GeoFlowNet: Fast and Accurate Sub-pixel Displacement Estimation From Optical Satellite Images based on Deep Learning

Tristan Montagnon, James Hollingsworth, Erwan Pathier, Mathilde Marchandon, Mauro Dalla Mura, Sophie Giffard-Roisin

GeoFlowNet: Fast and Accurate Sub-pixel Displacement Estimation From Optical Satellite Images based on Deep Learning

Tristan Montagnon, James Hollingsworth, Erwan Pathier, Mathilde Marchandon, Mauro Dalla Mura, Sophie Giffard-Roisin

***Abstract*—Optical satellite imagery is widely used for estimating ground movement in the aftermath of natural disasters such as earthquakes. This type of imagery enables detailed analysis of the factors and mechanisms that drive or influence these events. By using sub-pixel correlation algorithms, it provides precise displacement measurements (in the meter-to-centimeter range) and high spatial resolution (decimeter-to-centimeter level) by comparing images taken before and after the event.**

In this study, we present a deep neural network approach, trained on our new specific realistic dataset FaultDeform, to retrieve full-scale seismic ground motion displacement fields from optical satellite images with sub-pixel precision. The FaultDeform dataset, available at https://doi.org/10.57745/G02ZXZ, is the first satellite synthetic dataset tailored for ground motion estimation. We introduce the GeoFlowNet pipeline, utilizing a U-net architecture to solve the displacement estimation problem, delivering high-speed performance through GPU implementation, and outperforming current correlators in speed and precision. Comprehensive comparisons with state-of-the-art methods such as COSI-Corr, MicMac and CNN-DIS, and validation on real-world data from the 2019 Ridgecrest and 2013 Balochistan earthquakes showcases the robustness of our method. Codes are freely available: gricad-gitlab.univ-grenoble-alpes.fr/montagtr/GeoFlowNet.

***Index Terms*—Remote sensing, Deep Learning, Optical Image Correlation, Geodesy, Optical Flow, Dataset, Fault Deformation**

I. Introduction

Precise and accurate measurement of ground displacement at regional scales is fundamental for the study of natural hazards, such as earthquakes [1]–[3], volcanoes [4]–[7], [7], landslides [8]–[10], as well as monitoring of glaciers [11]–[14]. For earthquakes, precise and unbiased ground deformation measurement is essential for resolving the location, geometry, spatial distribution (e.g. on- vs off-fault), and down-dip slip distribution of seismogenic faults. Accurate characterization of the near-field displacement around surface ruptures, in turn, provides valuable constraints needed to understand the physics of earthquake slip [15], and ultimately to anticipate the seismic hazard posed to neighboring populations and infrastructure [16].

Optical image correlation (OIC) is an imaging technique used to estimate displacements between two optical satellite images acquired over the same area at different times (separated by hours up to years). It has been widely applied to the study of surface rupturing earthquakes, where displacements are generally small relative to the pixel size, and sub-pixel precision is required to accurately resolve the displacement field [1]–[3], [17]. Moreover, close to earthquake surface ruptures, the deformation may become complex (e.g. sharp discontinuities, distributed off-fault deformation, secondary faulting, along-strike slip variability), and InSAR often decorrelates close to the surface rupture.

Optical flow is a broader domain of computer vision that aims to estimate the flow, i.e. the full displacement field, between images of the same scene acquired at different times, and where objects have moved or deformed. A few studies using the optical flow problem formulation to retrieve ground or glacier displacement maps have been proposed [18], [19], highlighting their ability to monitor slow-to-rapid horizontal motions. However, a majority of studies have relied on Sentinel-2 (10 m) data, with only limited exploration of very high-resolution sensors (e.g., WorldView, Pléiades), with nominal precision decreasing exponentially with the temporal baseline between two acquisitions (reaching $>$2 pixels uncertainties [18]).

Recent preliminary studies employing deep learning techniques managed to retrieve clear coseismic displacement [20] or to detect slow earthquakes [21], [22], both in noisy Interferometric Synthetic Aperture Radar (InSAR) timeseries, but all relying on multiple acquisitions to remove noise components specific to InSAR. InSAR uses radar signals emitted and received by satellites or aircraft, measuring changes in the phase of these signals between acquisitions. This technique is effective in detecting both vertical and horizontal ground displacements at millimeter-scale precision over large areas. However, InSAR data is sensitive to noise from atmospheric variations, signal decorrelation, and topographic effects, which often necessitates multiple acquisitions to filter out noise and isolate significant displacement data. Additionally, InSAR often decorrelates near earthquake faults when displacement is large. In contrast, optical imagery relies on visible light to

Thanks to IDEX Université Grenoble Alpes, INSU PNTS, CNES, and CDP Risk for funding. For the computations, thanks to GRICAD infrastructure (gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, and to the HPC/AI resources of IDRIS under the allocation 2022-AD011012345R2 made by GENCI. ISTerre is part of Labex OSUG@2020 (ANR10 LABX56). This work has been partially supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003 and ANR-23-IACL-0006).

T. Montagnon, J. Hollingsworth, E. Pathier and S. Giffard-Roisin are with Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France.

M. Marchandon is with Department of Earth and Environmental Sciences, Ludwig-Maximillians-Universität München, Germany.

M. Dalla Mura is with Univ. Grenoble Alpes, CNRS, Grenoble INP (Institute of Engineering Univ. Grenoble Alpes), GIPSA-lab, 38000 Grenoble, France and Institut Universitaire de France (IUF), 75231 Paris.

capture high-resolution images for analysis, tracking changes in surface features over time. This method is well-suited for mapping distinct surface displacements, such as fault movements, and is generally less affected by atmospheric noise than InSAR. However, optical methods are constrained in their ability to detect subtle shifts and can be affected by lighting, shadows, occlusions. Montagnon et al. [23] presented a first deep learning method using a local approach for resolving the displacement field with a convolutional neural networks using optical imagery. This method assumes a uniform shift within localized segments (uniform-by-part, assuming uniform displacement within localized segments, instead of across the entire image as in traditional methods), trying to mitigate the bias near faults. However, the assumption remains simplistic with respect to the noise sources, the range of displacements it can accurately capture (constrained to +/-5 pixels), and the local sliding window method (i.e. computing one displacement vector at a time) remains computationally expensive.

Satellite images often suffer from noise, data corruption, and temporal decorrelation, causing bias and artifacts in displacement maps. Illumination changes between acquisitions (often separated by weeks or months), vegetation, anthropogenic factors, and sensor artifacts further degrade results [1], [10], [24], [25]. Pre-processing by data providers can introduce periodic artifacts and aliasing from resampling [24], [26]. Finally, processing imagery over large regions in remote sensing can be computationally intensive [27]. While large optical images cover vast areas, they involve significant computation time and often require tiling to manage memory. Different sensors, like high-resolution WorldView-3 or Pléiades (0.3-0.5 m resolution) and mid-resolution Sentinel-2 or Landsat 8 (10-15 m resolution) [28], [29], add complexity, each sensor capturing imagery with different spatial resolutions and noise characteristics, resulting in varying levels of detail and potential displacement estimates (with displacements varying from sub-pixel to multiple pixels). This variability demands extensive parameter tuning [1], [2], requiring expert intervention, complicating reproducibility and generalizability, and preventing a one-size-fits-all solution.

In this paper, we introduce a method based on a deep convolutional neural network, trained on our newly generated synthetic displacement dataset **FaultDeform**, for retrieving earthquake-induced ground displacements with sub-pixel precision and high accuracy. Our **GeoFlowNet** pipeline, illustrated in Figure 1, utilizes a U-net architecture to solve the full-scale displacement estimation problem, offering significant speed improvements through GPU implementation. This approach overcomes the typical limitation of uniform local shifts in state-of-the-art methods by estimating comprehensive displacement fields and leveraging a multi-scale architecture to capture features at various scales.

II. Related Works

A. Optical Image Correlation for Displacement Estimation

Over the past few decades, sub-pixel OIC-based methods utilizing satellite and aerial imagery have evolved into two primary approaches: spatial [17], [30] and frequency (Fourier) [1], [2], [26] domain techniques. In both cases, traditional methods [2], [17], [31], [32] usually make use of normalized cross correlation (NCC) or its equivalent in the frequency domain, involving estimating ground deformation by comparing a source s and a template t using a sliding window approach, assuming a uniform translation between them.

In the spatial domain, MicMac [17] estimates displacements using normalized cross-correlation (NCC), refining precision iteratively with a smaller search space and applying spatial regularization (isotropic and non-isotropic) to reduce noise. This regularization penalizes high-frequency components or unrealistic shifts based on surface regularity [33]. MicMac makes use of a non-linear cost [17] in order to limit the impact of the noisy signal on the whole measurement:

$$\text{Cost} = \left(1 - \left(\frac{\max(\text{Cor}, C^{\min}) - C^{\min}}{1 - C^{\min}}\right)^{\gamma}\right) \cdot (1 - C^{\min}) \tag{1}$$

Here, Cor represents the correlation score, and Cost is the resulting value of the cost function. C^{min} is the correlation threshold parameter; when the correlation score is below this threshold it has no influence. The parameter γ, controls the influence of correlation scores: the higher the value of γ, the higher the influence of correlation scores close to 1 [17]. While methods using NCC are robust to noise and large spatial differences [17], they are computationally expensive, especially with high-resolution data and large displacements requiring larger search spaces. Additionally, correlation can fail or be biased by noise (e.g., atmospheric changes, sensor artifacts, or geometric distortions) and struggles in uniform or low-contrast areas like snowfields or deserts, leading to poor estimations. The assumption of uniform local translation also breaks down near discontinuities, such as faults, causing biased estimations in rupture zones [34].

One of the most used spectral OIC method is COSI-Corr [24]. It estimates displacement purely in the frequency domain, minimizing residuals in the normalized cross-power spectrum through gradient descent, with an initial pixel-wise displacement estimation followed by a sub-pixel refinement step (typically made with a smaller correlation window). Iterative adaptive frequency masking reduces noisy high frequencies [2], and frequency masking is applied to only select parts of the cross-power spectrum where the phase information is valid, to reduce aliasing or optical aberrations [2]. Larger correlation windows improve robustness but reduce spatial detail [17], [24], with a common trade-off being a 32×32 window for accuracy and noise control. However, without regularization, the method can produce outliers and non-physical solutions, thus making it more sensitive to noise compared to spatial-domain methods. Both spatial and frequency approaches assume homogenous translation within the sliding window, leading to bias near discontinuities or in areas with complex displacements.

B. Deep Learning for Optical Flow Estimation

FlowNet [35] paved the way for the utilization of CNN-based methods to solve optical flow problems, by introducing

two models FlowNetS and FlowNetC, based on the U-Net architecture [36]. FlowNetS takes as input the two acquired frames stacked together and feeds them through the network in order to estimate the optical flow, while FlowNetC processes these two images separately before merging feature maps with a correlation layer. FlowNet 2.0 [37] extended FlowNet notably by stacking multiple networks in order to refine the displacement estimation, but also to tackle small displacements and correct artifacts in the estimated displacement fields. One of the blocks of the FlowNet 2.0 architecture, called FlowNet2-SD, is specifically designed to address noise in sub-pixel displacement field.

