



HAL
open science

Sub-Pixel Displacement Estimation with Deep Learning: Application to Optical Satellite Images Containing Sharp Displacements

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

► To cite this version:

Tristan Montagnon, Sophie Giffard-Roisin, Mauro Dalla Mura, Mathilde Marchandon, Erwan Pathier, et al.. Sub-Pixel Displacement Estimation with Deep Learning: Application to Optical Satellite Images Containing Sharp Displacements. 2023. hal-04256681v1

HAL Id: hal-04256681

<https://hal.science/hal-04256681v1>

Preprint submitted on 24 Oct 2023 (v1), last revised 2 Nov 2023 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sub-Pixel Displacement Estimation with Deep Learning: Application to Optical Satellite Images Containing Sharp Displacements

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

Abstract—Optical image correlation is an effective method for remotely assessing ground movement from satellite imagery, e.g. associated with natural disasters, such as earthquakes. This approach enables the characterization, and identification of the causal factors and mechanisms underlying such processes. By employing sub-pixel correlation algorithms, one can obtain highly accurate (m-to-cm level) displacement fields at high spatial resolution (dm-to-cm) by comparing satellite images acquired before and after a period of movement. However, this method generally assumes a homogeneous translation of all pixels within a given correlation window, which will lead to biased estimates of ground displacement if the real case is not well represented by such a simplification. This is particularly true when resolving ground displacements next to sharp gradients (or discontinuities) in displacement, such as those found in the near-field of earthquake surface ruptures.

In this paper, we present an innovative deep learning method estimating sub-pixel displacement maps from optical satellite images for the retrieval of ground displacement. From the generation of a realistic simulated database, comprising Landsat-8 satellite image pairs containing simulated sub-pixel shifts and sharp discontinuities, we developed a Convolutional Neural Network (CNN) able to retrieve sub-pixel displacements. The comparison to state-of-the-art correlation methods shows that our pipeline is able to significantly reduce the estimation bias around fault ruptures, thus leading to more accurate characterization of the near-field strain in surface rupturing earthquakes. Application of our model to the 2019 Ridgecrest earthquake (USA) demonstrates the ability of our model to accurately resolve ground displacement using real satellite images.

The numerical experiments proposed in this paper can be reproduced with Pytorch code available on Github at <https://github.com/tristanmtg/cnn4l-discontinuities>.

This manuscript was submitted on October 9, 2023, for review.

Thanks to IDEX Université Grenoble Alpes, INSU PNTS, CNES, and CDP Risk for funding. (corresponding author: James Hollingsworth.)

Tristan Montagnon, Sophie Giffard-Roisin, James Hollingsworth and Erwan Pathier are with Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France. (e-mail: tristan.montagnon@univ-grenoble-alpes.fr; sophie.giffard@univ-grenoble-alpes.fr; james.hollingsworth@univ-grenoble-alpes.fr; erwan.pathier@univ-grenoble-alpes.fr)

Mauro Dalla Mura is with CNRS, Inria, Grenoble Institute of Technology (Grenoble INP), Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab), Université Grenoble Alpes, 38000 Grenoble, France, and also with the Institut Universitaire de France (IUF), 75231 Paris, France (e-mail: mauro.dalla-mura@gipsa-lab.grenoble-inp.fr)

Mathilde Marchandon is with Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität München, Germany, and also with Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, 38000 Grenoble, France. (e-mail: m.marchandon@lmu.de)

The numerical experiments proposed in this paper can be reproduced with Pytorch code available on Github at <https://github.com/tristanmtg/cnn4l-discontinuities>.

Index Terms—optical image correlation, image registration, satellite imagery, deep learning, geodesy, surface rupture, bias

I. INTRODUCTION

A. General context

PRECISE estimation of ground displacement at regional scales is fundamental for the study of natural hazards, such as earthquakes, volcanoes, landslides, as well as monitoring of glaciers [1], [2], [3], [4], [5], [6]. In the case of earthquakes, an accurate and unbiased estimation of ground deformation, especially in the near-field of surface ruptures, is crucial to address the location, geometry, spatial distribution (on- vs off-fault) of slip, and the down-dip slip distributions of the causative fault. Accurate characterization of the near-field displacement around surface ruptures in turn provides valuable constraints needed to understand the physics of earthquake slip [7], and to anticipate the seismic hazard posed to neighboring infrastructure and populations [8]. Such observations are essential for addressing many areas of earthquake science, including fault mechanics, seismology, and structural/geological evolution of faulting, as well as earthquake engineering, including seismic hazard assessment, and seismic design. Recent observations have shown that the degree of slip localization can vary along the fault in individual earthquakes, as well as between different earthquakes in different settings [9], [10], [11]. The mechanism by which slip localization varies in a surface rupturing earthquake isn't well understood, although several controlling parameters have been proposed (e.g. lithology, rupture velocity, earthquake magnitude, fault geometry, topography, etc.) [12], [13], [14]. One difficulty in addressing this issue lies in the relative paucity of observational data currently available, rendering a comprehensive statistical analysis of the various parameters, and their trade-offs, challenging. Furthermore, of the modest number of case studies available, the observations are additionally complicated by the quality of the displacement maps obtained by image correlation. In particular, our ability to accurately resolve the degree of slip localization may be limited by the methodology used to generate the displacement map.

B. Optical image correlation

Optical satellite geodesy has revolutionized how we characterize ground deformation associated with natural hazards such

as earthquakes [15], [16]. Optical Image correlation (OIC) is an imaging technique able to estimate a full deformation field between two optical satellite images acquired over the same area at different times (separated by hours up to years). The main advantage of OIC, compared with more precise satellite geodetic techniques such as Interferometric Synthetic Aperture Radar (InSAR), lies in the ability to resolve large and high-strain displacements with very high spatial detail and using images with relatively long temporal baselines (up to many years). OIC has therefore been widely applied to the study of various sources of ground motion, such as surface rupturing earthquakes, volcanoes, landslides, and glacial movement [17]. In the study of earthquakes, where displacements are generally small relative to the pixel size, sub-pixel precision is critical for accurately capturing the displacement field, especially close to surface ruptures, where the deformation may become complex (e.g. sharp discontinuities, distributed off-fault deformation, secondary faulting, along-strike slip variability, etc). In this sense, earthquakes can represent a relatively challenging feature to measure with OIC, compared with large, fast-moving features such as glaciers.

In recent years, OIC has gained utility in constraining the near-field displacement around earthquake surface ruptures [18], [19], [20], [12], [11], [21], [14], in part due to the increasing availability of high resolution optical satellite datasets, as well as the difficulty in constraining these regions using InSAR (which typically decorrelate due to cycle skipping associated with the large strain gradients, [22], [23]). OIC can therefore provide valuable constraints on the localization (or distribution) of slip between the primary fault core and the neighboring (decimeter-scale) off-fault damage zone.

Over the past few decades, various OIC methods have been developed for quantifying image displacement [24], [25], [26], [27]. Focusing on the approaches largely used in the remote sensing community (involving satellite and aerial imagery), these can be broadly grouped in correlation-based techniques working in the spatial [28], [29], [30], [31] or frequency (Fourier) domain [1], [2], [29], [32]. Spatial cross-correlation is a fundamental technique for estimating ground deformation by comparing reference and target images through a sliding window approach. This technique has been used (and refined) to efficiently solve the correlation problem [33], [29]. Frequency-based correlation is considered to be more accurate and faster (given that convolution in the spatial domain is equivalent to multiplication in the frequency domain) than commonly used spatial correlation methods such as normalized cross correlation [34]. Both methods generally employ a block-based matching scheme, employing a sliding window to capture the local displacement field over the full image. However, this may lead to high frequency noise in the final displacement map (at wavelengths similar to the sliding window dimension). Therefore, some approaches also make use of multi-scale regularization schemes to mitigate this effect [2], [29].

The Semi-Global Matching (SGM) spatial-domain algorithm of [25], and similar stereo-matching techniques [35], employs a global optimization step that enforces consistency and regularization at both local and larger scales. This helps

to produce more accurate and smooth displacement maps, which can capture a broad range of displacement magnitudes. Such regularization schemes also permit smaller correlation windows to be used (as low as 3-by-3 pixels), thereby allowing more spatial detail to be resolved. It is these aspects which have made SGM popular for disparity estimation in stereo matching applications, i.e. for producing accurate depth maps, and generating digital elevation models from stereo imagery.

Optical Flow (OF) is a related approach which aims to track the motion of pixels from a sequence of images using local spatial gradients [36], [37], [38]. Nevertheless, SGM and OF generally have lower sub-pixel precision and accuracy than dedicated sub-pixel correlators employing larger windows, especially when using multi-temporal images acquired under different reflectance and incidence conditions.

Satellite images are subject to sources of noise leading to correlation bias or even temporal decorrelation. For example, the difference in illumination between two acquisitions, typically acquired with a difference of weeks to months, can introduce topographically-correlated artifacts in the resulting estimated displacement map [3], [39], [1], [6]. Reflectance changes due to vegetation or anthropogenic changes between acquisitions may also hinder the correlation over long time periods [1]. Noise present in the acquisitions can be also a source of errors in the correlation results. Low level image pre-processing carried out by the data providers, such as image resampling performed in the mosaicking of the raw acquisitions from single arrays for forming an image in staggered pushbroom sensors or the projection in a geographic reference system can introduce additional errors such as periodic artifacts appearing at specific spatial frequencies [3], [32]. Spatial regularization schemes [29], or frequency masking [3] have been used to help mitigate high frequency noise in the resulting displacement maps, although they can result in smoothing of the spatial detail. However, these approaches do not address limitations in the correlation process, such as the underlying assumption of rigid translations over the correlation window [40], or spatial smoothing associated with the correlation kernel [19], which may bias the displacements in the vicinity of sharp discontinuities typical with earthquake surface ruptures. In particular, smoothing of the displacement field in the vicinity of a sharp discontinuity will artificially distribute displacement into the neighboring areas. In the case of earthquakes, this effect may severely bias our estimates of slip localization, and the partitioning of slip between the fault plane and the surrounding damage zone, as well as estimates of near-field strain [14].

C. Data-driven image registration

Convolutional neural networks (CNNs) [41], [42] have emerged as a powerful tool in all image processing fields. A CNN is able to *learn* a task from a set of training images in order to then *estimate* a prediction for new test images. The architecture of a CNN consists of a succession of different layers. First, the input layer takes as input one or more images. Convolutional layers apply a set of *learnable* convolution filters (known as kernels) to the input. Each filter scans over

the input and computes a weighted sum of pixel values within its receptive field. Then, an activation function (commonly Rectified Linear Unit, or ReLU) is applied element-wise to introduce non-linearity into the network. Optional pooling layers reduce the spatial dimensions of the feature maps produced by the convolutional layers. Max-pooling layers retain the maximum value within a sliding window, effectively downsampling the outputs of convolutional layers (feature maps). The output layer produces the final estimations by linearly combining the outputs of the last layer. Training a CNN for a specific task such as image registration consists in updating the network’s weights (convolution filters) iteratively with a specific gradient descent optimization algorithm to minimize the loss function, here measuring the dissimilarity between the network’s predictions and the ground truth transformations. Therefore, training requires *labeled* data, which consists here of source and target image pairs (input) with known transformations (output, i.e. label). The data is typically split in 3 sets: training set for model learning, validation set for monitoring the performance during the training phase, and test set to assess the model’s performance on unseen examples once the model is trained. We call *hyper-parameters* the set of variables controlling the deep learning model architecture and algorithm, as opposed to the *parameters*, or weights, which are optimized during training. The hyper-parameters are the number and size of layers, the optimization algorithm (e.g., Adam, SGD), the learning rate (step size at each gradient descent iteration), etc.

