M. Kampffmeyer, A. N. Myronenko, and X. X. Zhu, "So2Sat LCZ42: A benchmark dataset for global local climate zones classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 178, pp. 191–203, 2021.

[40] Y. Wang, N. A. A. Braham, Z. Xiong, C. Liu, C. M. Albrecht, and X. X. Zhu, "SSL4EO-S12: A large-scale multimodal, multitemporal dataset for self-supervised learning in earth observation [software and data sets]," *IEEE Geosci. Remote Sens. Mag.*, vol. 11, no. 3, pp. 98–106, 2023.

[41] M. Schmitt, L. H. Hughes, and X. X. Zhu, "The SEN1-2 dataset for deep learning in SAR-optical data fusion," *ISPRS Ann. Photogramm. Remote Sens.*, vol. IV-1, pp. 141–146, 2018.

[42] W. Zhang, R. Zhao, Y. Yao, Y. Wan, P. Wu, J. Li, Y. Li, and Y. Zhang, "Multi-resolution SAR and optical remote sensing image registration methods: A review, datasets, and future perspectives," 2025. [Online]. Available: https://arxiv.org/abs/2502.01002

[43] R. Wenger, A. Puissant, J. Weber, L. Idoumghar, and G. Forestier, "MultiSenGE: A multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications," *ISPRS Ann. Photogramm. Remote Sens.*, vol. V-3-2022, pp. 635–640, 2022.

[44] C. González, M. Bachmann, J.-L. Bueso-Bello, P. Rizzoli, and M. Zink, "A fully automatic algorithm for editing the TanDEM-X global DEM," *Remote Sensing*, vol. 12, no. 23, 2020.

[45] J.-L. Bueso-Bello, M. Martone, C. González, F. Sica, P. Valdo, P. Posovszky, A. Pulella, and P. Rizzoli, "The global water body layer from TanDEM-X interferometric SAR data," *Remote Sensing*, vol. 13, no. 24, 2021.

[46] M. Martone, P. Rizzoli, C. Wecklich, C. González, J.-L. Bueso-Bello, P. Valdo, D. Schulze, M. Zink, G. Krieger, and A. Moreira, "The global forest/non-forest map from TanDEM-X interferometric SAR data," *Remote Sensing of Environment*, vol. 205, pp. 352–373, 2018.

[47] D. Zanaga, "ESA WorldCover 10 m 2021 v200," https://doi.org/10.5281/zenodo.7254221, Oct. 2022, zenodo.

[48] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[49] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, 2019.

[50] J. W. Goodman, "Some fundamental properties of speckle," *Journal of the Optical Society of America*, vol. 66, no. 11, pp. 1145–1150, 1976.

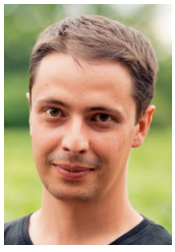
[51] E. Dalsasso, L. Denis, and F. Tupin, "SAR2SAR: A semi-supervised despeckling algorithm for SAR images," *IEEE J-STARS*, vol. 14, p. 4321–4329, 2021.

[52] C. Oliver and S. Quegan, *Understanding synthetic aperture radar images*, ser. SciTech Radar and Defense Series. SciTech Publishing, 2004, originally published by Artech House in 1998.

Simon Bertrand holds an engineering degree in physics and telecommunications and a M.S. degree in image and data analysis from the University of Strasbourg. His expertise primarily lies in software development and applied mathematics. He's currently pursuing a Ph.D. in machine learning at the University of Bordeaux.

Guillaume Bourmaud is an Associate Professor in computer vision and machine learning at the ENSEIRB-MATMECA School of Engineering. He conducts his research within the Signal and Image Group of the IMS Laboratory at the University of Bordeaux. His work mainly focuses on 3D computer vision, where he has worked on optimization methods, image matching, and neural network architectures.

Cornelia Vacar holds a Ph.D. in Bayesian approaches for image processing. With applied experience across a wide range of applications, she has successfully addressed image processing problems in biology, medicine, geology and non-destructive testing. Currently at the CEA, she focuses on terrain-aided navigation, applying her probabilistic modeling and image processing expertise to enhance autonomous navigation through terrain feature extraction and localization algorithms.



Lionel Bombrun holds an engineering degree from ENSIEG and a Ph.D in signal and image processing from the National Polytechnic Institute of Grenoble. He currently serves as an Associate Professor in statistics at Bordeaux Sciences Agro. He conducts its research activities within the Signal and Image Group of the IMS Laboratory, focusing on the MOTIVE theme (Models, Textures, Images, Volumes). His research primarily addresses computer vision with applications in remote sensing for agri-environmental domains.