Subsequent models, such as SpyNet [38], PWC-Net [39], and LiteFlowNet [40], introduced further enhancements, including iterative refinement through a coarse-to-fine pyramid, with a focus on large motion estimation, typically used in applications like video tracking, autonomous driving, or motion analysis in natural scenes. However, their performance on sub-pixel accuracy remains limited, as these methods prioritize handling larger displacements and fast-moving objects. RAFT [41] and IRR [42] improved refinement using recurrent units, but they still struggle with sub-pixel accuracy in noisy environments.

Recent advances in optical flow estimation have seen the introduction of Transformer-based deep learning architectures [43]. Perceiver IO [44] learns optical flow regression with a transformer-based architecture. GMFlow [45] make used of a customized Transformer for feature enhancement. FlowFormer [46] adopts an encoder-decoder architecture for cost volume encoding and decoding. These new transformer-based algorithms use global feature matching, to overcome the challenge of accurately measuring large displacements, meaning sub-pixel precision is often compromised.

CNN- and Transformer-based models are usually evaluated on different benchmark datasets of natural RGB images, such as MPI Sintel [47], KITTI [48] and FlyingThings3D [49]. The MPI-Sintel dataset contains large motions and complex effects like motion blur, with over 17.5% of its pixels having displacements greater than 20 pixels. The KITTI dataset consists of real-world driving scenes, featuring with varied lighting, and over 16% of pixels showing motion beyond 20 pixels. The Middlebury dataset [50] containing complex motions with less than 3% of the pixels having a displacement over 20 pixels, is considered to contain the smallest displacements by the community. The results attained on this latter dataset is the only one where the aforementioned state-of-the-art methods reaches sub-pixel precision (i.e. where the error is below the pixel size), and gives hope in adapting such models in other fields and reaching sub-pixel accuracy. However, these datasets and models are trained with natural images (i.e., everyday RGB photos or video frames of scenes), and focus on rigid-body motions, which differs drastically from our case of using satellite images, and detecting ground deformation. Additionally, the time between acquisitions in our case allows for changes beyond just deformation, which must also be accounted for.

C. Optical Flow Beyond Natural Image Applications

Previously proposed optical flow models have typically seen application across a range of research fields. For example, FlowNetS [51] and LiteFlowNet [52] have been used for fluid motion estimation. In medical imaging, GlueNet [53] adapted FlowNet 2.0 for displacement estimation in ultrasound elastography. StrainNet [54] adapted FlowNetS to estimate displacement fields occurring on the surface of deformed materials using speckle images. Finally, in the context of ground motion estimation, Montagnon et al. with CNN-DIS [23] developed a CNN architecture to retrieve accurate displacements from optical satellite images, using a two-step process to estimate, first, the initial pixel-wise displacements, and second, the sub-pixel component of displacement. All of these methods managed to reach sub-pixel accuracy on their data.

Using the same trained models for another problem without any refinement or fine-tuning is not always possible, because of differences in the data (e.g. different formats, ranges of displacement or noise and artifacts). Each method described above has been trained on a dataset representative of the addressed problem (e.g. a PIV dataset representing the fluid flow for [40], a speckle dataset for StrainNet [54], or a discontinuity remote sensing dataset for CNN-DIS [23] in order to optimize the model architecture and parameters to the specific problem. In addition, aforementioned datasets and most optical flow datasets (such as MPI Sintel and FlyingThings3D) use synthetics, creating reliable accurate ground truth. In the ground deformation measurement field, the dynamic nature of ground motion (involving small, often subtle displacements that occur over time) and the lack of dense and accurate ground-based measurement data makes creating a ground truth database exceptionally challenging. Therefore, to adapt models from the literature, we need a synthetic representative dataset suited for ground deformation estimation from optical satellite images. Such a dataset has been proposed by CNN-DIS [23], serving as a valuable preliminary approach; however, it lacks the size and complexity required for generating large comprehensive displacement maps and remains inherently simplified.

D. Contributions

In this study, we propose a deep convolutional neural network architecture trained on a new realistic synthetic earthquake displacement dataset **FaultDeform** to retrieve ground displacements produced by earthquakes from optical images with sub-pixel precision and high accuracy. We develop the **GeoFlowNet** pipeline shown on Figure 1 that uses a U-net architecture model to solve the full-scale displacement estimation problem, with significant speed performance thanks to its GPU implementation. We therefore overcome the basic limitation, i.e. spread across state-of-the-art methods, of a uniform local shift, by estimating extensive full-scale displacement fields, while leveraging the multi-scale architecture to capture valuable features at various scales.

This paper contributes to the literature in two major aspects:

- The creation of the first custom realistic dataset **FaultDeform** tailored for full-scale ground motion displacement

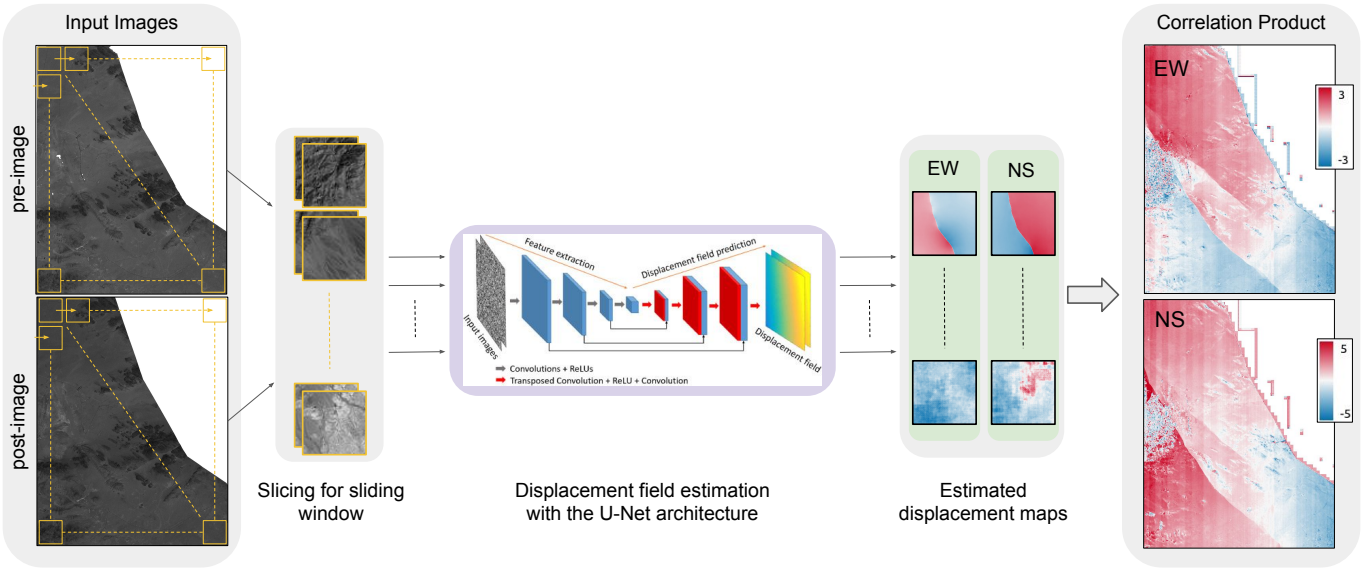

Fig. 1. Overview of the **GeoFlowNet** pipeline. Pre- and post-event satellite images are divided into 256×256 patches (sliding window), which are processed by the U-Net CNN to estimate displacement fields. The intermediate operations within the network, including convolutional layers and activation functions such as ReLU (Rectified Linear Unit), are illustrated in the middle panel. The resulting displacement fields are combined to generate full-scale displacement maps in the East-West (EW) and North-South (NS) directions. On the right, correlation products show the spatial variation of the displacement components.

estimation, offering the first ground-truth database based on displacements synthetically generated from realistic faults. The dataset is available to the community at https://doi.org/10.57745/G02ZXZ.

- The development of the first end-to-end U-net-based framework **GeoFlowNet** estimating accurate ground motion displacement maps able to give sub-pixel accuracy and with displacement ranging from several pixels to less than 0.1 pixels, radically surpassing speed of state-of-the-art correlators, and available at gricad-gitlab.univ-grenoble-alpes.fr/montagtr/GeoFlowNet.

III. Generation of the dataset FaultDeform

When applying existing trained models to very different data without fine-tuning, results are typically very poor, especially since our problem is notably distinct. Optical satellite images used for measuring ground movements face unique noise challenges compared to standard natural images. Also, the displacements from earthquakes, are typically small (less than 10 m) and often occur in the sub-pixel domain, decaying smoothly over large distances from sharp surface discontinuities.

Given these complexities and the lack of sufficient ground truth data for natural earthquakes, there is a need to create a training dataset accurately reflecting these challenges, incorporating real satellite images and realistic synthetic ground displacement maps.

Our FaultDeform database is generated in 3 steps. First, we download and prepare real satellite images acquired at different times for stable regions containing no natural ground displacements. Second, we generate a sequence of realistic faults and compute realistic earthquake displacement maps. Finally, we warp/resample the satellite images to include the synthetic displacement maps.

1) We downloaded Landsat 8 images from Collection 2 using the Landsatxplore interface [55], from 14 different areas, spanning stable regions in which no ground deformation has occurred. The locations cover diverse regions across Africa and Asia, including scenes from the deserts of Chad, Niger, and Libya; the rugged terrains of Afghanistan, Pakistan, and Iran; mountainous areas in Mongolia and Uzbekistan; and varied landscapes in Ethiopia, Sudan, and Yemen. We extracted the panchromatic band of Level 2 collection products, with a footprint of 185 km (cross-track) by 180 km (along-track), and ground resolution of 15 m (pixel dimensions of 12300×12000). Images from the same scene are globally co-registered (assuming a simple translation) using phase correlation (a Python implementation of the matrix-multiply DFT method available in the Scikit-image library [32]). It reduces global misregistrations that might remain after processing by the USGS. Images are then stacked onto the same grid. Every scene contains 25 acquisitions, ranging from 2015 to 2023. Iteratively, pairs of 1024×1024 pixels are extracted from the stacked images. For each scene, we randomly selected one window from any of the first 24 images in the acquisition sequence, which becomes the "pre-image." The corresponding "post-image" was then selected from the immediately following acquisition. For every scene, we collected 9k window pairs (126k pairs in total). We make sure the collected pairs respect a global correlation score minimum of 90%, in order to reduce images too heavily corrupted by noise (clouds, natural changes, etc.).
2) We developed a pipeline to randomly generate realistic fault discontinuities with fractally rough geometries and

slip distributions within a homogeneous, linear elastic half-space. Using analytical expressions [56], [57], we relate slip on triangular fault patches to surface displacement, calculating the resulting surface displacement field from the synthetic fault geometry and earthquake slip distribution. These faults adhere to natural earthquake length-displacement scaling laws [58], [59], rupture only the seismogenic part of the crust (<15 km), and exhibit geometric roughness (self-affine scaling, Hurst exponent 0.8 [60]), fractal slip distributions [61], [62], and reduced slip in the uppermost crust (shallow slip deficit [63], although shallow slip deficits are not systematic and can result from data resolution and inversion choices [64]). Fault lengths range from 5 to 150 km, down-dip widths from 5 to 17 km, strikes from 0 to 360°, and dips from 30 to 90°. We only generate strike-slip faults with dips varying +/-60° from vertical, as horizontal displacements are more resolvable in nadir-view satellite images. Fault models are discretized with triangular displacement elements (TDEs) using an unstructured meshing approach (Mesh2D [65]), achieving high surface resolution and coarser depth resolution (which is less sensitive to spatial details in the surface displacement field). Observation points are generated using an unstructured mesh, densifying near surface ruptures where we see more significant spatial variation in the surface displacements (displacements, and thus high strain gradients, attenuate rapidly with distance from the rupture). These displacements are resampled onto a regular grid using the griddata function in SciPy.