Image registration and displacement field estimation from optical images has been successfully addressed by recent data-driven approaches, and in particular CNNs, e.g. in medical-imaging [43], [44], and remote sensing [45]. In the domain of satellite images registration, CNNs have been shown to improve global image registration performance [46], [45]. Displacement field estimation between two images can also be efficiently solved by deep learning, e.g. treating optical flow estimation as a learning task [47]. However, the large majority of registration problems focus on the estimation of large displacements (> 1 pixel) from temporally dense datasets, while the estimation of sub-pixel shifts from temporally limited and distant acquisitions has been little studied. Several recent studies have demonstrated the potential of data-driven approaches to retrieve sub-pixel displacements [48], [49], [50], [51]. However, to the best of our knowledge, no application in remote sensing has yet been proposed for resolving sub-pixel displacements, particularly in the presence of local complexities such as sharp discontinuities, from optical satellite data, which are fundamental features for automatic slip characterization techniques.

D. Contributions

We developed a pipeline using a convolutional neural network (CNN) to solve the sub-pixel displacement estimation problem. CNNs have the capability (1) to be robust to variations (noise and lightning for example); (2) to learn relevant features from the data without manual feature engineering, which simplifies the modeling process and enhances the ability

to capture complex patterns; (3) to generalize well to unseen data with similar characteristics to the training set, making them suitable for tasks where correlation properties may vary. A novel aspect of the proposed technique is that we address the specific challenge faced when discontinuities are present within the sliding window, i.e. in the near-field of fault ruptures. We achieve this by training a sub-pixel model on a realistic synthetic dataset that includes samples mimicking real fault discontinuities (See Fig. 1). Our implementation includes a two-step pipeline that can effectively estimate (1) significant displacements exceeding one pixel by initially employing a coarse model in a pixel-scale registration step, and subsequently (2) refining the sub-pixel details centered around the pixel-scale displacement using a sub-pixel model.

This paper contributes to the literature in three major aspects:

- 1) The creation of large realistic synthetic training sets that allow data-based techniques to learn retrieving sub-pixel ground surface deformation.
- 2) The development of the first end-to-end neural network-based OIC method, as a 2-step procedure: first a coarse estimation (larger than one pixel), then a finer estimation (sub-pixel refinement).
- 3) Comparisons with state-of-the-art correlation methods COSI-Corr [3] and MicMac [29] to quantitatively (with synthetic realistic images) and qualitatively evaluate the results. The latter case is addressed through a study of the 2019 Ridgecrest earthquake [52], [21], where we aim at characterizing the co-seismic displacement.

II. RELATED WORKS

OIC is based on image registration principles, in which the goal is to estimate a displacement (or disparity) between common features (represented by pixels) present within two images [53]. Sub-pixel OIC refers to the estimation of the displacement field between two images with a precision less than the image resolution (i.e. < 1 pixel) [54]. Attaining sub-pixel precision requires interpolating into the sub-pixel domain. In order to solve this problem, one common approach (in block-based matching schemes) is to assume a homogeneous displacement between the two images (i.e. the displacement is smooth enough to consider a locally rigid deformation). In most remote sensing cases, the transformation between the pre- and post-images is dominated by a simple rigid translation in 2-dimensions, and rotations are assumed to be small. In these cases, we can thus simplify our registration problem to just 2 degrees of freedom.

1) *Spatial correlation methods*: A correlation measurement between two images of scalar values I_1 (reference) and I_2 (template) of same size consists of retrieving a similarity score, that indicates how well the content of the two images match and is typically used in template matching problems [55]. The cross-correlation C_{I_1, I_2} between image I_1 and I_2 is computed by integrating within the image domain the pixel-wise product between the two images for each 2D spatial shift, represented by a translation vector of components (x, y) :

$$C_{I_1, I_2}(x, y) = \sum_n \sum_m I_1(x + n, y + m) \cdot I_2(n, m). \quad (1)$$

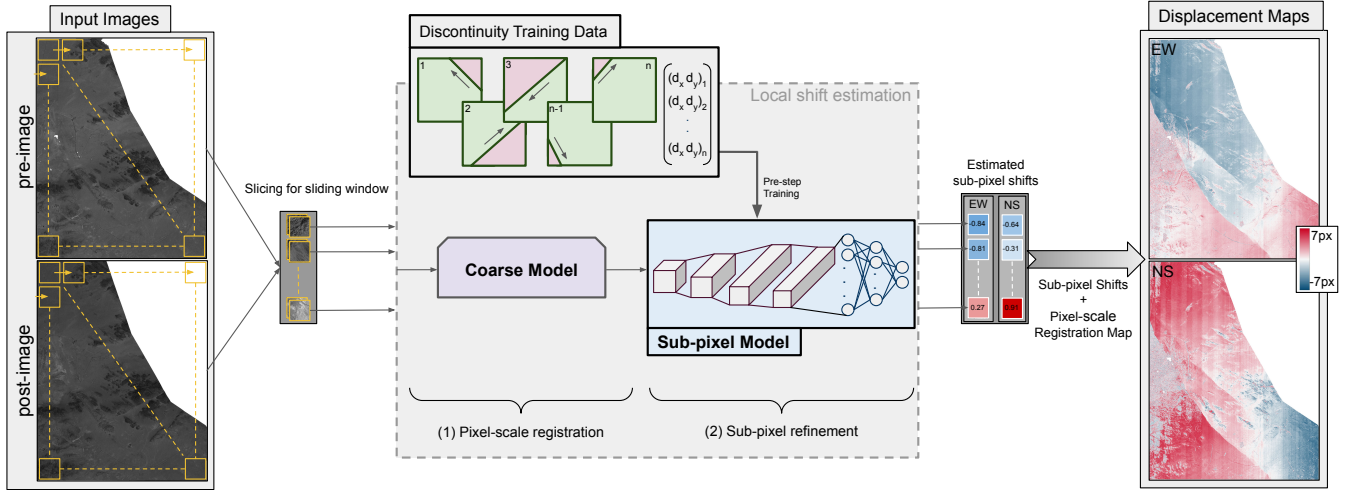


Fig. 1: From left to right: A pair of input images (pre and post) are acquired on two different dates. The sub-pixel displacement between the two images is retrieved using first a coarse model registering every pair of windows pixel-wise, followed by the sub-pixel model that extract the sub-pixel shift; displacements is given in the row (North-South) and column (East-West) direction. Our sub-pixel model is trained with data containing discontinuities.

The estimation of the integer 2D shift between the two images can be performed by finding the couple (x, y) , denoted by (x^*, y^*) , that maximises the correlation score:

$$(x^*, y^*) = \underset{x, y}{\operatorname{argmax}} C(x, y). \quad (2)$$

We recall that (x^*, y^*) might not be unique and that corresponds to a global translation of the two images, meaning that all pixels in the template image will get shifted by the same amount.

Traditional methods [3], [29] use the principle of cross-correlation to retrieve displacement maps that describe the surface displacement field between two satellite images in two directions: East-West (horizontal) and North-South (vertical). Let us assume we have two images I_1 and I_2 , and we want to retrieve the 2D displacement field between the two images. In contrast to the global registration presented previously, in this context the translation vector can be potentially different in each pixel. To produce a map of the displacement field based on cross-correlation, we consider a spatially local approach consisting of splitting the template image I_2 into a set of rectangular windows of a given size. Each window $\{W_{k,l}\}$ corresponds to the set of pixels of the image I_2 included by the window when centered at the I_2 coordinates (k, l) . We first compute the correlation coefficient map (function of (x, y)) for every W with origin (e.g., the window center) at coordinates (k, l) , named $C_{I_1, W}(k, l, x, y)$:

$$C_{I_1, W}((k, l, x, y) = \sum_n \sum_m I_1(x + n, y + m) \cdot W_{k,l}(n, m). \quad (3)$$

Note that in practise, we do not compute the correlation coefficient map on the whole reference image I_1 , but restrict the (x, y) to a research area around the coordinates of W . Then, from $C(k, l, x, y)$, that is a 4-way tensor, we can compute $(x_{k,l}^*, y_{k,l}^*)$, which correspond to the displacement

vector maximising the correlation score for every (k, l) :

$$(x_{k,l}^*, y_{k,l}^*) = \underset{x, y}{\operatorname{argmax}} C_{I_1, W}((k, l, x, y). \quad (4)$$

From this, we are able to retrieve the displacement field $\Delta(k, l)$:

$$\Delta(k, l) = (k - x_{k,l}^*, l - y_{k,l}^*). \quad (5)$$

In the state-of-the-art sub-pixel spatial correlation approach of MicMac ([29]), the displacement estimation is based on normalized cross correlation, in which the displacement precision is refined iteratively using a progressively smaller search space with each iteration. Additional refinements, such as the use of a non-linear cost function, and iterative spatial regularization (isotropic and non-isotropic) help to further reduce the impact of high frequency noise in the final displacement map. The spatial regularization takes into account the *a priori* knowledge of the surface regularity, by introducing parameters in the cross-correlation formula able to penalize certain aspects of the computation, such as high-frequency components or unrealistically large shifts [56]. MicMac can be less sensitive to image noise, and to large spatial heterogeneities due to large time differences between the acquisitions [29] than frequency methods. However, spatial correlation can be computationally expensive, as it involves a convolution process involving many computations. The correlation can also fail, or be heavily biased under particular noise conditions (e.g changes in atmospheric conditions, sensor artifacts, radiometric differences, geometric distortions, stereoscopic artifacts, aliasing, color saturation, etc) or very close to the fault, by the nature of the assumptions made.

