ALADDIn: Autoencoder-LSTM based Anomaly Detector of Deformation in InSAR

Anza Shakeel, Richard J Walters, Susanna K Ebmeier and Noura Al Moubayed

***Abstract*—In this study we address the challenging problem of automatic detection of transient deformation of the Earth's crust in time-series of differential satellite radar (InSAR) images. Detection of these events is important for a wide range of natural hazard and solid earth applications and InSAR is an ideal data source for this purpose due to its frequent and global observational coverage. However, the size of this dataset precludes systematic manual analysis and low signal to noise ratio makes this task difficult. We present a novel method to address this problem. This approach requires development of a novel network architecture to take advantage of the unique structure of the InSAR dataset. Our unsupervised deep learning model learns the 'normal' unlabeled spatio-temporal patterns of background noise signals in 3D InSAR datasets and learns the relationship between the input difference images and the underlying unknown set of individual 2D fields of noise from which the InSAR images are constructed. The detection head of our pipeline consists of two complementary methods, semivariogram analysis and density-based clustering. To evaluate, we test and compare three increasingly complex network architectures: compact, deep, and Bi-deep. The analysis demonstrates that the Bi-deep architecture is the most accurate and so it is used in the final detection pipeline (ALADDIn). The analysis of experimental results is based on detection of a synthetic deformation test case, achieving a 91.25% overall performance accuracy. Furthermore, we show that ALADDIn can detect a real earthquake of Magnitude 5.7 that occurred in 2019 in south-west Turkey.**

***Index Terms*—Unsupervised Deep Learning, Anomaly Detection, InSAR, Satellite Radar Data, Earthquake.**

I. Introduction

TWO new radar satellites, the Sentinel-1 constellation, are revolutionizing the way we measure deformation of the Earth's surface, generating high-spatial-resolution, near-global imagery of on-shore crustal deformation on a daily-to-weekly basis [1]. This new dataset of Sentinel-1 InSAR (Interferometric Synthetic Aperture Radar) images affords a major opportunity to investigate the prevalence of transient deformation phenomena that may have remained undetected in previous datasets [2]. InSAR datasets have a unique 3D structure, where individual images (interferograms) are in fact the difference in phase between two individual radar images taken of the same area but at different times. This 3D interferogram (referred as IFG) dataset is very different to the real-world video time-series that is a common dataset for deep-learning analysis, including anomaly detection, where an individual image instead captures information at a particular instant in time.

Detecting and measuring transient episodes of crustal deformation is important for a wide range of solid earth and natural hazard applications, e.g. for improving understanding of seismic and volcanological hazards and for monitoring anthropogenic deformation. It is important to characterize when and where such events have occurred in order to illuminate the basic physics of these deformation processes and to accurately estimate the hazards they pose to human populations. However, such a large dataset of satellite images (10TB/day, 1000-2000 images/day) [3] precludes systematic manual analysis, and the large magnitude of atmospheric and other nuisance signals relative to deformation signals of interest makes this task difficult. Therefore, this important objective requires the development of new automatic-detection tools based on cutting edge machine-learning methods.

Machine learning has been successfully applied to a wide variety of remotely-sensed satellite datasets for scene classification, object detection and mapping purposes [4], [5]. However, the application of machine learning and deep learning algorithms for analysis of InSAR data is still in its infancy. The majority of the early attempts to apply machine-learning to detection and extraction of deformation signals in InSAR datasets have involved relatively inflexible, off-the-shelf and supervised solutions, for example AlexNet [6] was used for supervised classification of volcanic signals in 2D images [7], a VGG [8] network was employed to detect volcanic unrest in 1D time-series [9], a supervised FCN [10] was designed based on UNet [11] to separate volcanic signals from time-consecutive InSAR-derived 2D time-series [12] and a supervised autoencoder [13] was trained to reconstruct accumulated ground deformation.

These existing approaches all have one or more of several major limitations: I) they are limited to analysis in space (2D) or time (1D) only, or else use higher-level InSAR-derived products that involve filtering or modelling constraints that make a priori assumptions about the signal; II) they require resource-intensive pixel-wise labelling on a limited dataset of real-world examples; III) they are restricted to focus on a single type of deformation only (e.g. volcanoes); IV) they preclude the important ability to detect deformation signals with previously unobserved spatial or temporal structure. To overcome all these issues we take full advantage of the unique, differential and multi-linked 3D nature of InSAR datasets. We have developed a new, unsupervised, event-agnostic, and state-of-the-art deep-learning based approach for automatic

A. Shakeel is with the Department of Earth Sciences, Durham University, UK, anza.shakeel@durham.ac.uk

R. J. Walters is with the Department of Earth Sciences, Durham University, UK

S.K. Ebmeier is with the School of Earth and Environment, University of Leeds, UK, s.k.ebmeier@leeds.ac.uk