With this process, we created 14k earthquake displacement maps (8000 × 8000 pixels), where every fault is centered. We extract nine image windows (1024 × 1024 pixels) from each displacement map, arranged in a 3×3 grid centered around the middle of the map, which gives 126k displacement maps in total.

3) To warp the second windows (all the post-images) we use a quintic-order spline re-sampling algorithm [66]. The precision of this resampling approach (∼ 1/100th pixel) is significantly higher than the precision of state-of-the-art correlators (<1/10th pixel). For each displacement map–satellite image pair, we make four different warps: we normalize every displacement map and multiply by 4 different scaling factors, each randomly selected from a different float range, in order to obtain four different scaled displacement maps. The four scaling factor intervals [0.01, 1], [1, 5], [5, 15] and [15, 50] indicate the ranges from which a scaling factor is randomly drawn for each sample to multiply its displacement map, controlling the amplitude of displacements seen by the model. These scaling factors are important for controlling the range of displacements we want a trained model to be able to retrieve. Therefore, we obtain 4 × 126k samples for our total dataset.

This procedure is summarized on Figure 2. We split our synthetic database that we name FaultDeform-126k (being the full dataset with 126k samples) into 3 datasets : (1) training with 90k samples, (2) validation with 18k samples, and (3) test datasets with 18k samples. The train dataset is only used for training our two models, the validation dataset is used to tune the parameters and evaluate them during training, and the test dataset is for comparing our trained models to other state-of-the-art methods, in order to have fair comparison metrics and criteria.

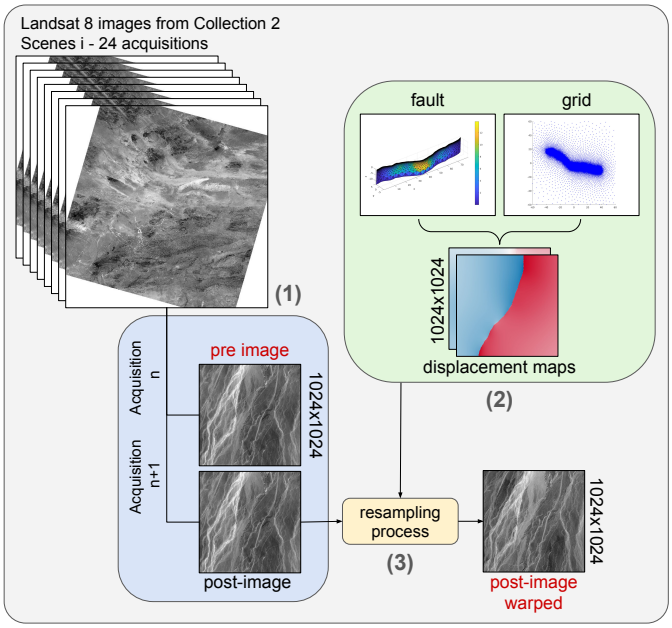

Fig. 2. Procedure to create one sample (pair of pre and post_warped in red, associated with the corresponding displacement maps) of the FaultDeform dataset.

TABLE I
SUMMARY OF FAULTDEFORM DATASET CONFIGURATION AND DISTRIBUTION

Component	Details	Values
Satellite Images	Number of Scenes	14
	Images per Scene	25
Extracted Window Pairs	Window Size	1024 × 1024 pixels
	Number of Pairs per Scene	9000
	Total Number of Pairs	126000
Displacement Maps	Fault Maps Size	8000 × 8000 pixels
	Number of Fault Maps	14000
	Displacement Maps Size	1024 × 1024 pixels
	Number of Displacement Maps per Fault Map	9
	Total Number of Displacement Maps	126000
Scaled Displacement Maps	Scaling Factors Intervals	[0.01, 1], [1, 5], [5, 15], [15, 50]
Final Dataset Splits	Training Dataset	4 × 10000 samples
	Validation Dataset	4 × 2000 samples
	Test Dataset	4 × 2000 samples

After a sensitivity study VI, we found that only 10k samples were necessary for training, and reduced the database by a factor or 9, using **10k samples for train**, **2k samples for validation** and **2k samples for test**, selected randomly from our FaultDeform-126k database. The dataset is available for the community on Data Repository Grenoble Alpes.

Table I shows the dataset configuration and distribution.

IV. PROPOSED DISPLACEMENT ESTIMATION METHOD

FlowNetS has shown its adaptability to different research fields ([51], [54]) and is a key reference in CNN-based optical flow estimation while also being computationally efficient.

The architecture named FlowNet2-SD from FlowNet 2.0 [37] is a modified FlowNetS architecture built to address the specific problem of sub-pixel motion tracking, due to the modifications they made (no stride and reduced kernel size at the beginning of the network). We created a similar model called GeoFlowNet-A with the same modifications.

We also made the same modifications StrainNet [54]) made to the FlowNetS architecture. Originally, the FlowNetS architecture provided an optical flow of 1/4 of the original image resolution, so interpolation was required to retrieve the full-scale displacement map. Simplifications of the architecture gave StrainNet-F, built with 4 downsamplings followed by 4 up-samplings, thereby outputting full-resolution maps. We made these modifications to GeoFlowNet-A and GeoFlowNet-B architectures to directly retrieve full-scale displacement maps.

We selected both GeoFlowNet-A (for its modifications specifically designed for tracking sub-pixel motion) and GeoFlowNet-B (for its ability to retrieve full-scale displacement maps and reaching sub-pixel precision) to be trained with our FaultDeform train database.

When training a data-driven model, a set of hyper-parameters must be chosen to control the learning process:

1) As with FlowNet, we selected the Adam optimizer [67], a gradient-based optimization algorithm with adaptive learning rates that achieves faster convergence compared to standard stochastic gradient descent with momentum.
2) We do not train our model using 1024×1024 pixel windows. Instead, we randomly select 256×256 crops at each epoch, employing this as a form of data augmentation to introduce more variability into the training process. This approach significantly reduces computational burden, thereby improving computation time and mitigating memory constraints. With this approach, we can raise the batch size to 32.
3) After a few tests, we initiate the learning rate at 5e-4 and divide it by 2 every 20 epochs.
4) We use endpoint error (EPE) as the training loss, defined as the average Euclidean distance between the predicted and ground truth flow vectors across all pixels. It is computed at the 5 scales of the network, and averaged with a weighting coefficient of 1, except the one of the last layer, which is set to 6 to give more importance to the last level forming the network. This loss corresponds to multi-scale EPE; the real EPE (EPE computed between the output and the target) is the metric used for validation.
5) We trained our two models GeoFlowNet-A and GeoFlowNet-B on 50k batch iterations (40 epochs), on $4 \times$ 10k train samples

With this strategy, training each of the networks on an NVIDIA TESLA V100 SXM2 32Go GPU required 9 hours 30 minutes for GeoFlowNet-A and 10 hours for GeoFlowNet-B.

A. Implemented pipeline: GeoFlowNet

To evaluate and compare our models on real data, we decided to implement a pipeline called GeoFlowNet (shown on Figure 1) that takes as input two large satellite images, and outputs directly the full-scale displacement maps.

GeoFlowNet constitutes a ready-to-use tool to retrieve displacement maps from satellite images, and its code is available online in a Git repository https://gricad-gitlab.univ-grenoble-alpes.fr/montagtr/GeoFlowNet. It utilizes PyTorch for deep learning and Rasterio for geospatial data processing and automates the process of computing displacement maps between a pair of satellite images using trained GeoFlowNet-A or GeoFlowNet-B models. An automated preprocessing step aligns input rasters prior to computing the displacement map.

We use a sliding window procedure to iteratively evaluate $k \times k$ large displacement maps from cropped $k \times k$ input pairs, extracted from the input satellite images. To reconstruct the full displacement fields corresponding to the satellite images, the procedure utilizes a stride, retaining only the unique (non-overlapping) displacements from all the maps, in order to minimize edge effects as much as possible.

Default parameters are $k = 256$ with a stride of $s = 128$. Models have thus been trained with 256×256 pixels windows, but can process any large window $k \times k$ pixels images to speed up the evaluation. While limited by memory and time consumption, performing a training on a larger window could allow the model to learn even larger spatial relations.

TABLE II
PARAMETERS USED FOR DIFFERENT MODELS.

Model	Parameters	Values
GeoFlowNet-A	Window size k	256×256 (512×512 for synthetics)
	Stride s	128
GeoFlowNet-B	Window size k	256×256 (512×512 for synthetics)
	Stride s	128
COSI-Corr	Initial window size	64×64
	Final window size	32×32
	Mask threshold	0.9
	Number of iterations	2
MicMac	C^{min}	0.5
	γ	2
	Window size	9×9
	Regularization term	0.3
CNN-DIS	Pixel window size	32×32
	Sub-pixel window size	16×16

V. Experimental Results

As described in Section IV-A, we use our GeoFlowNet pipeline with default parameters, a window size of 256×256, and a stride of $s = 128$. For our synthetic comparative and real data studies, we use COSI-Corr, MicMac and CNN-DIS with all default parameters. The correlation parameters for all four correlators are summarized in Table II. All methods use a step of 1, to retrieve full scale displacement maps.

A. Comparisons on synthetics

We first make a quantitative and qualitative analysis based on our FaultDeform test dataset. We compare our 2 trained models with COSI-Corr (using the recent python implementation [68]), MicMac, and CNN-DIS [23].

Our test sample dimensions are 512×512 pixels from the 1024×1024 pixels test samples. We take our 4×2k test samples, and compute the real EPE for each sub-dataset, for each method. Results are shown in Table III. The first step of the CNN-DIS procedure, and MicMac, with its limited search space set up at 5 pixels, are not able to retrieve displacements larger than 5 pixels, so the evaluations on the [5, 15] and [15, 50] sub-dataset are invalid.

TABLE III
ERROR (REAL EPE) FOR EACH SYNTHETIC TEST SUB-DATASET (DISPLACEMENT FACTOR RANGES [0.01, 1], [1, 5], [5, 15], [15, 50])

Model \ Sub-dataset range	[0.01, 1]	[1, 5]	[5, 15]	[15, 50]
GeoFlowNet-A	**0.169**	**0.238**	**0.302**	**0.780**
GeoFlowNet-B	**0.169**	0.243	0.326	0.888
COSI-Corr	0.261	0.300	0.425	1.45
MicMac	0.193	0.253	x	x
CNN-DIS	0.328	0.527	x	x

Both GeoFlowNet versions (GeoFlowNet-A and GeoFlowNet-B) outperform COSI-Corr, MicMac and CNN-DIS on all the synthetic test sub-datasets using the EPE criterion. In addition, GeoFlowNet-A gives marginally better results on three sub-datasets out of four compared to GeoFlowNet-B. Thanks to a built-in regularization, MicMac results remains very close to our models. On the smaller subset, we can reach a mean error of less than 0.17 pixel: this would correspond to less than 9cm on a Pleiades image, or 2.5m for Landsat.