2) *Frequency-based correlation methods*: Another way to retrieve displacement between two images is to work in the frequency domain. The general principle of frequency-based correlation is to compute the normalized cross-power spectrum Q_{I_1, I_2} (i.e. the complex conjugate of the Fourier transform $\mathcal{F}\{\cdot\}$ of an image multiplied element-wise by the Fourier

transform of a second image), and retrieve its inverse Fourier transform $\mathcal{F}^{-1}\{\cdot\}$, that will give us the normalized cross-correlation matrix $R_{I_1, I_2}(x, y)$ for a specific template image I_2 compared with I_1 :

$$R_{I_1, I_2}(x, y) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{I_1\} \mathcal{F}^*\{I_2\}}{|\mathcal{F}\{I_1\} \mathcal{F}\{I_2\}|} \right\}. \quad (6)$$

The relative displacement between the two images is then determined spatially from the position of the peak in this correlation matrix [26]. To then estimate a displacement field over the full image space, we split I_1 and I_2 into a set of template windows W_1 and W_2 of coordinates (k, l) which are used to obtain the local displacement values $\Delta(k, l)$ from the local correlation coefficient maps $R_{W_1, W_2}(k, l, x, y)$.

$$R_{W_1, W_2}(k, l, x, y) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}\{W_{1k,l}\} \mathcal{F}^*\{W_{2k,l}\}}{|\mathcal{F}\{W_{1k,l}\} \mathcal{F}\{W_{2k,l}\}|} \right\}. \quad (7)$$

Alternatively, for every set of windows, $\Delta(k, l)$ may be estimated directly in the frequency domain [1], [3], [32] from the 2-D slope of the normalized cross-spectrum's unwrapped phase.

In the state-of-the-art frequency-based correlator, COSI-Corr ([3]), the displacement is estimated purely in the frequency domain (thus avoiding expensive IFFT operations, where FFT denotes the Fast Fourier Transform, an algorithm used for converting a time-domain signal into its frequency-domain representation, and IFFT the corresponding inverse operation). Iterative adaptive frequency masking helps to mitigate the impact of noisy (high) frequencies, while maintaining a reasonable computational cost [2]. As with all frequency-based methods, the technique gives increasingly robust results with larger correlation windows [3], [29], albeit at the expense of spatial detail. Therefore, correlation windows of ~ 32 -by- 32 (equivalent to an effective width of 16 -by- 16 , accounting for the windowing function applied to reduce spectral leakage when computing the FFT) often represent the optimum trade-off between accuracy, acceptable noise levels, and spatial detail. Nevertheless, the assumption of a homogenous translation over this correlation window results in bias around sharp discontinuities or regions where the true displacement is more complex.

III. METHODOLOGY

A. Approach and pipeline of the proposed framework

Our proposed approach relies on the same principle as state-of-the-art OIC approaches. First, we work at the local scale with two small windows, W_1 and W_2 , of size $k \times k$ (with k the size of the sliding window in pixels; we use $k = 16$, as it has been demonstrated to be an effective window size from previous OIC studies of earthquake deformation, and allows us to compare with state-of-the-art approaches; k could be adapted). We also make the assumption of a locally rigid and non-rotating transformation, by evaluating the translational displacements Δ between the two windows. We develop an integrated 2-step pipeline able to estimate (1) the large scale displacements (> 1 pixel) with a coarse model in a pixel-scale

registration step, followed by (2) sub-pixel refinement centered on the pixel-scale displacement with a sub-pixel model. Both steps are implemented using a sliding window procedure. Fig. 1 summarizes the full pipeline, taking two satellite images (global scale) as input.

B. Sub-pixel model architecture

To perform the sub-pixel estimation, we developed a 4 convolutional layer CNN architecture `cnn4l` that takes two 16×16 pixel patches, W_1 and W_2 , as input. The model outputs a vector of two displacement values, representing the estimated shift (in pixels) between the two input windows, in the row and column directions (corresponding to north-south and east-west directions for UTM-projected imagery). The architecture of our network can be summarized as follows: the input is passed through four convolutional layers, with an increasing number of small kernels (64, 128, 256 and 256). The size of the kernels (3×3) was selected in order to extract small features in already small (16×16) windows, and is well suited to work on small displacements [48]. The output of each convolution is processed by the Rectified Linear Unit (ReLU) activation function. After the convolutions, two fully connected layers reduce the size of the data (from 16384 to 64 to 2) and outputs the estimated shift. The structure and its parameters are summarized on Fig. 2. Pooling operations are not used here; as the input images are already very small, the size reduction caused by the convolution operations without padding is sufficient to extract multi-scale features (the last feature maps shape is 8×8 , thus, two times smaller than the input). The proposed architecture is generic, could be used with other window sizes than $k = 16$.

C. Generation of the training database

In the Earth Science community, no suitable archive of synthetic earthquake displacements currently exists for the purposes of training a Convolutional Neural Network. Furthermore, the lack of precise, spatially dense ground displacement measurements in real earthquake cases hinders the creation of a relevant database of suitable ground truth data. Consequently, we generate our own synthetic satellite imagery with known displacements, to train and validate our network. This training dataset contains pairs of patches (input windows), together with the rigid deformation values linking them (targets). In order to create samples containing realistic reflectance noise (resulting from variations in illumination, changes in vegetation, etc), we use real Landsat-8 acquisitions acquired over the same region, but on different dates. We select satellite images from areas which are stable between the two acquisitions (i.e. no ground displacement is present). Additionally, we globally co-register the two acquisitions using phase correlation (a Python implementation of the matrix-multiply DFT method of [26] available in the the Scikit-image library), to reduce global mis-registrations that might remain after processing by the USGS.

We extract two windows, W_1 and W_2 , respectively, from two large satellite images, I_1 and I_2 , which are acquired over the same area on two different dates, t_1 and t_2 . We build a

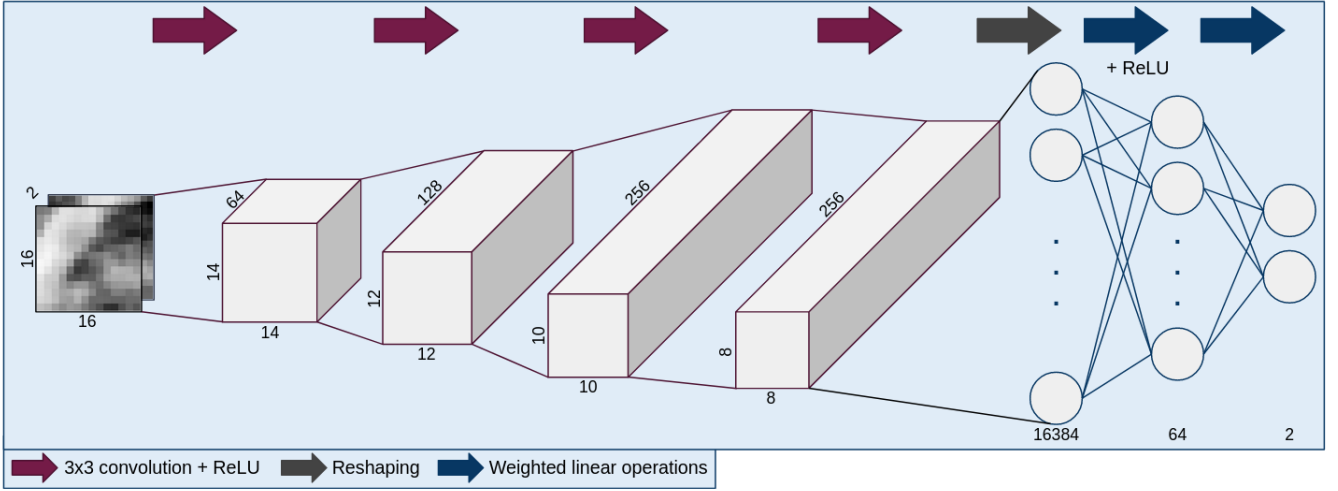


Fig. 2: Architecture of the model used to achieve the sub-pixel estimation (cnn41). The model takes the pre- and the post-windows as input, and outputs the estimated shift (EW and NS displacements between the two windows).

synthetic displacement field, Δ , used to warp W_2 to obtain W_{2s} , the distorted version of W_2 . One sample of the training dataset is the standardized pair (W_1, W_{2s}) (re-scaled with a 0-mean and a unit variance), with the target deformation Δ associated. A re-sampling algorithm is necessary during the warping process, because the shift applied is sub-pixel: we use Lanczos interpolation, with a kernel size of 6×6 .

This can be summarized as follows:

$$W_{2s}(x, y) = e_{\delta t}(f_{\Delta}(W_1(x, y)))(x, y)$$

with e the natural evolution of the ground during δt , and f the distortion operation associated to Δ . This process is repeated many times (with random Δ , random window extraction location, random pair (I_1, I_2) , following a uniform distribution), to create a large number of unique samples. The procedure is summarized on Fig. 3.

1) *UNI Dataset*: The first database that was generated assumes a perfectly uniform shift between the two patches, similar to the traditional correlation implementations. This uniform training dataset is called 'UNI'. Let's fix:

$$\Delta = (\Delta_x, \Delta_y)$$

where $(\Delta_x, \Delta_y) \in [-1, 1]^2$. Here, Δ simulates a uniform sub-pixel shift in W_{2s} . The UNI Dataset is created with this uniform distribution of displacement fields, and is made of 125,000 samples for training (80% for training; 20% for validation) and 25,000 for testing.

2) *DIS Dataset*: A second training database was created in order to address the incorrect assumption of a uniform shift in cases where a displacement discontinuity is present within the sliding window. Here, Δ used to warp W_2 is not uniform, but contains a discontinuity (see Fig. 3). This discontinuity dataset is called 'DIS'. Formally,

$$\Delta(x, y) = \begin{cases} \Delta_a(x, y) = (\Delta_{x_a}, \Delta_{y_a}) & \text{if } (x, y) \in A \\ \Delta_b(x, y) = (\Delta_{x_b}, \Delta_{y_b}) & \text{if } (x, y) \in B \end{cases} \quad (8)$$

where A and B are the two areas (green and pink on Fig. 3) created by intersecting a random line with the window square, with $area(A) > 1.05 \times area(B)$, and (x, y) are the coordinates of the pixels in W_2 . This means that the center pixel is always contained in A. Therefore, the target deformation for this sample is Δ_a . With this simulated fault discontinuity, the model needs to identify the displacement of the largest area A in a given pair of windows, while some smaller area B is allowed to move in a different direction. Again, 125,000 samples are created for training (80% for training; 20% for validation) and 25,000 for testing.

We separately trained the cnn41 model with both UNI and DIS datasets, respectively giving two models: cnn41-uni16 and cnn41-dis16, both made to do the sub-pixel estimation. Two other training datasets were tested: (1) using $e_{\delta t}(W_1) = Id(W_1)$, i.e. $t = 0$ using exactly the same acquisition, and (2) using $e_{\delta t}(W_1) = Id(W_1) + n$ with n a uniform random noise to train the model. However, the accuracy reached on real data was unsatisfactory.

D. Sensitivity study

In order to select the best hyperparameters, we evaluated our models on the test set (25,000 Landsat-8 16×16 patches, acquired on different dates and locations, to guarantee that there is no common data used in both training and evaluation stages).