N. Al Moubayed is with the Department of Computer Sciences, Durham University, UK, noura.al-moubayed@durham.ac.uk

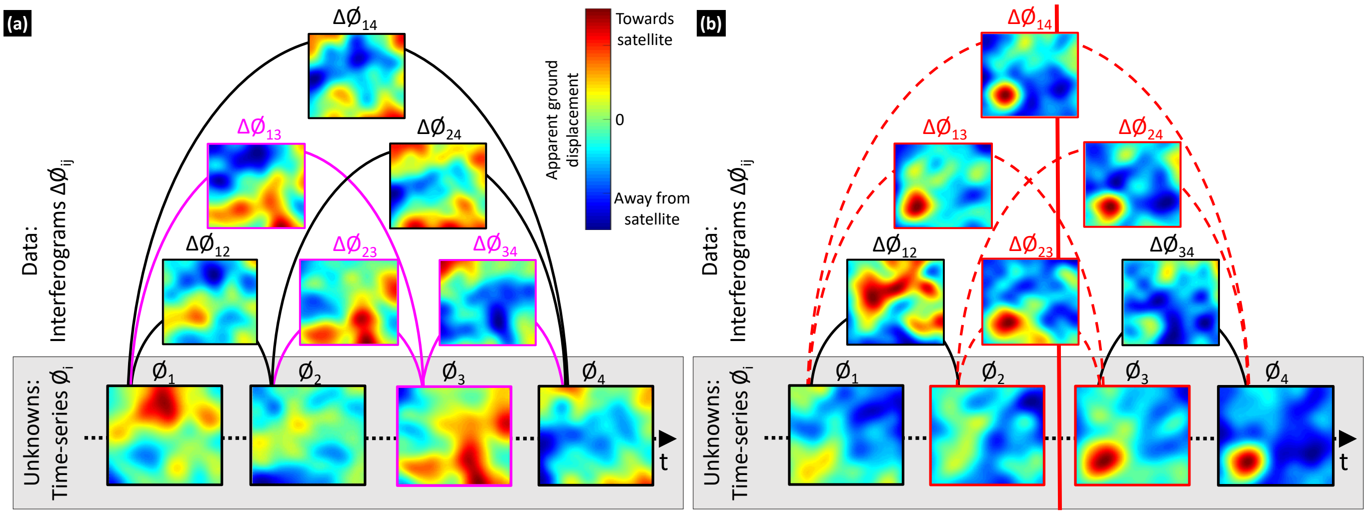


Fig. 1. Cartoon illustrating the unique structure of the InSAR dataset, in particular the relationship between the measured inteferograms (IFGs, phase-change $\Delta\emptyset_{ij}$ shown in top part) and the unknown epoch images (EP, $\emptyset_i$ shown in grey box in (a) and (b)). Pink outlined images in (a) show how nuisance signals associated with an individual epoch (e.g. the red signal in the bottom-right of the $\emptyset_3$ image) contribute to all linking interferograms, some in a positive sense (e.g. a similar red signal in $\Delta\emptyset_{13}$, $\Delta\emptyset_{23}$) and some in a negative sense (e.g. a blue signal of similar shape in $\Delta\emptyset_{34}$). Vertical red line in (b) represents a transient episode of deformation taking place between $\emptyset_2$ and $\emptyset_3$. Red outlined images in (b) show how this deformation contributes in a positive sense to any inteferogram that spans this event (e.g. a similar circular structure that is always red in the bottom left of $\Delta\emptyset_{13}$, $\Delta\emptyset_{23}$, $\Delta\emptyset_{14}$ and $\Delta\emptyset_{24}$).

detection of transient deformation.

In this novel approach we adopt an anomaly detection framework, based on convolutional neural networks (CNNs) and neural networks (NNs). Under this framework, anomalies correspond to any transient phenomena that deviates from the 'normal' spatio-temporal patterns in the dataset. Such 'normal' phenomena arise from a combination of atmospheric signals, satellite orbital errors and other unwanted 'nuisance' signals [14] [15]. We exploit the fact that the unknown 2D fields of nuisance non-deformation signals associated with individual SAR acquisition dates (these are termed 'epoch images' or *EP* here, following the domain nomenclature, and are not to be confused with the typical machine-learning definition of epoch) map into signals in interferograms with a fundamentally different temporal pattern to 'anomalous' deformation signals (Figure 1). By training a deep-learning algorithm to map common noise signals in inteferograms into the unknown EP time-series (Fig1a) we are then able to detect rare deformation events that map into the EP time-series differently (Fig1b). Our approach therefore allows us not only to estimate a background time-series of the unknown non-deformation signals, but also to identify deformation and effectively and accurately separate it from this background. In comparison to the past work, the main contributions of this work are:

- We have established a novel network architecture using CNNs and NNs that transforms InSAR data into an *EP* image sequence. It models the spatial and temporal patterns and the connection between interferograms and their corresponding *EP* images.
- Our model is unsupervised and is event-agnostic anomaly detection, where anomalies correspond to any transient phenomena that deviates from the 'normal' spatio-temporal pattern.
- We have successfully trained the framework on a set of interferogram sequences with multiple outputs. First, the automatic prediction of *EP* image responses (that are originally unknown). Second, the reconstruction of interferograms using these predicted *EP* responses, and last but not least the detection of anomalies within the sequence.
- We have developed a novel detection-and-extraction approach, that flags anomalies, estimates their spatial structure and separates them from noise.
- Finally, we present an accurate analysis of a test set with and without synthetic anomaly with spatial extent and amplitude similar to the background noise, achieving a true positive rate of 81.25% and an overall accuracy of 91.25%, and we also successfully demonstrate our method's ability to detect a real earthquake of Magnitude 5.7 that occurred in south east Turkey (a region outside the training set).

In this study, we first provide an introduction about InSAR data (II) and how it can be used with deep learning. Then we present our methodology explaining all three network architectures (Compact, Deep and Bi-Deep) and detecting mechanism in III. As ALADDIn is designed to cater for unique InSAR like data structures, it is erroneous to compare it with existing off-the-shelf deep learning based anomaly detectors that are trained on a frame-frame video data. Finally, as a proof of efficiency and accuracy, we put forward detailed experimental analysis of test results for real, normal and synthetic test cases when passed through all three models (Compact, Deep and Bi-Deep) in Section IV. In conclusion, we discuss the key contributions of this paper in Section V.