To have a deeper understanding of how the models behave on the synthetic test data, three examples of the test samples are shown on Figures 3, 4, 5 and 6. These examples are only 256×256 pixels because COSI-Corr does not pad boundary windows, retrieving 256×256 pixels displacement maps with a 512×512 pixels input and reporting NaN values for a 1/2 window dimension around the original extents of the image.

Figure 3 shows that both GeoFlowNet versions correctly retrieves the far-field and near-fault displacements, even when they are relatively small ($<$2.3 pixel for E-W and $<$0.9 pixel for N-S). COSI-Corr and CNN-DIS are significantly more noisy in the far-field (particularly on the N-S component), with common noisy features, which are suppressed by GeoFlowNet versions. MicMac manages to partially reduce the noise far-field thanks to a built-in regularization, yet it results in a very granulated fault trace (using default parameters). CNN-DIS has difficulty tracking such small displacements, particularly in the N-S component, which is due to the initial pixel-wise registration step that corrupts the refinement. In fact, the test images are more corrupted than the original CNN-DIS training dataset [23] which makes its first step unstable.

[0]Bold indicates the best results.

Figure 4 shows that with very large displacements ($>$30 pixels), COSI-Corr retrieves a highly inaccurate fault trace, especially in near-field. Even with different configurations for COSI-Corr, such as by raising the size of the initial window (128×128 and 256×256, to account for the larger displacement), we noticed no improvement. Aside from the boundary artifacts, MicMac is overall able to recover the fault and the displacement values, yet the fault angle is not correct (see Figure S2 in supp. for the residual maps). To note, for running MicMac on a +/- 45 pixel exploration (changing the Inc parameter from the default 2 to 45), the running time was increased by a factor of nearly 100. GeoFlowNet models manage to retrieve accurate displacements in both the far-field and near-fault zone, without any specific fine-tuning on the model. CNN-DIS results are not shown because the method is not designed to retrieve displacements larger than 5 pixels.

Focusing on the near-field estimation, we compare the GeoFlowNet-A method against COSI-Corr and MicMac on a third synthetic test sample (Figure 5 and Figure 6), using the same sample #3 but with two different scaling factors. We only show GeoFlowNet-A, as GeoFlowNet-B has very similar results). We zoom on the EW displacement results near the fault. We compute an averaged cross-profile (Figures 5(b) and 6(b)) by calculating the mean displacement across the fault over the yellow band from Figures 5(a) and 6(a). We finally estimate the Fault Zone Width (FZW), defined as the distance over which the fault-perpendicular strain exceeds 0.5% each side of the fault, in Figures 5(c) and 6(c). The FZW is estimated visually, based on the approximate inflection points of the horizontal displacements each side of the fault. For small displacements (example #3.A with scaling factor of 0.7, on Figure 5), COSI-Corr shows a moving, less realistic, and less accurate fault trace, with a wider FZW corrupted with outliers, and with noisier far field displacements. MicMac manages to delimit realistically the fault trace, but at the expense of spatial details, especially around the fault trace. This gives a granulated fault trace, which gives a smoothing of the discontinuity when computing the mean profile, resulting in a wider FZW. GeoFlowNet-A resolves a more accurate fault trace, with a narrower FZW, and smoother accurate far-field displacements (albeit with a slightly larger global bias). The estimated FZW is the closest to the ground truth, highlighting the superior performance of our model in resolving near-field characteristics of the fault zone (FZW of around 5 pixels, compared with 9 and 12 pixels for respectively MicMac and COSI-Corr). With larger displacements (example #3.B with scaling factor of 2.7, on Figure 6), we observe that COSI-Corr estimates a very sharp fault trace with a FZW of approx. 6 pixels (sharper than GeoFlowNet-A and MicMac, respectively with a FZW of 9 and 16 pixels). However, this is at the cost of an important localization error between the ground truth fault trace and that resolved with COSI-Corr: more than 5 pixels of offset are visible in the perpendicular distance. For MicMac, the granulated effect is even larger than in Figure 5 and the fault trace has a very unrealistic shape. The FZW of GeoFlowNet is thus better than MicMac, and it is correctly located (which is not the case for COSI-Corr).

These observations can also be clearly visible from the

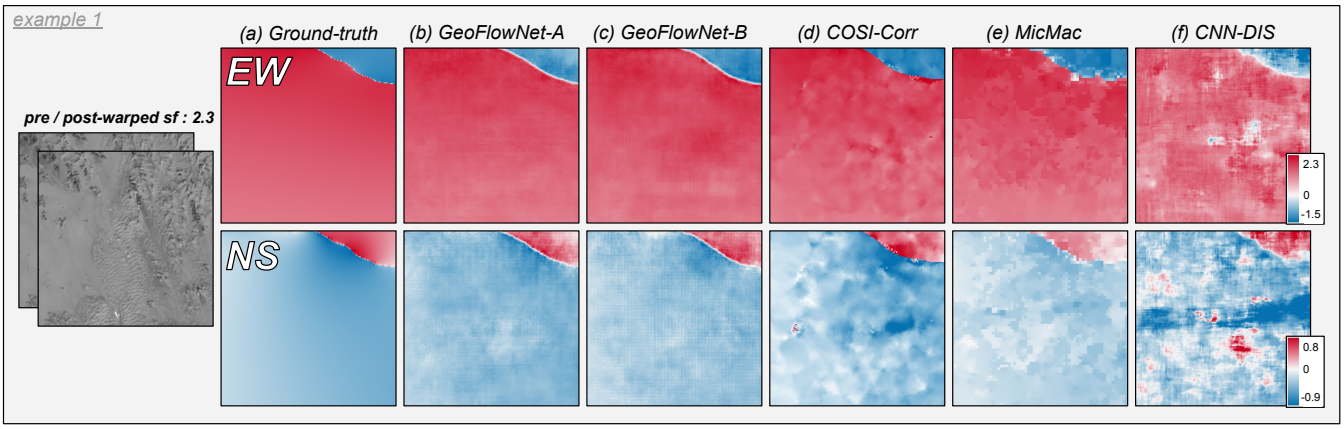

Fig. 3. Displacement fields (256×256 pixels) in East-West and North-South directions from (b) GeoFlowNet-A, (c) GeoFlowNet-B, (d) COSI-Corr, (e) MicMac and (f) CNN-DIS on the test synthetic example #1, with a scaling factor of 2.3. On the left, the pre and post-warped images and (a) the ground-truth displacement maps used to warp them. Results are expressed in pixels ; the scene corresponds to a desert-like environment with little vegetation cover, minimizing decorrelation effects

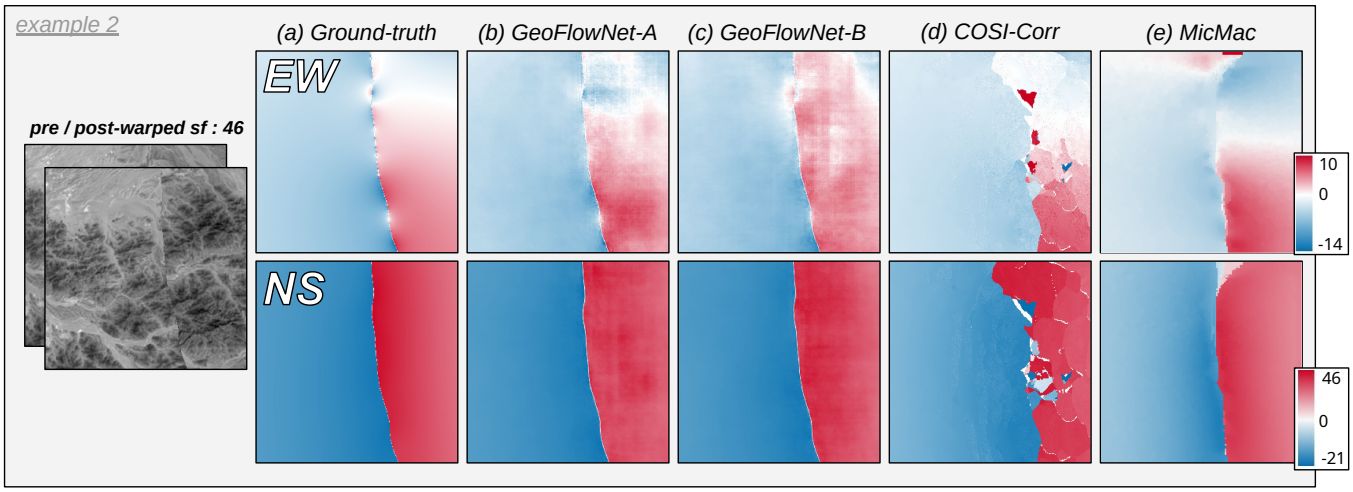

Fig. 4. Displacement fields (256×256 pixels) in East-West and North-South directions from (b) GeoFlowNet-A, (c) GeoFlowNet-B, (d) COSI-Corr, and (e) MicMac on the test synthetic example #2, with a scaling factor of 46. CNN-DIS is not optimized for such high deformations. On the left, the pre and post-warped images and (a) the ground-truth displacement maps used to warp them. Results are expressed in pixels.

residual maps (absolute error maps of each method) of the 3 examples, provided in the Supplementary Material.

B. Comparisons on real data

1) 2019 Ridgecrest and 2013 Balochistan cases: We aim to demonstrate that GeoFlowNet is effective at accurately resolving real earthquake displacements preserved in optical satellite images by applying our pipeline to two earthquake case studies: 2019 Ridgecrest, and 2013 Balochistan.

In July 2019, the Ridgecrest earthquake sequence in Southern California's Mojave desert produced cross-fault ruptures from a Mw 6.4 foreshock on a NE-SW-striking left-lateral fault, followed by a Mw 7.1 mainshock on a conjugate NW-SE-striking fault [69]–[71]. Multiple studies utilized optical image correlation (OIC) methods to quantify the surface displacement fields and assess the degree of surface slip localization (on- vs off-fault deformation) across the rupture zone [71]–[74]. Using the surface displacements to extract the strain field, these studies mapped the along-strike variation in fault zone width, which in turn correlates with features in the geology, fault structure, and seismic radiation of the earthquake. For Ridgecrest, we used pre- and post-event 50 cm resolution Pléiades satellite images [74], respectively acquired on 23rd June 2012 and 8th September 2019. This allowed us to assess our model's performance on data acquired with a different sensor to that used in the training (i.e. Landsat-8). Pléaides differs to Landsat-8 by being sensitive to different spatial wavelengths, different off-nadir acquisition angles, and different resolutions (0.5 m vs 15 m, thus spanning more than one order of magnitude difference). For this study, we used a 1024×1024 pixels crop located on the NW-SE-striking fault that ruptured in the Mw 7.1 earthquake, extracted from the $23k \times 30k$ pixels Pléiades images.