1) *The sub-pixel model architecture*: We evaluated the sensitivity of our model with respect to the two main parameters controlling the architecture: the number of convolutional layers, and the number of filters in each convolutional layer. We evaluated the mean absolute error when using 1, 2, 3, and 4 convolutional layers, and with three different levels of filters per layer (small, medium and large). Table I gives details on the architectures tested, and quantitative precision on the evaluations. We trained the different architectures with the UNI training set, and tested on the UNI test set four times

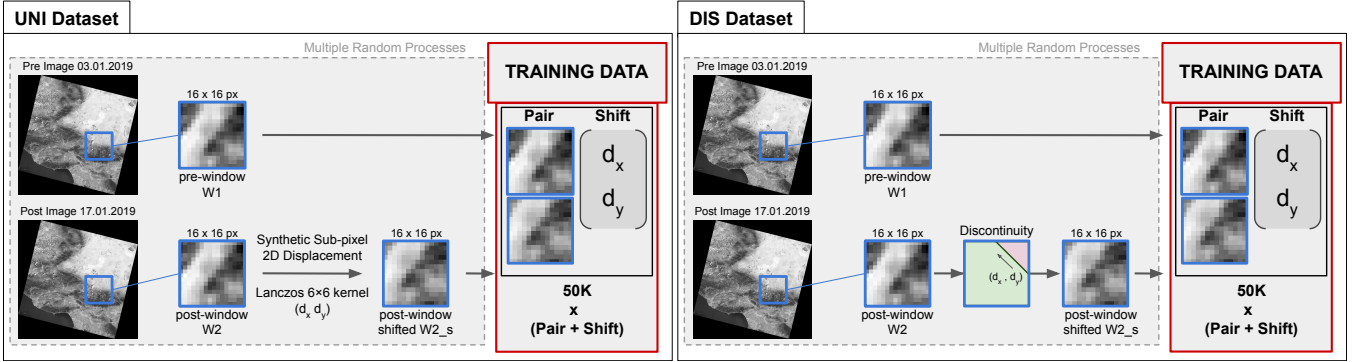


Fig. 3: Creation of our two training sets, UNI (left) and DIS (right). For the UNI dataset, the displacement incorporated in W_2 is uniform, when for the DIS dataset, the deformation includes a discontinuity.

TABLE I: Architectures details (number of layers and number of feature maps for each layer) and Mean absolute error (MAE) of the UNI test set (in px, mean over four runs) for the different models using 16×16 input patches (notation: s: small; m: medium; l: large).

Model name:	cnn-1l			cnn-2l			cnn-3l			cnn-4l		
	s	m	l	s	m	l	s	m	l	s	m	l
<i>conv1 (filters)</i>	16	48	128	16	24	64	16	24	64	16	24	64
<i>conv2 (filters)</i>	x	x	x	20	48	128	20	48	128	20	48	128
<i>conv3 (filters)</i>	x	x	x	x	x	x	24	72	256	24	72	96
<i>conv4 (filters)</i>	x	x	x	x	x	x	x	x	x	28	96	256
MAE (px)	0.127	0.137	0.133	0.107	0.106	0.106	0.110	0.110	0.106	0.114	0.111	0.105

for each architecture, so every evaluation is a mean of four runs.

We see that the main precision gain is obtained with an depth of at least 2 convolutional layers. The best precision is reached with the most complex architecture: 4 layers with the largest number of filters. With this architecture, the model is complex enough to be able to use all the information of the training samples. This architecture was developed using the UNI dataset, although the same trends are observed for the DIS dataset. We thus retain this architecture for the rest of our study. However, future studies may explore broader architectures.

2) *The window size:* The largest are the images, the more information the model has to estimate the displacement. Therefore, on our synthetics (that have a uniform displacement), the best precision is achieved for the largest window size (Table II). However, for real cases where the displacement is heterogeneous within an image, using large windows will bias the estimation. It is thus preferable to use a window size as small as possible. All trained models are performing well (Table II), but for size 8 pixels, the accuracy drops by 30% compared to size 16 pixels. We will focus the rest of our study on the 16×16 window size, known to be more accurate than larger resolutions on real (i.e. not uniform) deformations.

E. Pixel-scale registration with a coarse registration model

The above trained cnn4l model captures shifts that are smaller than 1 pixel in each direction. Even though Earth deformations are contained in this range for the majority of remote images acquired, larger shifts can happen. Yet,

TABLE II: Mean absolute error (MAE) on the UNI test sets (in px) for the different CNN models taking as input different window sizes.

	cnn4l-uni8	cnn4l-uni16	cnn4l-uni32	cnn4l-uni64
MAE (px)	0,130	0,105	0,0960	0,0884

estimating a shift at a pixel level is a different (and easier) task, and it requires a larger patch window. We thus trained another cnn3l model, called cnn3l-5px, with another dataset UNI-5px, very similar to UNI, but with $(\Delta_x, \Delta_y) \in [-5, 5]^2$ and $k = 32$. With this model, we register pixel-wise the sliding windows as a pixel-scale registration step, and then apply our sub-pixel model to refine our estimation. While the results of cnn3l-5px are more than satisfactory for estimating the rounded pixel-level shift (MAE=0.37px, notably smaller than 0.5 pixel), there is a clear advantage to perform the refinement with the specific sub-pixel cnn4l model. This procedure is detailed on Fig. 1, and allows an estimation of large and small displacements.

F. Implementation details

Our cnn4l-uni16 and cnn4l-dis16 models were implemented using PyTorch library in Python. We trained cnn4l-dis16 on 200 epochs and cnn4l-uni16 on 99 epochs (with early stopping), with an initial learning rate of 1×10^{-3} , decaying every 10 epochs with a factor of 0.8. We used the Adam optimizer, and the Mean Squared Error (MSE) loss. We chose a batch size of 128, averaging 10 minutes of computing per epoch

for both models. Computations were performed on a GPU NVIDIA Tesla V100 NVLink. The pipeline is implemented in Python (a combination of Numpy and PyTorch).

IV. RESULTS

A. Evaluation on realistic synthetic earthquake images

In the absence of substantial ground truth data from real earthquakes, we first generate realistic synthetic earthquake images using Landsat-8 satellite acquisitions re-sampled to include realistic synthetic displacement fields; we then quantitatively validate our model and compare with existing state-of-the-art correlators using these synthetic satellite images.

1) *Generation of realistic synthetic earthquake images:* We first develop a pipeline that randomly generates realistic fault discontinuities with rough geometries and slip distributions embedded in a homogeneous elastic half-space (assumed to behave in a linear elastic fashion). Using analytical expressions, relating slip on triangular fault patches (in the elastic material) to displacement at the surface, we then compute the resulting surface displacement field produced by the prescribed fault geometry and associated earthquake slip distribution [57], [58]. The faults obey a length-displacement scaling consistent with natural earthquake faults [59], rupturing only the seismogenic part of the crust, with geometric roughness consistent with natural faults (self-affine scaling, with a Hurst exponent of ~ 0.8 , [60]), fractal slip distributions [61], [62], and the potential for reduced slip in the uppermost crust (shallow slip deficit, [63]). We only consider strike-slip faults (fault dip varies $\pm 60^\circ$ from vertical), in which the displacement is mainly contained within the horizontal plane, as vertical displacements are highly challenging to resolve using orthorectified nadir-view satellite images. The fault models are discretized with an unstructured meshing approach (Mesh2D, [64]) using triangular displacement elements (TDEs), with sub-pixel resolution at the surface, which increases to several kilometers at depth; this depth-dependent spatial sampling accounts for the rapid decay of static displacements with distance to the source ($1/r^2$), which prevents fine spatial details at depth to be resolved by surface measurements alone (i.e. through inverse modelling - see [65]), while also significantly reducing the number of TDEs when computing the surface displacement field (generated using *cutde*, a GPU-accelerated implementation of [58]). Nevertheless, the high resolution of the fault model near the surface allows us to produce realistic displacements at the resolution of the satellite imagery. Finally, we also generate observation points (locations where we calculate a surface displacement) using an unstructured mesh, densifying points with proximity to the surface rupture, thereby concentrating the number of observation points to where there is significant variation in the displacement, further reducing redundant calculations with *cutde*; the irregular grid of output displacements are subsequently resampled onto a regular grid using the *griddata* function within the Python library: SciPy.

We then extract 1024×1024 patches from Landsat-8 satellite images acquired on two different dates, but for a stable region in which no ground deformation has occurred. We

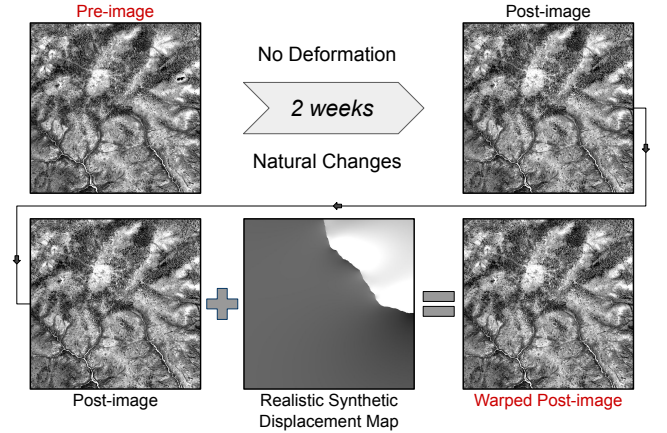


Fig. 4: Creation of a pair of realistic synthetic earthquake images. Here, the warped post-image (bottom right) contains natural differences in reflectance due to the different acquisition time of pre-image (top left), and carries the synthetic displacement map (bottom center), that our model should retrieve.

perform an initial global co-registration step (using phase correlation), to ensure there is no mis-registration between the two images. The displacement field Δ is then used to warp the second (post) satellite image using a quintic-order spline re-sampling algorithm [66], to simulate an earthquake where the displacement field is fully known. The precision of this resampling approach ($\sim 1/100$ th to $1/100$ th pixel) is significantly higher than the precision of state-of-the-art correlators ($< 1/10$ th pixel). Three different 1024×1024 images were created for evaluation; the creation of the input images is summarized on Fig. 4.

2) *Comparisons with MicMac and COSI-Corr:* The pre- and post-images are then compared using our new deep learning approach, COSI-Corr, and MicMac, to retrieve the displacement map. We then compute the absolute residual maps (using the ground truth displacement). In these examples, all displacements are kept in the range $[-1, 1]$, so we just focus on sub-pixel performance.

Using a sliding $k \times k$ pixels window with a stride of 1 pixel, we compute the full displacement maps for the three evaluation examples using our two sub-pixel models, *cnn4l-uni16* and *cnn4l-dis16*, as well as MicMac and COSI-Corr. For consistency, COSI-Corr was applied with a 32×32 window, because the effective correlation window width is reduced \sim half due to the windowing function applied [4]. MicMac typically uses smaller window sizes (the default is 9×9). However, we increase this to 15×15 to allow a more meaningful comparison with COSI-Corr and our pipeline. (Although COSI-Corr can use 16×16 windows, thus yielding an effective window of 8×8 , the displacements start to be biased too low). We also remove the spatial regularization option in MicMac, to keep consistency with our model and COSI-Corr (however, noise is still mitigated in MicMac by using a non-linear cost function, and in COSI-Corr by using adaptive frequency masking). Thus, we focus on the core sub-

TABLE III: Mean residuals (mean absolute error) of cnn4l-uni16, cnn4l-dis16, Mic-Mac and COSI-Corr on the three synthetic simulated 1024×1024 maps.