II. InSAR data

A Synthetic Aperture Radar (SAR) satellite image is a 2D array of complex numbers encoding amplitude and phase

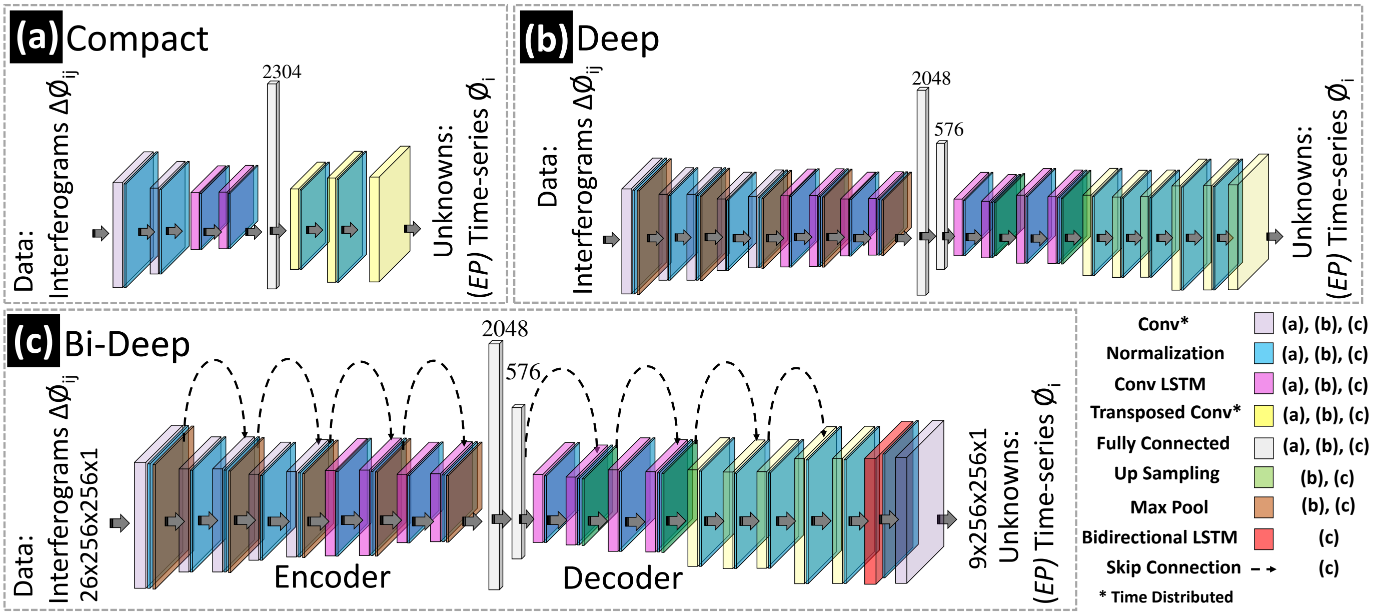

(a) Compact
Data: Interferograms ΔØij
2304
Unknowns: (EP) Time-series Øi
(b) Deep
Data: Interferograms ΔØij
2048
576
Unknowns: (EP) Time-series Øi
(c) Bi-Deep
Data: Interferograms ΔØij 26x256x256x1
2048
576
Encoder
Decoder
9x256x256x1 Unknowns: (EP) Time-series Øi
Conv* (a), (b), (c)
Normalization (a), (b), (c)
Conv LSTM (a), (b), (c)
Transposed Conv* (a), (b), (c)
Fully Connected (a), (b), (c)
Up Sampling (b), (c)
Max Pool (b), (c)
Bidirectional LSTM (c)
Skip Connection (c)
* Time Distributed

Fig. 2. The network architecture of (a) Compact, (b), Deep and (c) Bi-Deep models are shown. Compact includes time-distributed 2D convolutional layers with stride in place of maxpooling, followed by layer normalization, 2D convolutional LSTM, a fully connected layer and transpose convolutions to decode the encoded features. Where as the Deep includes greater number of conv-LSTM layers in the ecoder as well as the decoder, with an extra fully connected layer, maxpooling layers replaced by stride in convolutions and upsampling layers instead of transpose convolution. The Bi-Deep architecture is similar to the Deep one but with a major difference in input of each layer, here skip connections are placed to merger features and a bi-directional LSTM layer is added in the end.