On 24th September 2013, a powerful earthquake struck

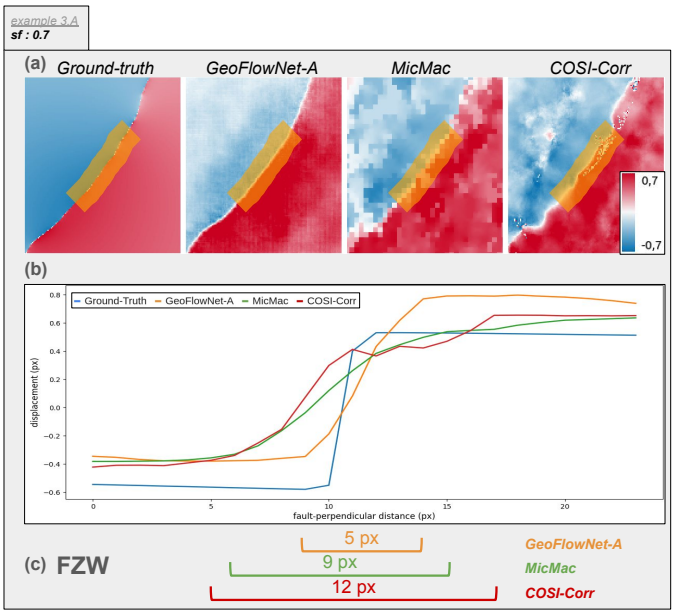

Fig. 5. Synthetic test sample #3-A (with scaling factor of 0.7). On (a), EW displacement maps from GeoFlowNet-A, MicMac and COSI-Corr. On (b) the mean of displacement profiles across the fault at the location represented in yellow. On (c) an estimation of the fault zone width corresponding to the three displacements results. Results are in pixels.

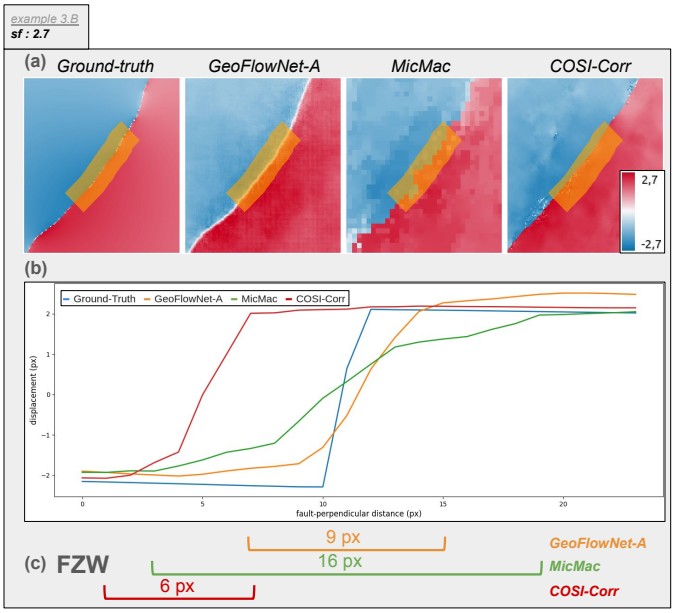

Fig. 6. Synthetic test sample #3-B (with scaling factor of 2.7). On (a), EW displacement maps from GeoFlowNet-A, MicMac and COSI-Corr. On (b) the mean of displacement profiles across the fault at the location represented in yellow. On (c) an estimation of the fault zone width corresponding to the three displacements results. Results are in pixels.

the Awaran district in the Balochistan province of Pakistan, generating widespread surface ruptures and displacement fields across the region [75], [76]. The earthquake, with a magnitude of Mw 7.7, nucleated south of the Chaman strike-slip fault and propagated southwestward along the Hoshab fault, producing a 200 km-long strike-slip surface rupture. For the Balochistan earthquake, we employed $15k \times 15k$ pixels Landsat imagery at its native 15 m resolution, matching the sensor type and resolution of our FaultDeform dataset. This provided an opportunity to test our model on real earthquake data with similar imaging characteristics, but in a different geological and tectonic context, and for an event where the deformation is mostly sub-pixel relative to the 15 m Landsat pixel size [77]. The pre-image of our example was acquired on 10th September 2013, and the post-image on 26th September 2013. For this second case study, the full $15k \times 15k$ pixel images are processed and compared, enabling to test GeoFlowNet on a large zone scenario. For both real case studies, all estimations were performed with a step of 1, i.e. a full dense estimation.

2) Qualitative comparisons: Both earthquakes involved significant ground displacements (up to 6 m for Ridgecrest, and up to 15 m for Balochistan [77]). Given the very high resolution imagery used for the Ridgecrest case (50 cm Pléiades imagery), and the medium resolution imagery used for Balochistan (15 m Landsat-8 imagery), we are able to test our full pipeline's capability at resolving both supra-pixel and sub-pixel displacements. Although we lack ground truth displacement maps of similar density for these real cases, very high-resolution rupture maps and dense field offset datasets exist for both Ridgecrest and Balochistan. A more extensive comparison, ensuring the accuracy and consistency of these ground truth datasets, would represent a valuable complement to our study, but remains beyond its current scope. Comparing our results with established methods such as COSI-Corr and MicMac helps us evaluate consistency. Visual inspection of the displacement fields near the fault ruptures also allows us to identify potential errors, artifacts, or bias based on their spatial distribution, offering another means for validating our approach.

Across both cases, GeoFlowNet produced displacement maps with noticeably less high frequency noise near the fault zones and in the far field compared to COSI-Corr, MicMac, and CNN-DIS (Figures 7 and 8). The smoother outputs might raise concerns about losing small fault features, but they offer an advantage in terms of stability, particularly in areas where displacement data are noisy or difficult to correlate.

Close to the fault rupture, both GeoFlowNet-A and B provide a more consistent and realistic representation of the displacement field, with fewer artifacts along the rupture traces compared to COSI-Corr on Figure 7. This is especially visible on the (b) segment (comparing 1.b and 2.b with 3.b) where COSI-Corr exhibits many outliers and unrealistic geometric roughness. Although MicMac yielded relatively clean maps, they are more pixelated (Figure 7.4.ab compared to the GeoFlowNet smoother displacement profiles, which preserve a sharper fault trace (however, we note that this pixelation effect can be mitigated by tuning the SsResolOpt option, albeit at the cost of much increased runtime). We also notice a rather uniform difference on each part of the fault between the GeoFlowNet-A and GeoFlowNet-B results, especially in the far field. In the North-East part of the fault, GeoFlowNet-B estimates positive displacements with a magnitude of around 6% (0.05 meters, or 0.025 pixels) higher than GeoFlowNet-A. In the South-East part of the fault, GeoFlowNet-A estimates negative displacements with a magnitude of 2% (0.013 meters,

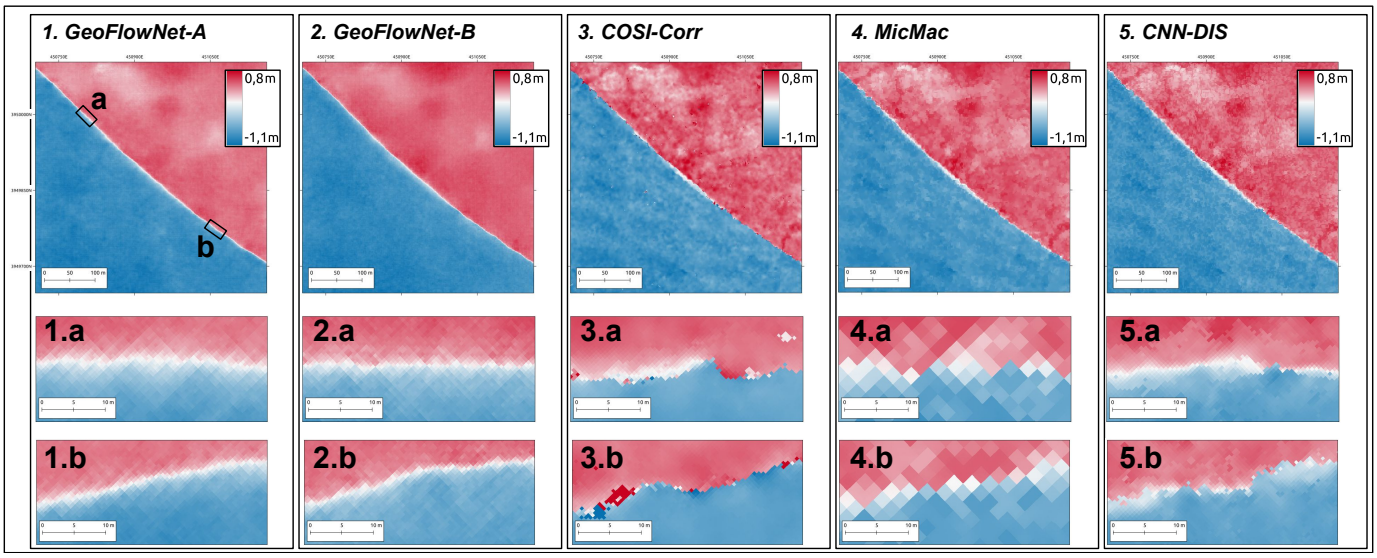

Fig. 7. EW component of the displacement maps from Pléiades images for the Ridgecrest earthquake. From left to right: GeoFlowNet-A, GeoFlowNet-B, COSI-Corr, MicMac and CNN-DIS displacement maps. Two segments a and b cropped from each displacement map are shown for every method. Results are expressed in meters, with one pixel being 50 cm.