	Mean residual (px)	Mean residual around fault (px)
cnn4l-uni16	0.0919	0,208
cnn4l-dis16	0.0909	0.150
MicMac	0.0716	0.231
COSI-Corr	0.0689	0.211

pixel performance of the various algorithms.

On Fig. 5, the four EW displacement maps are presented for one of the three realistic earthquake examples (NS displacement maps in Appendix A). To first order, the four models have similar results, with some small differences in texture, and in the sharpness of the fault boundary. Stacked displacement profiles spanning the fault (Fig. 6) highlight a much sharper discontinuity (i.e. narrower displacement width) for our cnn4l-dis16 model, which compares most accurately to the ground truth of all the models; even compared to the default MicMac parameters (9x9 with 0.3 regularization) or MicMac using the smallest available window size (3x3 with 0.3 regularization; note that such parameters yield noisier correlation maps than we obtain with our CNN approach, e.g. see Appendix B), see Fig. 6b. This sharper, more accurate representation of the fault trace is also clear from the 2-D residual maps (Fig. 7). The displacement maps for the two other synthetic earthquake images are available in the Github repository.

The mean residuals (MAE) for each of the 4 methods over the full extent (1024×1024) of the displacement map, as well as a narrow zone extending 16 pixels from the fault trace are given in Table III. The global mean errors of the four models are similar, ranging from $6.89e - 2$ to $9.19e - 2$ pixel. Our models perform slightly worse than MicMac and COSI-Corr (MAE increases by up to $2.30e - 2$ pixel). This can be seen visually in the displacement maps (Figure 5): our cnn4l-uni16 and cnn4l-dis16 models yield displacement maps with slightly more high frequency noise. This is in part because we have made no effort to mitigate high frequency noise, unlike MicMac and COSI-Corr (simple post-processing such as Total Variation - L1 smoothing (weight 0.2) (see [67]) allows our results to attain comparable level of smoothness with MicMac, while preserving sharper fault discontinuity (see Figure B)). Nevertheless, the stacked displacement maps spanning the fault displacement yield very similar displacements (Figure 6).

The error in the near-field of the fault trace is higher than the global residuals in each case. However, for cnn4l-dis16, the residual is almost one third that of the other methods (MAE drops by $\sim 5.91e - 2$), reflecting a substantial improvement in resolving the displacement close to the fault rupture.

B. Evaluation on real case: Ridgecrest

On 4th and 7th July 2019, a large foreshock-mainshock earthquake sequence struck the town of Ridgecrest, in the

Mojave desert region of southern California [52], [68], [69]. The foreshock broke a NE-SW-striking left-lateral fault, while the subsequent mainshock broke a conjugate NW-SE-striking right-lateral fault. Using very high resolution satellite imagery spanning both earthquakes, various studies used OIC to characterize the near-field displacements close to the surface rupture, and assess the extent of off-fault deformation associated with the event [13], [70], [69], [21]. These displacement maps also allowed an assessment of the local strain field, which turn helps to provide new constraints on the mechanics of fault slip in surface rupturing earthquake, and the distribution of slip between the fault core and neighboring damage zones.

Comparisons of our model with COSI-Corr and MicMac for optical data spanning the Ridgecrest earthquake is valuable two reasons: (1) to test our model on satellite images containing real earthquake displacements, and (2) to test our model on different imagery to that used to train our model. In this case, our models are trained with Landsat-8 imagery (15 m), while we test on pre- and post-event Pleiades satellite images (from [21]; resampled here from 0.5 m to 1.5 m for memory constraints); thus, we assess the performance of our model on data with acquired by a different sensor (with different characteristics, e.g. sensitive to different spatial wavelengths), different incidence angles (Pleiades imagery is typically acquired off-nadir, unlike Landsat-8), and at highly different resolutions (spanning ~ 1 order of magnitude). Furthermore, we avoid any artifacts potentially introduced by warping the satellite imagery, and we are not constrained by an artificial synthetic earthquake displacement model (however realistic we attempt to make it). Finally, the displacements involved during the earthquake can reach a magnitude of ~ 6 meters, while the input imagery is 1.5 m resolution; therefore, we can test our full pipeline, which solves for displacements at the pixel-scale and then the sub-pixel scale (Fig. 1). These comparisons therefore allow us to assess the generalizability of our technique. However, we note that for real cases of earthquakes, like Ridgecrest, we do not have a ground truth map of known displacements. Comparing different techniques is nevertheless useful to check consistency with existing and more established OIC methods. Furthermore, visual inspection of the displacement field close to the fault rupture can often reveal errors or bias which are identifiable from their spatial distribution, and which offer another means for comparison.

We compute displacement maps in 3 different areas using (1) our CNN-framework (step-1: pixel-scale registration with cnn3l-5px using 32×32 windows, followed by step-2: sub-pixel refinement with cnn4l-dis16 using 16×16 pixels windows), (2) COSI-Corr (32×32 windows), and MicMac (15×15 windows). A stride/step of 4 pixels is used for the CNN-pipeline and COSI-Corr at large scales (Fig. 8a,c), and a stride of 1 for finer scales (Fig. 8b,d); MicMac (Fig. 8e,f) can only stride at 1 pixel increments. The displacement maps for the first area are shown in Fig. 8 (two results for additional areas are shown in Appendix C). The pre-image of this example was acquired on 23rd June 2012, and the post-image on 8th September 2019 [21].

All four methods give very similar displacement values globally, with small differences in the high frequency content

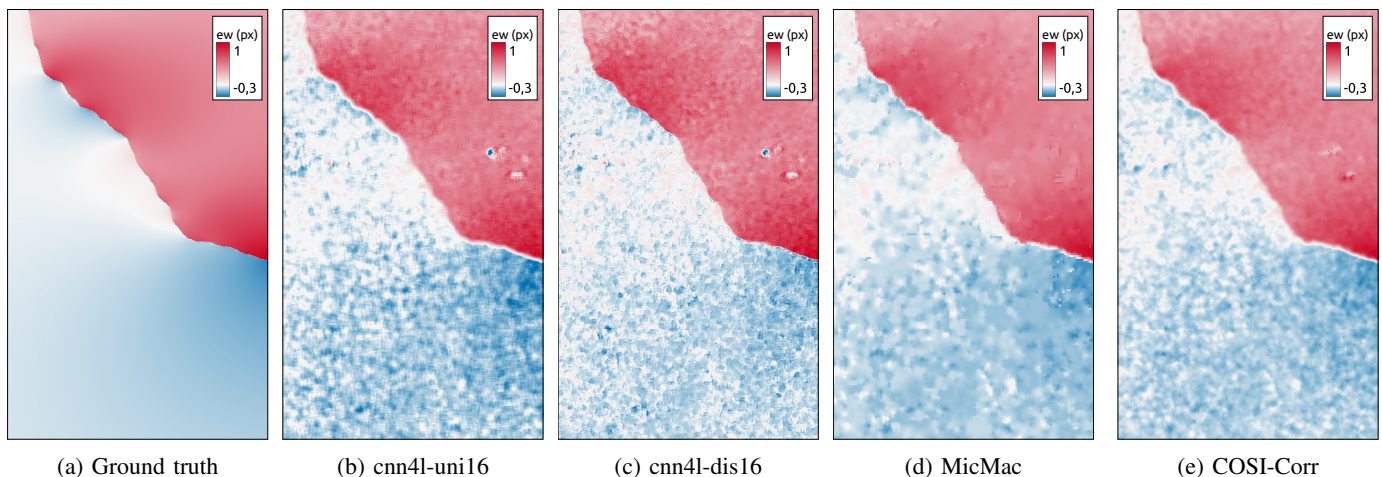


Fig. 5: EW displacement map for the four models `cnn4l-uni16`, `cnn4l-dis16`, `MicMac` and `COSI-Corr` on one of the three examples. On the left, the synthetic displacement maps used to warp the two satellite images. Results are expressed in pixels.

(Fig. 8). Both `MicMac` (and to a lesser extent `COSI-Corr`) give slightly smoother results, and in several small areas smooth over high frequency noise or regions of decorrelation (e.g. from buildings or roads), compared with our CNN pipeline. Nevertheless, globally the displacements are consistent with each other at medium and long wavelengths (Fig. 8).

Close to the fault rupture (Fig. 8b,d,f), where displacements are dominated by high-frequency components (i.e. sharp discontinuities), we see more significant differences between the various methods, consistent with our observations for the synthetic earthquake cases (e.g. Fig. 7 bottom panel). `MicMac` yields a relatively clean, albeit slightly pixelated displacement map (Fig. 8d), which is generally consistent with our CNN-pipeline result (Fig. 8b). `COSI-Corr` yields a slightly smoother, lower resolution displacement map (Fig. 8f), with occasional outliers, and apparent fattening/adhesion artifacts [71] along the rupture trace, indicating an artificially induced component of geometric roughness (which likely results from the frequency masking approach used to mitigate noisy frequencies). Therefore, in real case examples, our CNN-pipeline features significantly less bias in the near-field of the fault rupture.

V. DISCUSSION

In our experiments, we demonstrate that a CNN-based optical registration scheme can achieve similar performance to existing state-of-the-art optical sub-pixel correlators, and in the case of sub-pixel displacements characterized by sharp discontinuities, we exceed it. Reduction of near-field displacement bias in the proximity of sharp discontinuities represents a significant break-through in the characterization of ground displacements using satellite images. Recent applications of OIC in the study of earthquake surface ruptures using high resolution (<1 m) satellite data have tended to focus on (1) characterizing the fault zone width (from the spatial width of the displacement step, e.g. see Fig. 6), (2) examining the near-field strain components (e.g. shear and dilatation), to spatially assess the transition between elastic and inelastic deformation (e.g. a yield strength of 0.5% is often assumed), and (3) to

quantify the total (on- and off-fault) displacement, which can then be compared with highly localized field-based measurements of slip (assumed to represent purely on-fault displacement), in order to quantify the extent of off-fault displacement. All of these applications require an unbiased correlator, able to accurately and precisely resolve the near-field displacement, which has so far remained elusive (though perhaps not well realized by the community). Therefore, potentially all previous studies investigating near-field strain and off-fault deformation based on optical correlation may be biased by to some degree by limitations of the correlation process used to retrieve the displacements. This will be especially apparent when sharp discontinuities become artificially smoothed over a wider zone by the correlation kernel, thereby artificially attributing on-fault slip to the off-fault region. Consequently, estimates of near-field strain will be biased too high, leading to incorrect interpretations about the mechanical behaviour of fault zones. Our approach helps to address this issue, and will lead to more accurate estimates of near-fault displacement in the future; this is the primary value of OIC methods for the study of earthquakes, where the medium and far-field displacements can be more precisely characterized using InSAR techniques).