information of microwave radar waves that are emitted by satellite, backscattered from the Earth's surface, and recorded again by the satellite's antenna. In order to measure movement of the Earth's surface, two SAR images of the same location but captured at different times can be used to construct an unwrapped Interferometric SAR (InSAR) image. This image is called an unwrapped interferogram (hereafter referred to simply as an interferogram or InSAR image) and represents a map of how the ground has moved towards or away from the satellite (i.e. a 1D displacement in the satellite's 'line-of-sight') in the time interval between the two SAR measurements. The largest nuisance signals in interferograms arise from uncertainties in satellite orbits and from changes in atmospheric conditions, and are commonly considered as noise when trying to measure ground motion. In this study we use unwrapped Sentinel-1 interferograms obtained from the global LiCSAR processing system developed by the UK's Centre for the Observation and Modelling of Earthquakes, Volcanoes and Tectonics (COMET) [3]. These images typically cover a region of the Earth's surface $\sim$250 km $\times$ 250 km, with pixels of size 80×80 m. There is a major difference between regular video data that is commonly analysed using deep-learning methods and InSAR data. In video data an individual image contains information on the position of objects at a single acquisition time, but an individual interferogram instead contains information on the difference in position between two acquisition times. The unique structure of the dataset is illustrated by Figure 1a; in this simple example, six interferograms (curved lines) capture the differences between four epoch images (circles) that are each associated with an individual SAR satellite image. These epoch images are always unknown; due to the way in which unwrapped interferograms are constructed, these 2D fields cannot be directly calculated from the SAR images themselves. Nuisance signals associated with an epoch (e.g. Ep_3 in Figure 1a) can be mapped into associated interferograms (pink outlined images and lines) via a simple spatio-temporal relationship. But permanent ground displacement that takes place between two Epochs (e.g. between Ep_2 and Ep_3 in Figure 1b) maps into a different set of interferograms (red outlined images and lines) according to a different relationship. In the following section we describe how it is possible to use CNNs to exploit this key difference and therefore to detect deformation.

III. Methodology

Autoencoders are a type of artificial neural network that are designed to understand the underlying distribution of a dataset by learning to reconstruct the data from a transformed version of them. The transformation of the input is referred to as an encoding and usually results in a compressed representation of the data. The reconstruction of the data is referred to as decoding and usually involves up-sampling of the encoded data. This has proven to be a very powerful approach with applications in image denoising, segmentation, 2D-reconstruction and image generation purposes [16], [17], [18]. Autoencoder-based anomaly detection in deep learning often refers to training a model to learn normality underlying a given labelled dataset. The autoencoder learns to reconstruct the input, which can be an image or a video sequence [19] or multivariate sequence data [20] [21]. So when fed with new input data the anomalies are identified with high error. The autoencoder based anomaly detection can be designed using

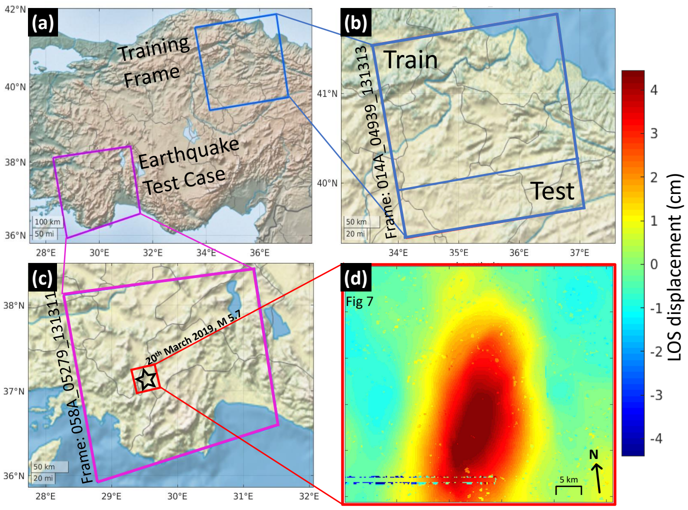


Fig. 3. (a) The map shows the geo location of the train and test data. (b) The training frame is located in the north-east of Turkey, where as the real earthquake test case, shown in (c) and (d) is located in the south-western part of Turkey. (d) Shows the zoomed-in image of the region where earthquake occurred on the 20th of March 2019. The spatial structure of earthquake which is estimated by our model is shown here in (d) (shown in 7).

different combinations of deep learning layers for example convolutional (for spatial data) [22], LSTM (for temporal data) [23] or combined Convolutional-LSTM (for spatio-temporal data) [24].

The solution to an anomaly detection problem can be developed by first understanding the data (InSAR data in this case) and defining the 'anomolous' class (crustal deformation) in it. The low absolute numbers of interferograms containing transient deformation, and even lower numbers where deformation has been labelled makes this problem better suited to an unsupervised learning.

A. Deep Learning For InSAR

InSAR data have inherent inter-dependent spatial and temporal patterns associated with background nuisance signals due to the unique data structure (Figure 1a). This prominent feature of the data can be learned so that anomalous signals corresponding to deformation (Figure 1b) are identified. Long Short Term Memory (LSTM) cells [25] are often used in similar cases to learn from time dependent data. LSTMs are a type of Recurrent Neural Network (RNN) that directly model the temporal dynamics in the data stream for more accurate prediction. RNNs are commonly used for speech recognition, language modeling, translation and image captioning [26], [27], [28], but they suffer from a vanishing-gradients problem which limits how much memory they can hold. Information is propagated through each time-point in a RNN and so gradients are computed for each hidden layer (all across time) using backpropagation starting from final layer to the initial layer. Depending on the length of time and the number of layers, the small derivatives when multiplied together (causing a ripple-effect) decreases (vanishes) the gradients exponentially [29]. In contrast, LSTMs are capable of retaining information for longer intervals so they have been successfully used for capturing changes that are prolonged in time, e.g. CCTV surveillance [30]. LSTM models tend to learn from the temporal dynamics of the sequence in 1D, whereas multiple filters in a convolutional layer span and perform convolutions on 2D or 3D data, preserving pixel-based spatial information. For problems like the one tackled in this study, where, to detect the time-stamp and location of anomaly it is important to learn both the spatial structures and temporal patterns of input data, the LSTM cells are applied with convolutions as a mathematical operator. The internal $1D$ matrix multiplications in the LSTM layer are converted in convolution operations. These are termed as convolutional LSTMs [31] and are able to maintain the dimensions of the input data for images or videos. Where there is no ground truth available and it is both expensive and time consuming to mark, label or caption abundant video or image data, unsupervised or semi-supervised deep learning techniques involving convolutional LSTMs [32] are used to understand changes and track object movements [33].