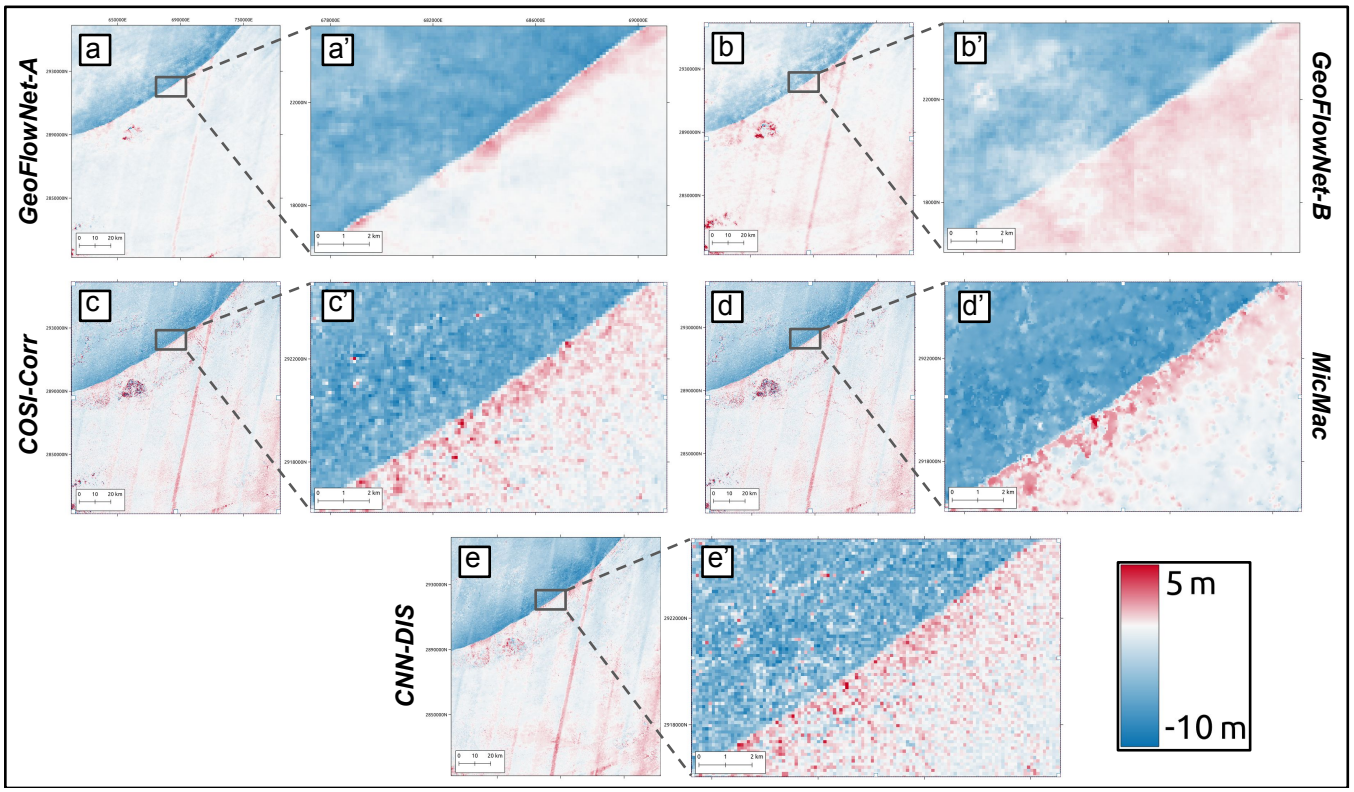

Fig. 8. EW component of the displacement maps from Landsat images for the Balochistan earthquakeof GeoFlowNet-A, GeoFlowNet-B, COSI-Corr, MicMac and CNN-DIS displacement maps . A zoom on the Hoshab fault is shown for every method. Results are expressed in meters, with one pixel being 15 meters.

or 0.0066 pixels) higher than GeoFlowNet-B.

For the Balochistan rupture (Figure 8), the fault trace is very well resolved by GeoFlowNet, despite some longer-wavelength patches of displacement that are likely to be slightly underestimated (whiter patches on the left side of GeoFlowNet-B result in b.). As in the Ridgecrest case, in the South-East part of the fault, GeoFlowNet-B estimates positive displacements with a magnitude of around 116% (0.41 meters, or 0.027 pixel) higher than A, and in the North-West part of the fault, GeoFlowNet-A estimates negative displacements with a magnitude of 43% (0.68 meters, or 0.045 pixel) higher than B. Comparing with MicMac and COSI-Corr, GeoFlowNet-A seems more accurate in both close- and far-field than GeoFlowNet-B. From the different results, we don't

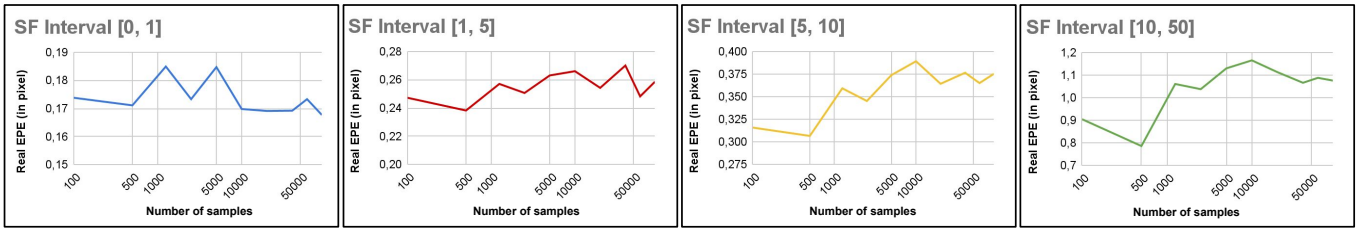

Fig. 9. Sensitivity plot showing the real EPE (in pixels) function of the size of the training dataset, for the four scaling factor intervals (SF) test sub-datasets.

TABLE IV
COMPUTATION TIME[1] OF AN 8000X8000 REAL IMAGE PAIR FOR GEOFLOWNET-A, CNN-DIS, COSI-CORR AND MICMAC: NUMBER OF DISPLACEMENT ESTIMATIONS (PROCESSED WINDOW PAIRS) PER SECOND. PARAMETERS FOR THE COMPARISONS ARE THE DEFAULT (SEE TABLE II)

	1 GPU	*32 CPU cores*	Displacement est. / s
GeoFlowNet-A	**44 s**	x	**1,454,545**
2-step CNN-DIS	2030 s	x	31,527
COSI-Corr	x	460 s	139,130
MicMac	x	720 s	88,889

see a particular pattern where GeoFlowNet-B would be clearly better than GeoFlowNet-A. Yet, we recommend users to test both architectures as a robustness check of the results.

We acknowledge the difficulty in interpreting such observations for such a complex case, where the displacements are all below the pixel size, with the majority being below 0.5 pixels (7.5 m), and especially in the far-field where displacements quickly attenuate to 0 m. Nevertheless, our results closely match those of state-of-the-art methods, and show that we are able to retrieve the deformation field across a range of displacements and imaging conditions.

3) Computation time: The GeoFlowNet pipeline is designed to run on GPU and processes large image patches, which allows for much faster processing compared to traditional CPU-based methods with local window estimation. We cropped a 8000×8000 pixels from the original Pléiades images (22976×29782 pixels) and processed this image pair with a step of 1 with the 4 methods. GeoFlowNet-A produces a full-scale displacement field in 44 s, while CNN-DIS (the 2-step pipeline [78]) takes 2030 s on the same GPU, COSI-Corr (using its ENVI implementation) 460 s, and MicMac 720 s, both on 32 CPU cores. Metrics on table IV give the number of displacement estimations per second for all 4 methods presented. The GeoFlowNet pipeline is drastically faster than all existing methods, notably 10 times faster than the fastest COSI-Corr, allowing the computation of full-scale displacement maps on areas covering 11.5×14.8 km² (22976×29782 pixels Pléiades images) in less than 8 minutes.

VI. SENSITIVITY STUDY

A. Selection of the amount of data

The number of training data samples is crucial for model performance. Too little data can lead to poor generalization, while too much can increase computational costs without improving accuracy. To address this, we conduct a sensitivity analysis on GeoFlowNet-A (similar results were found for GeoFlowNet-B) to determine the minimum number of samples required for accurate displacement estimations, assessed through both real End Point Error (EPE) metric on the synthetic test dataset and visual criteria on the real case of Balochistan. (Results on the real case of Ridgecrest showed no significant differences ; for the EPE, the lower the result, the more accurate the model.)

Increasing the number of samples does not yield any improvement in accuracy on our test data (Figure 9). In fact, it seems that more data lowers the accuracy for all scaling factors. However, it is important to note that these observations may not fully reflect performance in real-world scenarios.

We also show the results of the displacement maps on a 880×1256 pixels crop of Balochistan, shown on Figure 10, for a model trained with $4 \times 1k$ samples and a model trained with $4 \times 10k$ samples (respectively named GeoFlowNet-A 1k and GeoFlowNet-A 10k). This figure shows that the amount of data directly impacts the quality of the results on real data, with a significant reduction in noise correlated with topography. We performed other tests with a higher number of data (20k, 40k, 60k, and 90k), noticing negligible differences in the resulting maps, concluding that 10k training samples represents a good trade-off between generalizability and over-kill.

B. Selection of sub-datasets

In this section, we evaluate the choice of the synthetic training samples based on their displacement size, i.e. the scaling factor that was applied to the displacement maps before warping. Training exclusively on the sub-dataset with scaling factors limited to the range [0, 1] fails to converge to an optimal solution; therefore, results from this subset are excluded from the analysis. As expected, the first four lines of Table V show that the best results are achieved when a model is trained and tested on sub-datasets with the same scaling factor interval. However, these models perform poorly when evaluated on the sub-datasets with different scaling factor ranges. In particular, training with small scaling factor ranges limits the model's ability to produce accurate results on larger scaling factor ranges. Yet, when training on the four sub-datasets (last column), the drop in accuracy on each test set is minimal, with a decrease of around 0.05 pixels in EPE for the smallest test set range. Given these observations, we selected the model trained on the full set of sub-datasets.

[1]Bold indicates the best results.

TABLE V
TEST PERFORMANCE ACROSS DIFFERENT TRAINING AND TESTING INTERVALS. BOLD INDICATES THE BEST OF EACH LINE. S.F. = SCALING FACTOR

	Training sets					
Testing sets	[1, 5]	[5, 15]	[15, 50]	[1, 5] & [5, 15]	[1, 5] & [5, 15] & [15, 50]	[0, 1] & [1, 5] & [5, 15] & [15, 50]
s.f. [0, 1]	**0.131**	0.227	0.31	0.17	0.203	0.153
s.f. [1, 5]	**0.183**	0.228	0.326	0.209	0.233	0.236
s.f. [5, 15]	1.6	**0.254**	0.37	0.258	0.298	0.306
s.f. [15, 50]	11.4	6.14	**0.622**	6.21	0.722	0.771
All scaling factors	13.31	6.85	1.63	6.85	**1.47**	**1.47**

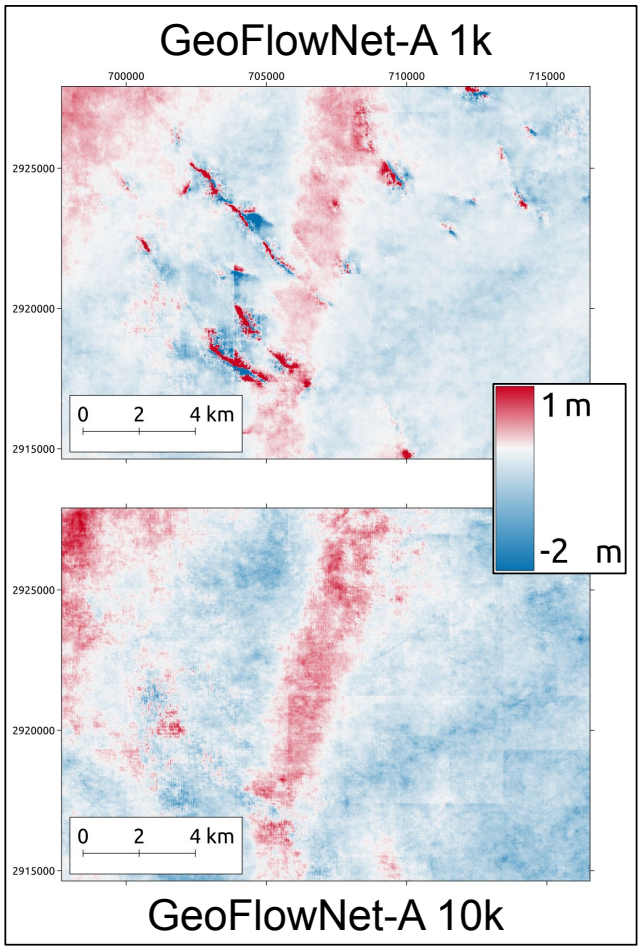

Fig. 10. 880×1256 pixels from the Balochistan displacement estimations using GeoFlowNet-A: (top): trained with $4 \times 1k$ samples and (bottom) trained with $4 \times 10k$ samples (respectively named GeoFlowNet-A 1k and GeoFlowNet-A 10k).