Additionally, our pipeline, although trained with Landsat-8 satellite imagery, is shown to have general applicability to other optical datasets (e.g. Pleiades). This is likely due to the scale-invariant nature of very small (e.g. 16-by-16 pixel) image windows, which look similar regardless of the sensor used to acquire the data. Nevertheless, our simple pipeline represents a first step towards using data-drive approaches to retrieve sub-pixel displacement maps using optical satellite data.

Several limitations remain in our approach, which will be addressed in future studies. Firstly, to create the training datasets, we used the Lanczos resampling algorithm to incorporate sub-pixel shifts, which can ultimately can add sub-pixel bias resulting from high-frequency ringing artifacts. The magnitude of these artifacts varies (in part) as a function of spatial contrasts between image pixels, and which, in turn, may correlate with topography; the magnitude of this bias

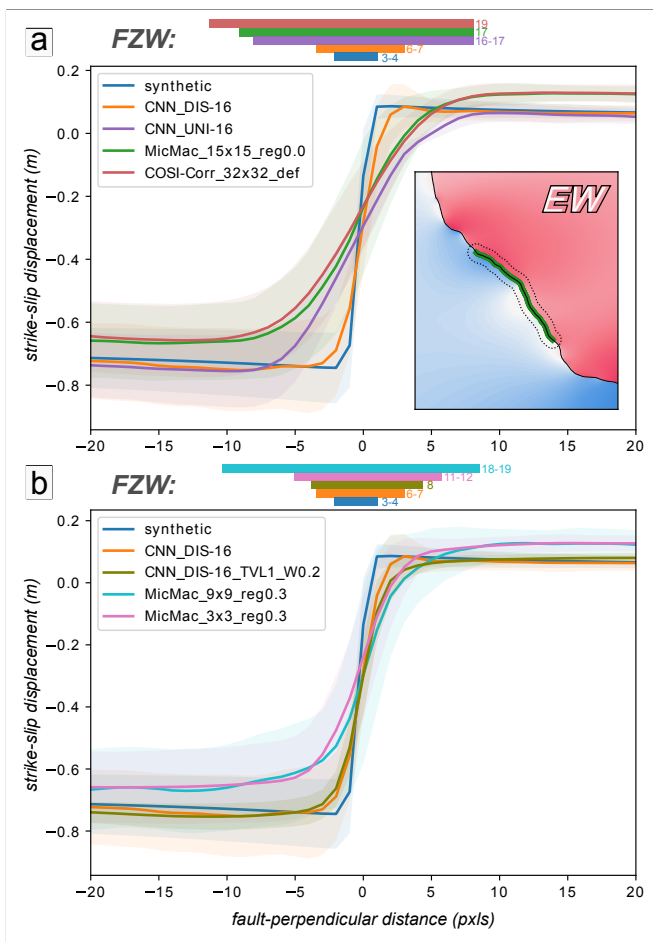


Fig. 6: Stacked displacement profile on a portion of the fault for (a) the synthetic ground truth displacement field, the `cnn4l-dis16` and `cnn4l-uni16` models, `COSI-Corr`, and `MicMac` (15×15 , no regularization), and (b) `MicMac` (9×9 and 3×3 with 0.3 regularization), and `cnn4l-dis16` with TV-L1 smoothing (weighting 0.2; see [67]). The fault zone width (FZW) is estimated in each case (colored bars, with the width in pixels displayed), based on the inflection points of the horizontal displacements each side of the fault.

is typically $< 1/10$ th pixel. When evaluating the error over larger synthetic displacement maps, we are also subject to interpolation bias introduced by our warping procedure, used to create synthetically displaced satellite images; the magnitude of these resampling errors on the resulting displacement map is generally substantially less than $1/50$ th pixel. We are also subject to initial mis-registration of the satellite images used to generate our training data, which can add bias to our synthetic displacements. We try to mitigate this error by applying a global mis-registration correction to the full pre and post images (i.e. not just on the two windows), based on sub-pixel phase correlation, to bring the two images into better alignment; by aligning the two images over a large area, we obtain a more accurate global alignment exceeding $1/10$ th pixel. Nevertheless, this global alignment step may become less accurate if the illumination conditions between the two images are strongly different.

In addition, we also face the same basic problem of solving for an overly simplistic displacement transform. Although we can now solve for a displacement step function contained within a sliding window, we still ignore higher order terms, rotations, scale changes, etc, which may be important in characterizing the displacement (especially close to sharp discontinuities). Furthermore, insufficient calibration of the CCD sensors, and resampling errors introduced during the ortho-rectification process (prominent in topographically rough regions, and when the DEM is of lower resolution than the image) both introduce sub-pixel bias in the training data. Therefore, the sub-pixel precision of our CNN-approach is likely limited to $\sim 1/10$ th pixel, reflecting the combined contribution of all these error sources.

Additional development of our CNN-based registration approach will likely further enhance performance, particularly in mitigating high frequency noise in the final estimated displacement map. Numerous tests were performed in tuning the model parameters; however, further investigation may lead to improved sub-pixel capability. We also limit the amount of training data to 100k samples (based on sensitivity studies for this parameter, striking a balance between accuracy and computation time); nevertheless, increasing the number of samples may also leave to additional improvements. Our sub-pixel CNN-based model is fixed to window sizes of 16×16 pixels, thus restricting the level of spatial detail we can resolve; it could be possible to further reduce the window size to obtain additional spatial detail, while relying on multi-scale spatial regularization techniques to mitigate increased noise associated with smaller sliding windows.

The reduction of noise introduced during the generation of training data will also directly enhance the sub-pixel accuracy and precision which we can obtain (and consequently the high frequency spatial noise in the output displacement maps). A significant component of the high frequency noise is likely related to systematic bias resulting from variations in surface reflectance related to differing illumination conditions between the two images. In the case of widespread and freely available optical datasets, such as Landsat-8 or Sentinel-2, where images are acquired at approximately the same time of day (~ 10 am local time), these sources of noise should vary in a predictable manner, thereby enabling this source of bias to be learned and potentially removed using data-driven approaches [6]. Here we make no effort to apply spatial regularization techniques to smooth high frequency noise; therefore, the fact that our simple data-driven approach already performs on a par with state-of-the-art OIC approaches is encouraging for future development of CNN-based optical registration techniques. One obvious direction for future development will be in the use of multi-scale approaches, such those offered by U-Net architectures, and which have been shown to achieve sub-pixel performance in particular cases [48], [49]. In the case of earthquake, the displacement fields are generally smooth over longer wavelengths, and thus should benefit from multi-scale spatial regularization techniques.

Finally, our CNN-based registration method is developed within a GPU environment, which offers the potential for significant speed improvements in generating displacement maps

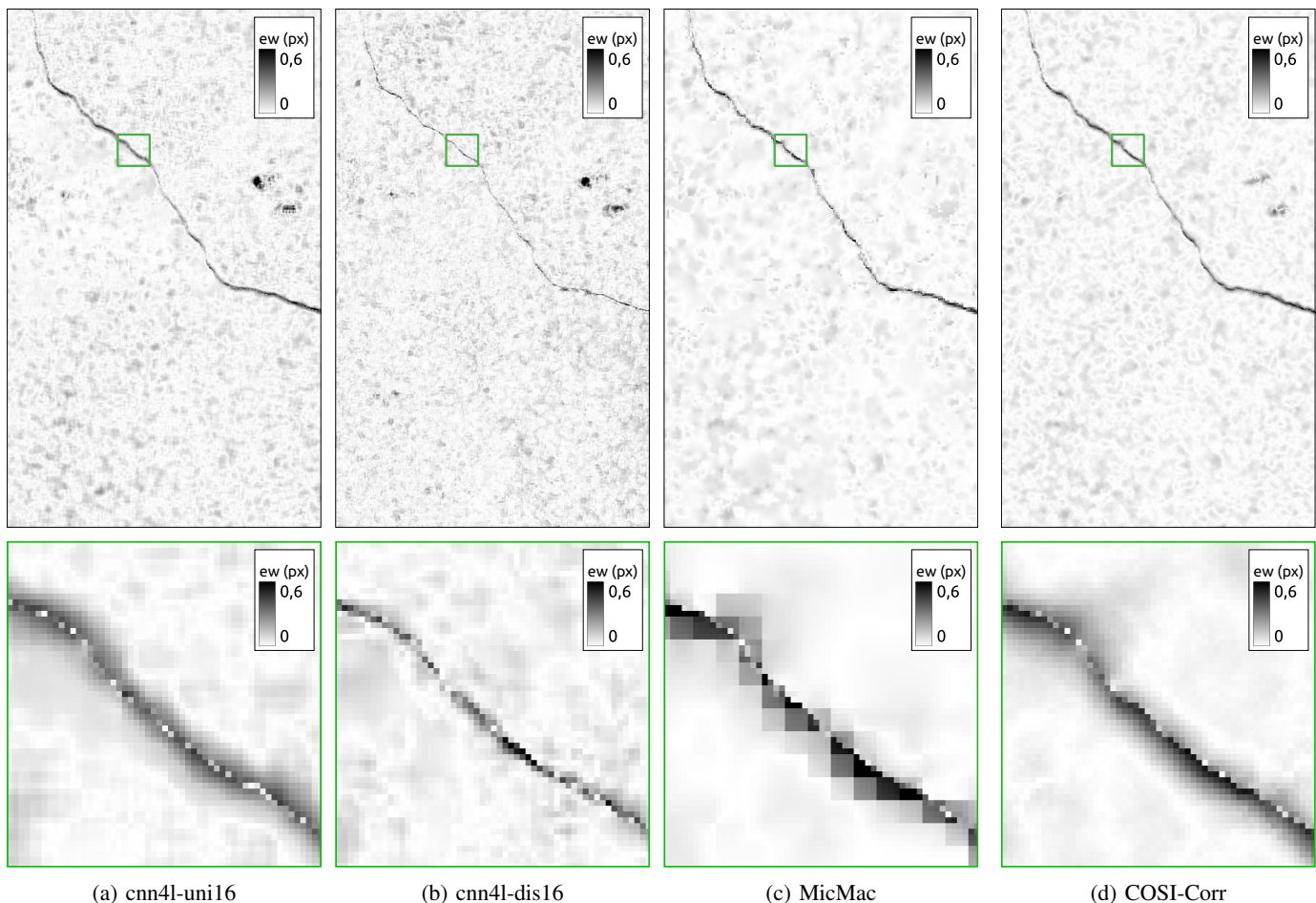


Fig. 7: Residual maps of the displacements from Fig. 5 for the four models (EW component) cnn41-uni16, cnn41-dis16, MicMac and COSI-Corr.

from optical data. Speed optimization is beyond the scope of this current study, which represents a proof of concept for the technique. Nevertheless, we obtain computation speeds within ~ 1 -order of magnitude of COSI-Corr and MicMac, which have both been optimized for speed (e.g. developed in C++, multi-threaded); e.g. for a pre/post image pair (1024×1024 pixels), the computation time is ~ 10 minutes on a GPU NVIDIA Tesla V100 NVLink, compared with 30 seconds for COSI-Corr and 1m30 for MicMac on 4 CPU cores. However, for larger images, where the computational overhead of loading images into memory is small relative to the correlation time, the run-times of COSI-Corr and MicMac become similar (for a stride of 1). It is for these larger image cases where we expect significant improvement in the runtime for our CNN-based method. Our pipeline would drastically increase in speed by processing the windows in batches to reduce data transfer between CPU and GPU, and using parallelism with multiple GPUs. In addition, future efforts should explore multi-scale deep learning architectures (e.g. U-Net), which offer the best potential for speed gain; such approaches work with much larger image windows, thereby reducing the computational burden associated with processing many correlation windows.