In contrast to the existing approaches to machine-learning analysis of InSAR data [7], [12], [9], [13], we develop our method starting with three building blocks: the encoder that models spatio-temporal patterns in the interferogram sequence; the fully connected (FC) layers that transition these encoded features to corresponding epoch responses; and the decoder that then up-samples these epoch responses. A FC layer is a 1D layer containing feed forward neurons. Each neuron in them is connected to every single feature encoding of preceding and succeeding layers, representing every pixel in time and space. This strictly ensures that the model learns: (I) the spatio-temporal patterns within the interferogram set (while encoding); (II) the relationship within epoch responses and their difference (while transitioning using the fully connected layers and also constrained by the loss function, that is defined in eq 1); and (III) the spatio-temporal patterns within the sequence of epochs (while decoding). In order to encode the distribution of an input sequence, so that the LSTM layers can learn spatio-temporal patterns from it, the input images are fed to convolution blocks each block includes a time-distributed convolution and a layer normalization. The benefit of time-distributed layer is that it ensures the same convolution is applied on each temporal instance in the input sequence. For example, as we have X IFG in the sequence, then in the time-distributed layer there would be X number of convolutional filters applied individually on each interferogram, giving us X number of features in time that each contain 2D spatial information. The weights of this layer are distributed among the X filters, this helps to connect the features learned for each temporal instance within an input sequence and makes it computationally manageable for backpropagation.

The approach of [13] is in some respects similar to ours, although the time-series (in our case epochs that are generated by our model) they used for training is computed using SBAS inversion and topography is added as an extra channel mid-way in the model. Although, the same length of epoch time-series is used, [13] select a much lower (48 x 48) spatial resolution. Unlike our approach, the autoencoder is supervised and trained on synthetic data. Although [13] predicts the accumulated ground deformation of an InSAR time-series, but the model does not know if that deformation is anomalous or not. In addition, the model prediction in [13] provides only the spatial structure of cumulative deformation, without allowing the exact timing or duration of the event to be retrieved.

1) Compact Model: We start by training a Compact model (figure 2a), with just four of these blocks, two with time-distributed convolution layers with strides and two with conv-LSTM. For an input of $X\ IFG$ that are made from Y *EP*, the encoded features, which are ordered in time and are passed to LSTM blocks that learn and retain the spatio-temporal pattern. To reconstruct the *EP* responses from these features, they are first transformed to Y representations in temporal order by the FC layers, where Y is the number of *EP* responses. At this stage the features are of size $X \times 16 \times 16 \times F$ (where F is the number of filters) and the number of neurons in the FC layers are used in accordance with the size of features we require after transformation, i.e. $Y \times 16 \times 16 \times F$. Likewise, in the decoder, transpose convolution layers are used to upsample the reshaped NN features. The NN with 2304 neurons are used for the transformation of $X\ IFG$ to Y *EP*, where X is 26, Y is 9, so, the number of neurons are $9 \times 16 \times 16 = 2304$.

2) Deep model: We further developed these building blocks and increase the overall depth of our model, creating what we call the Deep model (figure 2b). This is designed with a balance of time-distributed convolutions and the conv-LSTM layers in the encoder as well as the decoder. Each of the convolution blocks in the Deep model includes a maxpooling layer that removes invariances like shift and scale in the feature representation and ensure computationally manageable trainable parameters across the model by down-scaling and extracting most important features. Apart from depth, other major changes are the addition of upsampling layers instead of the stride, smaller filter size (to precisely capture local features) in the convolutions and $tanh$ [34] as an activation layer after each layer except the neural network. An activation function is applied on the top of layers to introduce non-linearity and output of $tanh$ is zero-centered and ranges from -1 to 1, hence strongly mapping both negative and positive inputs.

3) Bi-Deep model: Finally, for the Bi-Deep model (figure 2c) we added separate skip connections for encoder and decoder and also included a bidirectional-conv-LSTM layer. These skip connections ensure the forwarding of any residual feature representation in the previous layer also it helps in merging features learned by a time-distributed convolution and a conv-LSTM layer and other combinations like a FC layer and a time-distributed convolution layer etc. Unlike U-Net [11], long skip connections cannot be used because of different feature sizes, as our model reconstructs Y *EP* from the encoded $X\ IFG$. So, to ensure flow of information between layers of our autoencoder, we perform feature concatenation via short skip connections separately for encoder and decoder. The bidirectional-conv-LSTM layer retains information by spanning the features propagating both forwards and backwards in time, to prevent bias in the predicted *EP* sequence associated with their order in time.

TABLE I
DATASET DETAILS

Data	No. of patch locations	No. of sequences for each patch location	Total sequences
Train	365	25	9125
Validation	62	25	1550
Synth Test	1	10	10
Real EQ	1	7	7

B. Experimental Details

In general all these models (Compact, Deep and Bi-Deep) attempt to learn the relationship between normal spatio-temporal patterns of background noise in a set of related interferograms and the unknown 2D fields of that same noise in their constituent epochs (e.g. Figure 1a). Our model transforms $X\ IFG$ in the encoding feature space to Y *EP* in the decoding feature space. The transformation is constrained through the loss function

$$Loss = MSE(IFG2TS(Output), Input) \tag{1}$$

, where $IFG2TS$ is a custom layer that converts the sequence of Y estimated epochs into X interferograms by simple subtraction in each case of the 1st constituent epoch from the 2nd, as per Figure 1a. Therefore, the mean squared error (MSE) is computed between the input set of interferograms and the reconstructed ones.