Although training models on multiple sub-datasets with different scaling factors allows for greater generalizability and better performance across the total displacement range, the importance of sub-pixel precision becomes resolution-dependent. For example, in 10 m resolution imagery, achieving sub-pixel precision of 1/10th of a pixel translates to a spatial precision of 1 m. In contrast, for 0.5 m resolution data, sub-pixel precision of 1/10th of a pixel corresponds to a higher precision of 5 cm. Therefore, for very high-resolution datasets, such as those from Pléiades or WorldView, achieving extremely fine sub-pixel precision may be less critical: we recommend in this case to use a specific small-scale model.while GeoFlowNet-A and B already perform well across a wide range of displacement magnitudes, a dedicated model trained on a narrower displacement range could further optimize performance for specific scenarios if required.

VII. LIMITATIONS AND FUTURE WORKS

The dataset used to train the network is critical for the quality of the synthetic experiments and the resulting displacement maps on real scenarios. Enhancing the quality of the training data could help to further mitigate inherent biases and improve accuracy and precision (particularly important for cases where displacement is largely sub-pixel, such as for Balochistan). Improvements such as better data scheduling (highlighted by FlowNet 2.0 [37]) and more diverse and realistic fault simulations would also help to reduce bias.

Moreover, our model operates as a single-step estimation process, relying on a single forward pass through a convolutional neural network. In contrast, implementations like Flownet 2.0 utilize a sequence of models to progressively refine the output using intermediate warping operations, which has been shown to enhance precision [37]. Such an approach presents a promising avenue for future research, as it could potentially lead to improved accuracy in deformation estimates.

Finally a significant limitation of our current work is the lack of quantitative validation in real-world scenarios. This absence of empirical validation hinders a comprehensive understanding of the models' actual performance, emphasizing the need for further studies to validate our findings under various conditions. In particular, we have for now trained and tested our model on regions without too much change between the acquisitions (with correlation score minimum of 90%). We would need to incorporate pairs with more natural changes (i.e. by taking pairs with a lower correlation threshold when generating the synthetics) to be able to work in vegetated environments.

VIII. CONCLUSION

In this work, a deep learning-based scheme, GeoFlowNet, has been proposed for ground motion estimation from satellite optical images with various resolutions, containing a wide range of displacements, and attaining sub-pixel precision. Two CNN-based U-net versions, GeoFlowNet-A and GeoFlowNet-B have been tested to extract detailed shallow-level features

across multiple scales. Based on our experiments (see Table III), both architectures achieve comparable performance, with GeoFlowNet-A slightly outperforming GeoFlowNet-B on most sub-datasets. We therefore recommend GeoFlowNet-A as the default choice, while GeoFlowNet-B remains available for robustness checks or further experiments. Additionally, we presented our new realistic dataset FaultDeform generated from realistic faults, used for training the models to retrieve full-scale ground motion displacement fields. A major upgrade is the significantly reduced computation time required to obtain full-scale (pixel-resolution) displacement maps, especially valuable when using optical imagery covering large regions, with a high number of pixels, along with a marked improvement in near-fault displacement accuracy compared to COSI-Corr and MicMac.

The growing volume of optical imagery from satellite providers and government agencies increases the opportunity for data-driven methods to emerge. Our model's speed and robustness across several displacement scales makes it well-suited for quickly processing this data, particularly in applications involving time-series analysis that require rapid processing. Furthermore, automation and cloud-based systems, combined with GPU processing, will streamline the management of these large datasets.

ACKNOWLEDGMENT

This work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02), and the SaTellite project (ANR-22-CE01-0028). The authors would like to thank Université Grenoble Alpes, INSU PNTS, CNES, and CDP Risk for funding, and GRICAD infrastructure (gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, for the computations. This work was granted access to the HPC resources of IDRIS under the allocation 2024-AD011012345 made by GENCI. We also thank Solène Antoine for kindly providing the Pléiades orthoimages of the Ridgecrest region. The authors would like to thank Juliette Bertrand for her fruitful suggestions during discussions.

REFERENCES

[1] N. Van Puymbroeck, R. Michel, R. Binet, J.-P. Avouac, and J. Taboury, "Measuring earthquakes from optical satellite images," *Applied Optics*, vol. 39, no. 20, pp. 3486–3494, 2000.

[2] 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 Transactions on Geoscience and Remote Sensing*, vol. 45, no. 6, pp. 1529–1558, 2007.

[3] R. Zinke, J. Hollingsworth, J. F. Dolan, and R. Van Dissen, "Three-dimensional surface deformation in the 2016 mw 7.8 kaikōura, new zealand, earthquake from optical image correlation: Implications for strain localization and long-term evolution of the pacific-australian plate boundary," *Geochemistry, Geophysics, Geosystems*, vol. 20, no. 3, pp. 1609–1628, 2019.

[4] R. Grandin, A. Socquet, R. Binet, Y. Klinger, E. Jacques, J.-B. De Chabalier, G. King, C. Lasserre, S. Tait, P. Tapponnier *et al.*, "September 2005 manda hararo-dabbahu rifting event, afar (ethiopia): constraints provided by geodetic data," *Journal of Geophysical Research: Solid Earth*, vol. 114, no. B8, 2009.

[5] J. Hollingsworth, S. Leprince, F. Ayoub, and J.-P. Avouac, "Deformation during the 1975–1984 krafla rifting crisis, ne iceland, measured from historical optical imagery," *Journal of Geophysical Research: Solid Earth*, vol. 117, no. B11, 2012.

[6] A. Bonforte, P. J. González, and J. Fernández, "Joint terrestrial and aerial measurements to study ground deformation: application to the sciara del fuoco at the stromboli volcano (sicily)," *Remote Sensing*, vol. 8, no. 6, p. 463, 2016.

[7] T. R. Walter, "Low cost volcano deformation monitoring: optical strain measurement and application to mount st. helens data," *Geophysical Journal International*, vol. 186, no. 2, pp. 699–705, 2011.

[8] P. Lacroix, "Landslides triggered by the gorkha earthquake in the langtang valley, volumes and initiation processes," *Earth, Planets and Space*, vol. 68, no. 1, pp. 1–10, 2016.

[9] P. Lacroix, G. Bièvre, E. Pathier, U. Kniess, and D. Jongmans, "Use of sentinel-2 images for the detection of precursory motions before landslide failures," *Remote Sensing of Environment*, vol. 215, pp. 507–516, 2018.

[10] P. Lacroix, G. Araujo, J. Hollingsworth, and E. Taipe, "Self-entrainment motion of a slow-moving landslide inferred from landsat-8 time series," *Journal of Geophysical Research: Earth Surface*, vol. 124, no. 5, pp. 1201–1216, 2019.

[11] D. Scherler, S. Leprince, and M. R. Strecker, "Glacier-surface velocities in alpine terrain from optical satellite imagery—accuracy improvement and quality assessment," *Remote Sensing of Environment*, vol. 112, no. 10, pp. 3806–3819, 2008.

[12] F. Herman, B. Anderson, and S. Leprince, "Mountain glacier velocity variation during a retreat/advance cycle quantified using sub-pixel analysis of aster images," *Journal of Glaciology*, vol. 57, no. 202, pp. 197–207, 2011.

[13] T. Heid and A. Kääb, "Evaluation of existing image matching methods for deriving glacier surface displacements globally from optical satellite imagery," *Remote Sensing of Environment*, vol. 118, pp. 339–355, 2012.

[14] F. Herman, O. Beyssac, M. Brughelli, S. N. Lane, S. Leprince, T. Adatte, J. Y. Lin, J.-P. Avouac, and S. C. Cox, "Erosion by an alpine glacier," *Science*, vol. 350, no. 6257, pp. 193–195, 2015.

[15] M. Marchandon, J. Hollingsworth, and M. Radiguet, "Origin of the shallow slip deficit on a strike slip fault: Influence of elastic structure, topography, data coverage, and noise," *Earth and Planetary Science Letters*, vol. 554, p. 116696, 2021.

[16] A. M. Rodriguez Padilla and M. E. Oskin, "Displacement Hazard from Distributed Ruptures in Strike-Slip Earthquakes," *Bulletin of the Seismological Society of America*, vol. 113, no. 6, pp. 2730–2745, 09 2023. [Online]. Available: https://doi.org/10.1785/0120230044

[17] A.-M. Rosu, M. Pierrot-Deseilligny, A. Delorme, R. Binet, and Y. Klinger, "Measurement of ground displacement from optical satellite image correlation using the free open-source software micmac," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 100, pp. 48–59, 2015.

[18] F. Provost, D. Michéa, J.-P. Malet, E. Boissier, E. Pointal, A. Stumpf, F. Pacini, M.-P. Doin, P. Lacroix, C. Proy *et al.*, "Terrain deformation measurements from optical satellite imagery: The mpic-opt processing services for geohazards monitoring," *Remote Sensing of Environment*, vol. 274, p. 112949, 2022.

[19] C. Vogel, A. Bauder, and K. Schindler, "Optical flow for glacier motion estimation," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 1, pp. 359–364, 2012.

[20] C. Zhu, X. Li, C. Wang, B. Zhang, and B. Li, "Deep learning-based coseismic deformation estimation from insar interferograms," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[21] B. Rouet-Leduc, R. Jolivet, M. Dalaison, P. A. Johnson, and C. Hulbert, "Autonomous extraction of millimeter-scale deformation in insar time series using deep learning," *Nature communications*, vol. 12, no. 1, p. 6480, 2021.

[22] E. C. Reinisch, C. J. Abolt, E. M. Swanson, B. Rouet-Leduc, E. E. Snyder, K. Sivaraj, and K. C. Solander, "Advancing the limits of insar to detect crustal displacement from low-magnitude earthquakes through deep learning," *Remote Sensing*, vol. 16, no. 11, p. 2019, 2024.

[23] T. Montagnon, S. Giffard-Roisin, M. Dalla Mura, M. Marchandon, E. Pathier, and J. Hollingsworth, "Sub-pixel displacement estimation with deep learning: Application to optical satellite images containing sharp displacements," *Journal of Geophysical Research: Machine Learning and Computation*, vol. 1, no. 4, p. e2024JH000174, 2024.

[24] S. Leprince, F. Ayoub, Y. Klinger, and J.-P. Avouac, "Co-registration of optically sensed images and correlation (cosi-corr): An operational methodology for ground deformation measurements," in *2007 IEEE*

international geoscience and remote sensing symposium. IEEE, 2007, pp. 1943–1946.