VI. CONCLUSION

Our paper presents a complete deep learning framework that extracts ground motion, from pairs of optical satellite images (OIC), with sub-pixel accuracy. By addressing the presence of sharp discontinuities within sliding windows, we propose a new approach that mitigates displacement estimation bias, exceeding the performance of state-of-the-art methods, in the near-field of earthquakes. We demonstrate this improvement quantitatively using evaluations on high quality synthetic data generated using realistic fault slip models. We also demonstrate that our technique is transferable, achieving state-of-the-art performance on satellite images acquired with different sensors and resolutions; which is not trivial for such data-driven approaches. Future work will focus on leveraging multi-scale deep learning approaches to improve accuracy (including reduction of high frequency noise, and systematic bias), and speed.

ACKNOWLEDGMENT

The authors would like to thank IDEX Université Grenoble Alpes, INSU PNTS, CNES, and CDP Risk for funding, and GRICAD infrastructure (gricad.univ-grenoble-alpes.fr), which

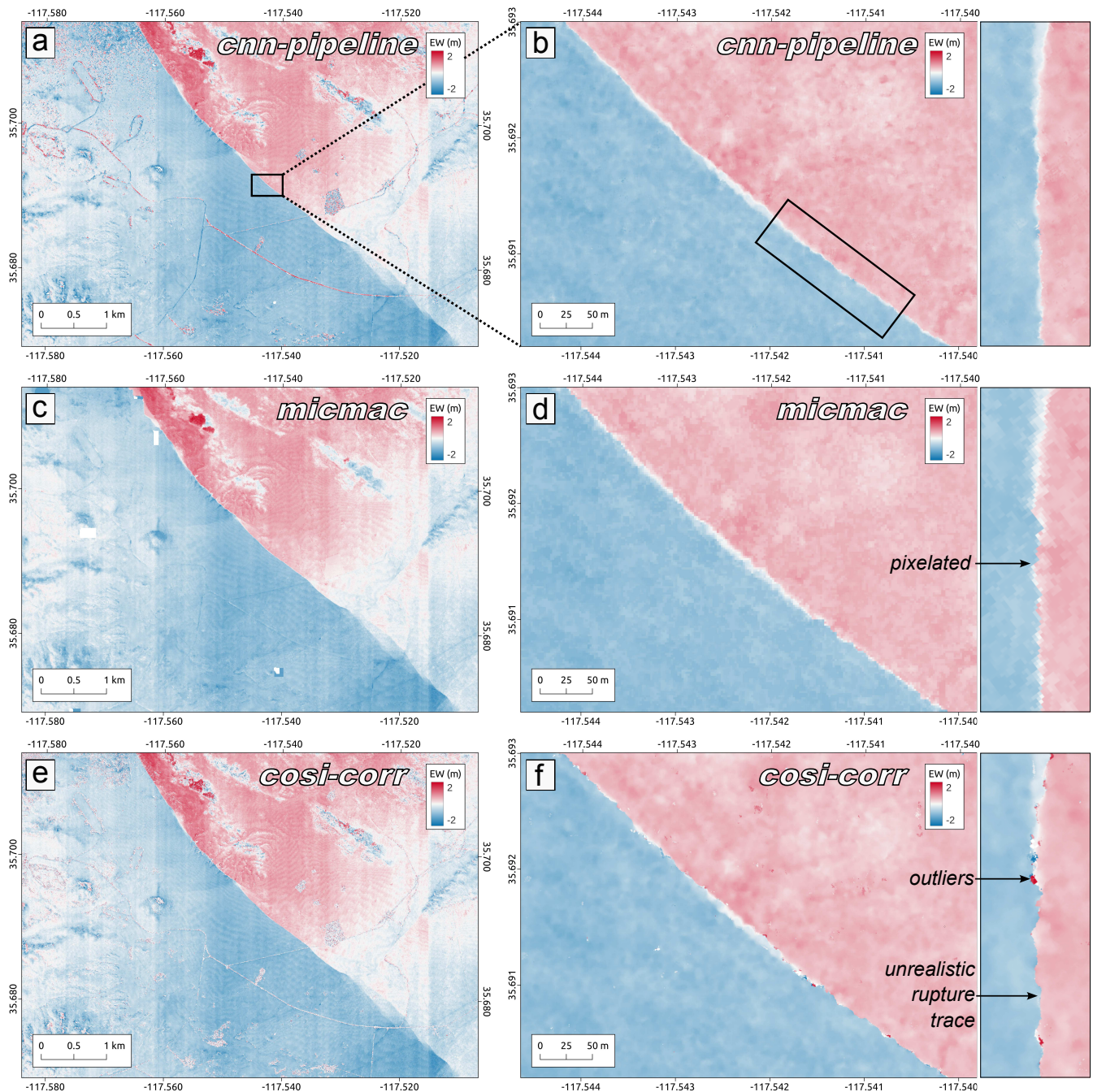


Fig. 8: EW component of the displacement maps from Pleiades images for the Ridgecrest earthquake. (left-column:) (a) Our 2-step CNN-pipeline, (c) MicMac, and (e) COSI-Corr. Positive pixel values indicates movement to the south and negative to the north. (right-column:) Close-up region immediately around the fault rupture for (b) CNN-pipeline, (d) MicMac, and (f) COSI-Corr. Inset shows a further zoom of the rupture details.

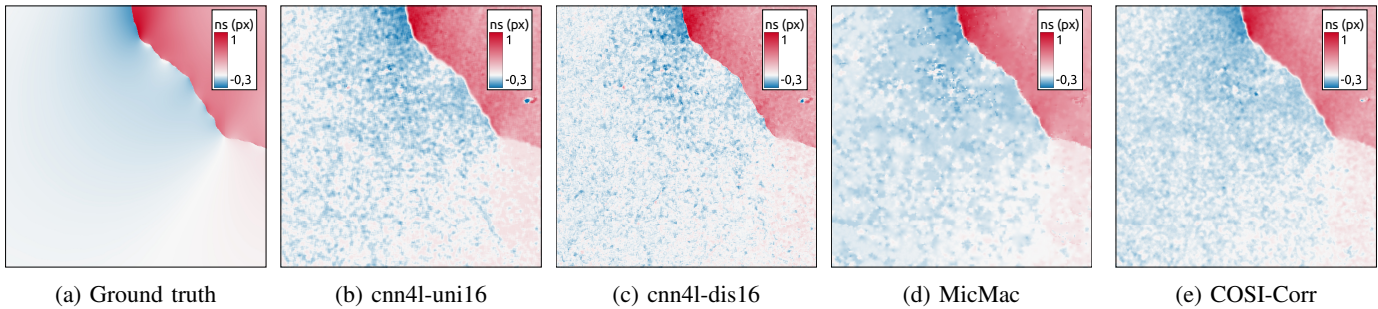


Fig. A: NS displacement map for the four models `cnn4l-uni16`, `cnn4l-dis16`, `MicMac` and `COSI-Corr` on the first example. On the left, the synthetic displacement maps used to warp the two satellite images. Results are expressed in pixels.

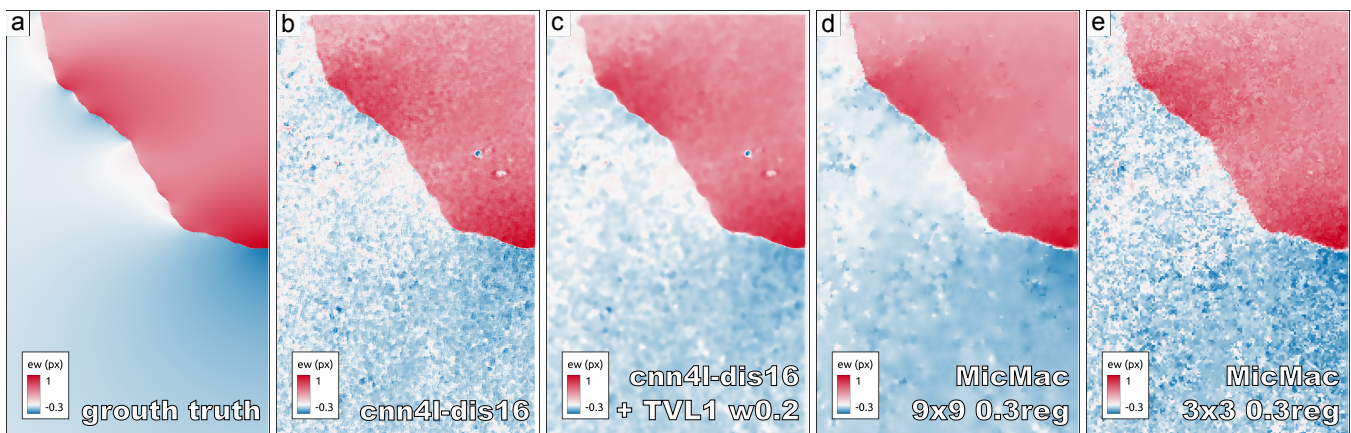


Fig. B: EW displacement maps for the (a) ground truth, (b) `cnn4l-dis16`, (c) `cnn4l-dis16` with TVL1 smoothing (weight 0.2), (d) `MicMac` (9x9 window, 0.3 reg), and (e) `MicMac` (3x3 window, 0.3 reg). Inclusion of simple TVL1 smoothing on our `cnn4l-dis16` model highlights the potential of fast and simple post-processing approaches that may be used to reduce high frequency noise, giving smooth displacement maps comparable to default `MicMac` parameters, yet with reduced near-field bias.