The data values in the interferograms are of varying range and positive and negative values hold equal importance, so activation layers like $ReLU$ (ranges from 0 to ∞) and $Sigmoid$ (ranges from 0 to 1) cannot be used. Therefore a $tanh$ activation function is applied for all convolutions except the output time distributed convolution layer, so that the prediction of *EP* values are not bound by any range.

We train and test our method using Sentinel-1 InSAR data from Turkey (figure 3), obtained from the COMET LiCSAR processing system [3]. Turkey has been a focus area for initial development of this processing system, so has the largest dataset available for training and testing our method [35].

The data frame on the northern coast of Turkey (LiCSAR Frame name: $014A\_04939\_131313\_2214\_2147$, spatial extent $\sim 250km \times 250km$) is used for training. In order to manage the complexity of model and memory required to train large number of parameters, the frame is divided into cubes of size $256 \times 256 \times 26$ pixels (a spatial extent of approximately $20.5km \times 20.5km$) with a fifty percent spatial overlap (in both E-W and N-S directions) and instead of passing the whole time-series in every training iteration, a set of 26 interferograms that cover 9 *EP* is passed. The temporal sliding window is 9 *EP* in length, and moves with a stride of 4 ensuring a temporal overlap of $> 50\%$ between successive input sequences. The 26 interferograms link each *EP* with all successive and preceding *EP* within the sequence, up to a maximum distance of 4 forwards and backwards in time. For example, the central *EP* is linked by 8 interferograms to all

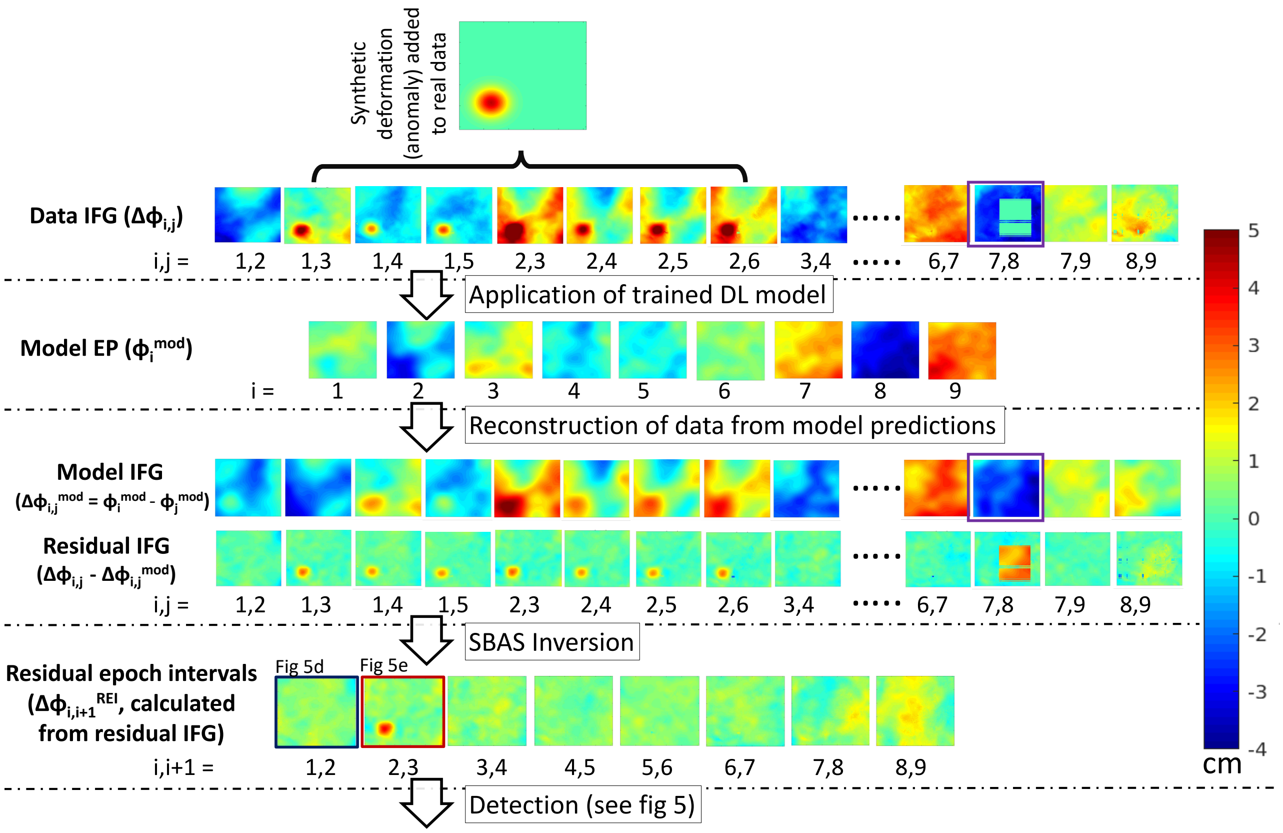

Synthetic deformation (anomaly) added to real data
Data IFG (Δϕi,j)
i,j = 1,2 1,3 1,4 1,5 2,3 2,4 2,5 2,6 3,4 ····· 6,7 7,8 7,9 8,9
Application of trained DL model
Model EP (ϕimod)
i = 1 2 3 4 5 6 7 8 9
Reconstruction of data from model predictions
Model IFG
(Δϕi,jmod = ϕimod - ϕjmod)
Residual IFG
(Δϕi,j - Δϕi,jmod)
i,j = 1,2 1,3 1,4 1,5 2,3 2,4 2,5 2,6 3,4 ····· 6,7 7,8 7,9 8,9
SBAS Inversion
Residual epoch intervals
(Δϕi,i+1REI, calculated from residual IFG)
Fig 5d
Fig 5e
i,i+1 = 1,2 2,3 3,4 4,5 5,6 6,7 7,8 8,9
Detection (see fig 5)
5
4
3
2
1
0
-1
-2
-3
-4
cm