[25] J. Hollingsworth, F. Ayoub, M.-P. Doin, S. Daout, H. Perfettini, G. Peltzer, and S. Samsonov, "Characterization and removal of shadow bias from optical image correlation; application to the 2013 balouchistan earthquake," in *AGU Fall Meeting Abstracts*, vol. 2017, 2017, pp. G43A–0909.

[26] X. Tong, Z. Ye, Y. Xu, S. Liu, L. Li, H. Xie, and T. Li, "A novel subpixel phase correlation method using singular value decomposition and unified random sample consensus," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 8, pp. 4143–4156, 2015.

[27] A. J. Plaza and C.-I. Chang, *High performance computing in remote sensing*. CRC Press, 2007.

[28] Z. Malenovskỳ, H. Rott, J. Cihlar, M. E. Schaepman, G. García-Santos, R. Fernandes, and M. Berger, "Sentinels for science: Potential of sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land," *Remote Sensing of environment*, vol. 120, pp. 91–101, 2012.

[29] M. A. Wulder, T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Allen, M. C. Anderson, A. S. Belward, W. B. Cohen *et al.*, "Current status of landsat program, science, and applications," *Remote sensing of environment*, vol. 225, pp. 127–147, 2019.

[30] E. Rupnik, M. Daakir, and M. Pierrot Deseilligny, "Micmac–a free, open-source solution for photogrammetry," *Open geospatial data, software and standards*, vol. 2, no. 1, pp. 1–9, 2017.

[31] H. Foroosh, J. B. Zerubia, and M. Berthod, "Extension of phase correlation to subpixel registration," *IEEE transactions on image processing*, vol. 11, no. 3, pp. 188–200, 2002.

[32] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, "Efficient subpixel image registration algorithms," *Optics letters*, vol. 33, no. 2, pp. 156–158, 2008.

[33] M. Pierrot-Deseilligny and N. Paparoditis, "A multiresolution and optimization-based image matching approach: An application to surface reconstruction from spot5-hrs stereo imagery," *Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, no. 1/W41, pp. 1–5, 2006.

[34] J. Hollingsworth, S. Daout, M.-P. Doin, and M. Cantraine, *Optical Geodesy and the Measurement of Ground Deformation by Image Correlation*. Cham: Springer International Publishing, 2024, pp. 89–129. [Online]. Available: https://doi.org/10.1007/978-3-031-59306-2_5

[35] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. Van Der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 2758–2766.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[37] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "Flownet 2.0: Evolution of optical flow estimation with deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2462–2470.

[38] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4161–4170.

[39] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8934–8943.

[40] T.-W. Hui, X. Tang, and C. C. Loy, "Liteflownet: A lightweight convolutional neural network for optical flow estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8981–8989.

[41] Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 402–419.

[42] J. Hur and S. Roth, "Iterative residual refinement for joint optical flow and occlusion estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5754–5763.

[43] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[44] A. Jaegle, S. Borgeaud, J.-B. Alayrac, C. Doersch, C. Ionescu, D. Ding, S. Koppula, D. Zoran, A. Brock, E. Shelhamer *et al.*, "Perceiver io: A general architecture for structured inputs & outputs," *arXiv preprint arXiv:2107.14795*, 2021.

[45] H. Xu, J. Zhang, J. Cai, H. Rezatofighi, and D. Tao, "Gmflow: Learning optical flow via global matching," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 8121–8130.

[46] Z. Huang, X. Shi, C. Zhang, Q. Wang, K. C. Cheung, H. Qin, J. Dai, and H. Li, "Flowformer: A transformer architecture for optical flow," in *European conference on computer vision*. Springer, 2022, pp. 668–685.

[47] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12*. Springer, 2012, pp. 611–625.

[48] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 3354–3361.

[49] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovitskiy, and T. Brox, "A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4040–4048.

[50] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International journal of computer vision*, vol. 92, pp. 1–31, 2011.

[51] S. Cai, S. Zhou, C. Xu, and Q. Gao, "Dense motion estimation of particle images via a convolutional neural network," *Experiments in Fluids*, vol. 60, pp. 1–16, 2019.

[52] S. Cai, J. Liang, Q. Gao, C. Xu, and R. Wei, "Particle image velocimetry based on a deep learning motion estimator," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 6, pp. 3538–3554, 2019.

[53] M. G. Kibria and H. Rivaz, "Gluenet: Ultrasound elastography using convolutional neural network," in *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation: International Workshops, POCUS 2018, BIVPCS 2018, CuRIOUS 2018, and CPM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16–20, 2018, Proceedings*. Springer, 2018, pp. 21–28.

[54] S. Boukhtache, K. Abdelouahab, F. Berry, B. Blaysat, M. Grediac, and F. Sur, "When deep learning meets digital image correlation," *Optics and Lasers in Engineering*, vol. 136, p. 106308, 2021.

[55] Y. Forget, G. Riembauer, and MrChebur, "yannforget/landsatxplore: v0.10," https://doi.org/10.5281/zenodo.4543601, 2021, zenodo, Version 0.10.

[56] B. J. Meade, "Algorithms for the calculation of exact displacements, strains, and stresses for triangular dislocation elements in a uniform elastic half space," *Computers & geosciences*, vol. 33, no. 8, pp. 1064–1075, 2007.

[57] M. Nikkhoo and T. R. Walter, "Triangular dislocation: an analytical, artefact-free solution," *Geophysical Journal International*, vol. 201, no. 2, pp. 1119–1141, 2015.

[58] S. G. Wesnousky, "Displacement and geometrical characteristics of earthquake surface ruptures: Issues and implications for seismic-hazard analysis and the process of earthquake rupture," *Bulletin of the Seismological Society of America*, vol. 98, no. 4, pp. 1609–1632, 2008.

[59] C. H. Scholz, *The mechanics of earthquakes and faulting*. Cambridge university press, 2019.

[60] T. Candela, F. Renard, Y. Klinger, K. Mair, J. Schmittbuhl, and E. E. Brodsky, "Roughness of fault surfaces over nine decades of length scales," *Journal of Geophysical Research: Solid Earth*, vol. 117, no. B8, 2012.

[61] C. Milliner, C. Sammis, A. Allam, J. Dolan, J. Hollingsworth, S. Leprince, and F. Ayoub, "Resolving fine-scale heterogeneity of co-seismic slip and the relation to fault structure," *Scientific reports*, vol. 6, no. 1, p. 27201, 2016.

[62] R. Amey, L. Wedmore, L. C. Gregory, A. J. Hooper, K. J. McCaffrey, M. W. Wilkinson, H. Goodall, and J. Woodman, "Not just rough around the edges: Fractal properties of exhumed fault surfaces in the italian apennines," in *AGU Fall Meeting Abstracts*, vol. 2017, 2017, pp. T52C–08.

[63] Y. Fialko, D. Sandwell, M. Simons, and P. Rosen, "Three-dimensional deformation caused by the bam, iran, earthquake and the origin of shallow slip deficit," *Nature*, vol. 435, no. 7040, pp. 295–299, 2005.

[64] S. L. Antoine, Y. Klinger, K. Wang, and R. Bürgmann, "Coseismic shallow slip deficit accounted for by diffuse off-fault deformation," *Geophysical Research Letters*, vol. 51, no. 24, p. e2024GL110798, 2024.

[65] D. Engwirda, "Locally optimal delaunay-refinement and optimisation-based mesh generation," PhD thesis, Faculty of Science, School of Mathematics and Statistics, University of Sydney, 11 2014.
[66] R. Wiemker, "Registration of airborne scanner imagery using akima local quintic polynomial interpolation," in *Presented at the Second International Airborne Remote Sensing Conference and Exhibition*, vol. 24, 1996, p. 27.
[67] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017. [Online]. Available: https://arxiv.org/abs/1412.6980
[68] S. Aati, C. Milliner, and J.-P. Avouac, "A new approach for 2-d and 3-d precise measurements of ground deformation from optimized registration and correlation of optical images and ica-based filtering of image geometry artifacts," https://doi.org/10.1016/j.rse.2022.113038, 2022, software.
[69] Z. E. Ross, B. Idini, Z. Jia, O. L. Stephenson, M. Zhong, X. Wang, Z. Zhan, M. Simons, E. J. Fielding, S.-H. Yun *et al.*, "Hierarchical interlocked orthogonal faulting in the 2019 ridgecrest earthquake sequence," *Science*, vol. 366, no. 6463, pp. 346–351, 2019.
[70] W. D. Barnhart, G. P. Hayes, and R. D. Gold, "The july 2019 ridgecrest, california, earthquake sequence: Kinematics of slip and stressing in cross-fault ruptures," *Geophysical Research Letters*, vol. 46, no. 21, pp. 11 859–11 867, 2019.
[71] Y. Magen, A. Ziv, A. Inbal, G. Baer, and J. Hollingsworth, "Fault rerupture during the july 2019 ridgecrest earthquake pair from joint slip inversion of insar, optical imagery, and gps," *Bulletin of the Seismological Society of America*, vol. 110, no. 4, pp. 1627–1643, 2020.
[72] W. D. Barnhart, R. D. Gold, and J. Hollingsworth, "Localized fault-zone dilatancy and surface inelasticity of the 2019 ridgecrest earthquakes," *Nature Geoscience*, vol. 13, no. 10, pp. 699–704, 2020.
[73] C. Milliner, A. Donnellan, S. Aati, J.-P. Avouac, R. Zinke, J. F. Dolan, K. Wang, and R. Bürgmann, "Bookshelf kinematics and the effect of dilatation on fault zone inelastic deformation: Examples from optical image correlation measurements of the 2019 ridgecrest earthquake sequence," *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 3, p. e2020JB020551, 2021.
[74] S. L. Antoine, Y. Klinger, A. Delorme, K. Wang, R. Bürgmann, and R. D. Gold, "Diffuse deformation and surface faulting distribution from submetric image correlation along the 2019 ridgecrest, california, ruptures," *Bulletin of the Seismological Society of America*, vol. 111, no. 5, pp. 2275–2302, 2021.
[75] J.-P. Avouac, F. Ayoub, S. Wei, J.-P. Ampuero, L. Meng, S. Leprince, R. Jolivet, Z. Duputel, and D. Helmberger, "The 2013, mw 7.7 balochistan earthquake, energetic strike-slip reactivation of a thrust fault," *Earth and Planetary Science Letters*, vol. 391, pp. 128–134, 2014.
[76] A. Vallage, Y. Klinger, R. Grandin, H. Bhat, and M. Pierrot-Deseilligny, "Inelastic surface deformation during the 2013 mw 7.7 balochistan, pakistan, earthquake," *Geology*, vol. 43, no. 12, pp. 1079–1082, 2015.
[77] R. D. Gold, N. G. Reitman, R. W. Briggs, W. D. Barnhart, G. P. Hayes, and E. Wilson, "On- and off-fault deformation associated with the september 2013 mw 7.7 balochistan earthquake: Implications for geologic slip rate measurements," *Tectonophysics*, vol. 660, pp. 65–78, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0040195115004461
[78] T. Montagnon, "cnn4l-discontinuities [software]," Jun. 2024.