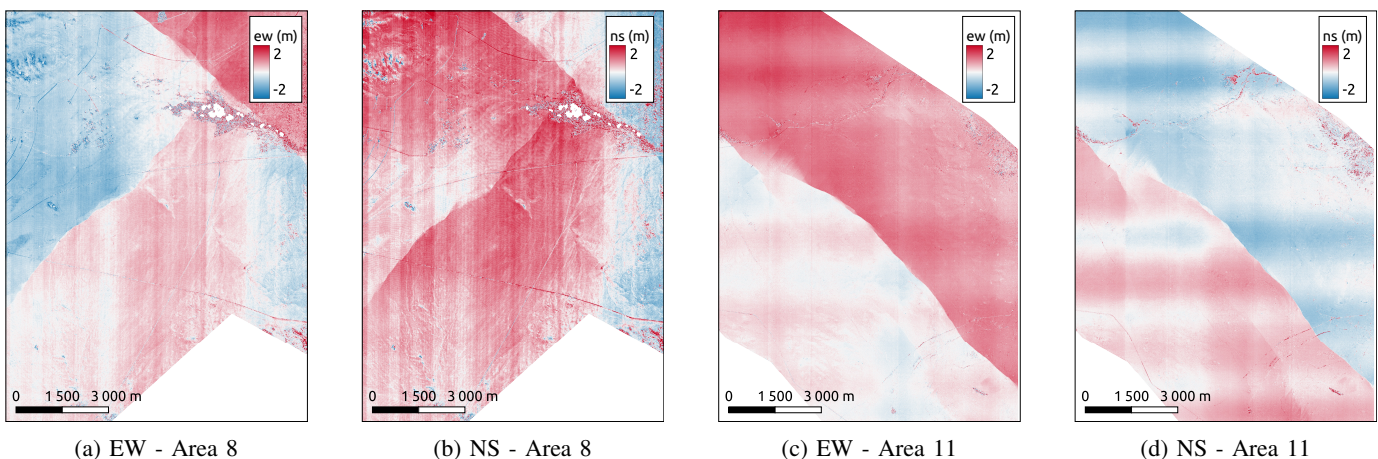


Fig. C: EW and NS components of the displacement maps from real Pleiades images (Ridgecrest area 8 and 11), acquired on 23rd June 2012 (pre-image) and on 26th August 2019 (post-image), computed with our 2-step CNN pipeline. Positive pixel values indicates movement to the south, negative to the north.

is supported by Grenoble research communities, for the computations. We also thank Solène Antoine for kindly providing the Pleiades orthoimages of the Ridgecrest region.

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] 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.
- [4] T. Heid and A. Käab, "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.
- [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] 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.
- [7] 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.
- [8] A. M. R. Padilla and M. E. Oskin, "A probabilistic displacement hazard assessment framework for distributed ruptures from strike-slip earthquakes," 2023.
- [9] J. F. Dolan and B. D. Haravitch, "How well do surface slip measurements track slip at depth in large strike-slip earthquakes? the importance of fault structural maturity in controlling on-fault slip versus off-fault surface deformation," *Earth and Planetary Science Letters*, vol. 388, pp. 38–47, 2014.
- [10] C. Milliner, J. Dolan, J. Hollingsworth, S. Leprince, and F. Ayoub, "Comparison of coseismic near-field and off-fault surface deformation patterns of the 1992 mw 7.3 landers and 1999 mw 7.1 hector mine earthquakes: Implications for controls on the distribution of surface strain," *Geophysical Research Letters*, vol. 43, no. 19, pp. 10–115, 2016.
- [11] G. Cheng and W. D. Barnhart, "Permanent co-seismic deformation of the 2013 mw7.7 baluchistan, pakistan earthquake from high-resolution surface strain analysis," *Journal of Geophysical Research: Solid Earth*, vol. 126, no. 3, p. e2020JB020622, 2021.
- [12] 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.
- [13] 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.
- [14] N. Ajorlou, J. Hollingsworth, Z. Mousavi, A. Ghods, and Z. Masoumi, "Characterizing near-field surface deformation in the 1990 rudbar earthquake (iran) using optical image correlation," *Geochemistry, Geophysics, Geosystems*, vol. 22, no. 6, p. e2021GC009704, 2021.
- [15] J. Avouac and S. Leprince, "Geodetic imaging using optical systems," in *Geodesy*. Elsevier Inc., 2015, pp. 387–424.
- [16] M. Marchandon, T. J. Wright, and J. Hollingsworth, "Remote sensing of the earthquake deformation cycle," *Surface Displacement Measurement from Remote Sensing Images*, pp. 191–246, 2022.
- [17] S. Leprince, E. Berthier, F. Ayoub, C. Delacourt, and J.-P. Avouac, "Monitoring earth surface dynamics with optical imagery," *Eos, Transactions American Geophysical Union*, vol. 89, no. 1, pp. 1–2, 2008.
- [18] R. Michel and J.-P. Avouac, "Deformation due to the 17 august 1999 izmit, turkey, earthquake measured from spot images," *Journal of Geophysical Research: Solid Earth*, vol. 107, no. B4, pp. ETG–2, 2002.
- [19] C. W. Milliner, J. F. Dolan, J. Hollingsworth, S. Leprince, F. Ayoub, and C. G. Sammis, "Quantifying near-field and off-fault deformation patterns of the 1992 mw 7.3 landers earthquake," *Geochemistry, Geophysics, Geosystems*, vol. 16, no. 5, pp. 1577–1598, 2015.

- [20] C. P. Scott, J. R. Arrowsmith, E. Nissen, L. Lajoie, T. Maruyama, and T. Chiba, "The m7 2016 kumamoto, japan, earthquake: 3-d deformation along the fault and within the damage zone constrained from differential lidar topography," *Journal of Geophysical Research: Solid Earth*, vol. 123, no. 7, pp. 6138–6155, 2018.
- [21] 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.
- [22] C. Lasserre, G. Peltzer, F. Crampé, Y. Klinger, J. van Der Woerd, and P. Tapponnier, "Coseismic deformation of the 2001 mw= 7.8 kokoxili earthquake in tibet, measured by synthetic aperture radar interferometry," *Journal of Geophysical Research: Solid Earth*, vol. 110, no. B12, 2005.
- [23] B. Pinel-Puysségur, F. De Zan, and J. Champenois, "The interferometric phase: Unwrapping and closure phase," *Surface Displacement Measurement from Remote Sensing Images*, pp. 155–189, 2022.
- [24] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l 1 optical flow," in *Pattern Recognition: 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29*. Springer, 2007, pp. 214–223.
- [25] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 30, no. 2, pp. 328–341, 2007.
- [26] 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.
- [27] A. V. Nefian, K. Husmann, M. Broxton, V. To, M. Lundy, and M. D. Hancher, "A bayesian formulation for sub-pixel refinement in stereo orbital imagery," in *2009 16th IEEE international conference on image processing (ICIP)*. IEEE, 2009, pp. 2361–2364.
- [28] P. A. Rosen, S. Hensley, G. Peltzer, and M. Simons, "Updated repeat orbit interferometry package released," *Eos, Transactions American Geophysical Union*, vol. 85, no. 5, pp. 47–47, 2004.
- [29] 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.
- [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] C. Milliner and A. Donnellan, "Using daily observations from planet labs satellite imagery to separate the surface deformation between the 4 july m w 6.4 foreshock and 5 july m w 7.1 mainshock during the 2019 ridgecrest earthquake sequence," *Seismological Research Letters*, vol. 91, no. 4, pp. 1986–1997, 2020.
- [32] 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.
- [33] T. A. Scambos, M. J. Dutkiewicz, J. C. Wilson, and R. A. Bindshadler, "Application of image cross-correlation to the measurement of glacier velocity using satellite image data," *Remote sensing of environment*, vol. 42, no. 3, pp. 177–186, 1992.
- [34] X. Tong, Z. Ye, Y. Xu, S. Gao, H. Xie, Q. Du, S. Liu, X. Xu, S. Liu, K. Luan *et al.*, "Image registration with fourier-based image correlation: A comprehensive review of developments and applications," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 10, pp. 4062–4081, 2019.
- [35] G. Facciolo, C. De Franchis, and E. Meinhardt, "Mgm: A significantly more global matching for stereovision," in *BMVC 2015*, 2015.
- [36] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [37] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *IJCAI'81: 7th international joint conference on Artificial intelligence*, vol. 2, 1981, pp. 674–679.
- [38] J. Sánchez Pérez, E. Meinhardt-Llopis, and G. Facciolo, "TV-L1 Optical Flow Estimation," *Image Processing On Line*, vol. 3, pp. 137–150, 2013, <https://doi.org/10.5201/iplol.2013.26>.
- [39] 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.
- [40] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE transactions on pattern analysis and machine intelligence*, vol. 16, no. 9, pp. 920–932, 1994.
- [41] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.
- [43] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration—a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.
- [44] M. Simonovsky, B. Gutiérrez-Becker, D. Mateus, N. Navab, and N. Komodakis, "A deep metric for multimodal registration," in *International conference on medical image computing and computer-assisted intervention*. Springer, 2016, pp. 10–18.
- [45] S. Wang, D. Quan, X. Liang, M. Ning, Y. Guo, and L. Jiao, "A deep learning framework for remote sensing image registration," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 148–164, 2018.
- [46] F. Ye, Y. Su, H. Xiao, X. Zhao, and W. Min, "Remote sensing image registration using convolutional neural network features," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 2, pp. 232–236, 2018.
- [47] 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.
- [48] 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.
- [49] 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.
- [50] R. Yang, Y. Li, D. Zeng, and P. Guo, "Deep dic: Deep learning-based digital image correlation for end-to-end displacement and strain measurement," *Journal of Materials Processing Technology*, vol. 302, p. 117474, 2022.
- [51] T. Montagnon, J. Hollingsworth, E. Pathier, M. Marchandon, M. Dalla Mura, and S. Giffard-Roisin, "Sub-pixel optical satellite image registration for ground deformation using deep learning," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 2716–2720.
- [52] 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.
- [53] M. A. Sutton, J. J. Ortu, and H. Schreier, *Image correlation for shape, motion and deformation measurements: basic concepts, theory and applications*. Springer Science & Business Media, 2009.
- [54] B. Pan, K. Qian, H. Xie, and A. Asundi, "Two-dimensional digital image correlation for in-plane displacement and strain measurement: a review," *Measurement science and technology*, vol. 20, no. 6, p. 062001, 2009.
- [55] J. P. Lewis, "Fast template matching," in *Vision interface*, vol. 95, no. 120123. Quebec City, QC, Canada, 1995, pp. 15–19.
- [56] 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.
- [57] 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.
- [58] M. Nikkhou and T. R. Walter, "Triangular dislocation: an analytical, artefact-free solution," *Geophysical Journal International*, vol. 201, no. 2, pp. 1119–1141, 2015.
- [59] C. H. Scholz, "Scaling laws for large earthquakes: consequences for physical models," *Bulletin of the Seismological Society of America*, vol. 72, no. 1, pp. 1–14, 1982.
- [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. Lerpince, 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] D. Engwirda, "Locally optimal delaunay-refinement and optimisation-based mesh generation," 2014.
- [65] M. T. Page, S. Custódio, R. J. Archuleta, and J. Carlson, "Constraining earthquake source inversions with gps data: 1. resolution-based removal of artifacts," *Journal of Geophysical Research: Solid Earth*, vol. 114, no. B1, 2009.
- [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] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, pp. 89–97, 2004.
- [68] 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.
- [69] 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.
- [70] 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.
- [71] G. Blanchet, A. Buades, B. Coll, J.-M. Morel, and B. Rougé, "Fattening free block matching," *Journal of mathematical imaging and vision*, vol. 41, pp. 109–121, 2011.