Fig. 4. The results for the synthetic test case (detailed discussion in section IV-A). The synthetic test signal - a 2D Gaussian with peak amplitude of 4.34 cm and exponential length-scale of 10.5 km; this signal was added to the interferogram time series. Following the synthetic input, are the *EP* predictions, using them the interferograms are reconstructed and residuals are computed. These residuals are used in least square inversion to compute residual based epoch intervals.

other *EP*, but all other *EP* in the sequence are linked with less than 8, to a minimum of 4 interferograms for the *EP* at the start of the sequence and the *EP* at the end of the sequence. The order with which the 26 IFG are passed to the model is sequential, for example, the first sequence passed is in order:

- 26 Interferograms: IFG_{12}, IFG_{13}, IFG_{14},IFG_{15}, IFG_{23}, IFG_{24}, IFG_{25}, IFG_{26}, IFG_{34}, IFG_{35}, IFG_{36}, IFG_{37}, IFG_{45}, IFG_{46}, IFG_{47}, IFG_{48}, IFG_{56}, IFG_{57}, IFG_{58}, IFG_{59}, IFG_{67}, IFG_{68}, IFG_{69}, IFG_{78}, IFG_{79}, IFG_{89}.
- Covering initial 9 Epochs: EP_1, EP_2, EP_3, EP_4, EP_5, EP_6, EP_7, EP_8, EP_9.

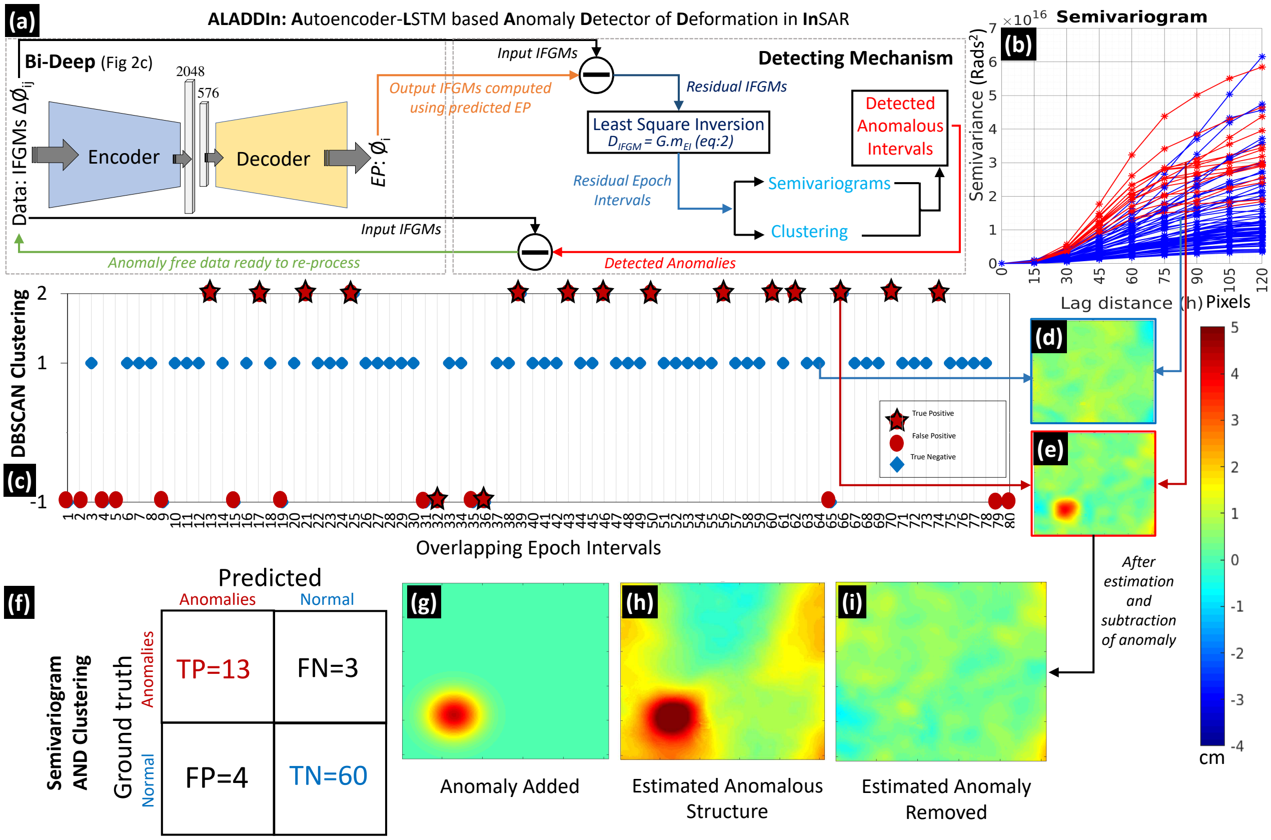
The dataset details are given in Table I. The model is trained using Keras with TensorFlow backend. Due to the large size of the images in memory the batch size was set to 1. Adam optimizer [36] was used with a learning rate of 0.00001. A lower learning rate gives the model a chance to learn features through steady changes in loss instead of rapid fluctuations.

C. ALADDIn: Autoencoder-LSTM based Anomaly Detector of Deformation in InSAR

The deep learning models are trained on background atmospheric noise, so that in case of an anomaly there must be a misfit. The training data has been manually reviewed for any anomolous events, thus confirming that the training patch sequences contain only the 'normal' background atmospheric noise. Once the test data are passed through the models, the residuals are computed between the reconstructed interferograms and original data. Because any anomaly will appear in multiple interferograms (and therefore also multiple residuals), we first reduce the residual dataset down to a mutually exclusive set of N_{EI} (figure 4) "epoch intervals". An epoch interval is a difference image that spans two successive epochs, so therefore N_{EI} is equal to one less than the number of epochs. In order to estimate this set of N_{EI} residual epoch intervals, we perform a linear least squares inversion on a pixel by pixel basis of our N_{IFG} residuals as follows (based on the SBAS approach from [37]):

$$d_{IFG} = G.m_{EI}, \tag{2}$$

where d_{IFG} is a $N_{IFG} \times 1$ array containing pixel values of all N_{IFG} residual interferograms, m_{EI} is the $N_{EI} \times 1$ vector of epoch intervals that we wish to solve for, and G is the $N_{IFG} \times N_{EI}$ sized design matrix for this system of equations, containing 1s and 0s only. Eq 3 shows an example of matrix G for a set of six residual interferograms (IFG_{12}, IFG_{13}, IFG_{14}, IFG_{23}, IFG_{24}, IFG_{34}) corresponding to the simplified cartoon structure shown in Figure 1, constructed from four epochs (EP_1, EP_2, EP_3, EP_4), which will output three epoch intervals (EI_{12}, EI_{23}, EI_{34}) based on the residuals. These intervals are essentially equivalent to the shortest



Semivariogram AND Clustering	Predicted: Anomalies	Predicted: Normal
Ground truth: Anomalies	TP=13	FN=3
Ground truth: Normal	FP=4	TN=60

Fig. 5. The ALADDIn pipeline is shown here. (a) shows the Bi-Deep model (sec III-A3) plugged in with the detecting mechanism that involves semivariogram analysis (sec III-C1) and DBSCAN clustering (sec III-C2), summing up the ALADDIn pipeline. (b) shows the semivariogram plot, red lines are the synthetic anomalies- all have high RMSE with respect to the majority of semivariograms corresponding to 'normal' epoch intervals (blues lines). (c) shows the results of clustering when it is applied on residual epoch intervals. The results for the synthetic test case (detailed discussion in section IV-A). (d) and (e) are one of the normal and anomolous residual epoch intervals (also shown in figure 4). (f) Shows the confusion matrix for the synthetic test results for the Bi-Deepmodel, where 13 out of a total of 16 are correctly detected. (g) The synthetic test signal - a 2D Gaussian with peak amplitude of 4.34 cm and exponential length-scale of 10.5 km; this signal was added to 8 different time intervals in the interferogram time series, which makes a total of 16 anomalies due to the overlap between successive sequences. (h) shows the estimated spatial structure of synthetic anomaly for one of the intervals and (i) shows the undetected output of same interval when the estimated structure is subtracted and the data is reprocessed.

spanning set of residual interferograms (e.g. EI_{12} is equivalent to IFG_{12} = EP_2 - EP_1) but are instead estimated from the full set of residual interferograms so are more robust to noise in any one residual image.

$$\begin{pmatrix} IFG_{12} \\ IFG_{13} \\ IFG_{14} \\ IFG_{23} \\ IFG_{24} \\ IFG_{34} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ET_{12} \\ ET_{23} \\ ET_{34} \end{pmatrix} \tag{3}$$

Due to the overlap between successive sequences, most epoch intervals occur twice in the time series and some appear three time. Epoch intervals for every sequence are computed by solving equation 2 for m_{ET} and then they are concatenated together to make one overlapping time-series of residual epoch intervals. These intervals are then automatically analysed for the presence of spatial anomalies. This is achieved by two complementary analysis methods: semivariogram analysis [38] and density based clustering [39].

1) Semivariogram: An empirical semivariogram is an estimate of how pairs of samples within a dataset differ as a function of distance. The semivariance $\gamma(h)$ for distance h is:

$$\gamma(h) = \frac{\sum_{N(h)} [Z_i - Z_{i+h}]^2}{2(|N(h)|)}, \tag{4}$$

where Z_i is the value at pixel location i and $N(h)$ is the total number of pairs that lie at distance h. The spatial variability measured by the semivariogram can account for deformation that affects only certain spatial frequencies, e.g. capturing deformation that is spread over small regions, and separating such anomalies from larger areas that are 'normal'. These small but significant changes are less likely to be detected by simply computing bulk differences between actual and reconstructed images (e.g. by a Mean Squared Error).

We expect that epoch intervals containing no anomaly will all have similar spatial structure and therefore will also have similar empirical semivariograms, whilst epoch intervals containing anomalies will have semivariograms that differ substantially from this normal structure. This can be seen in Figure 5b, where the semivariograms for residual EI that

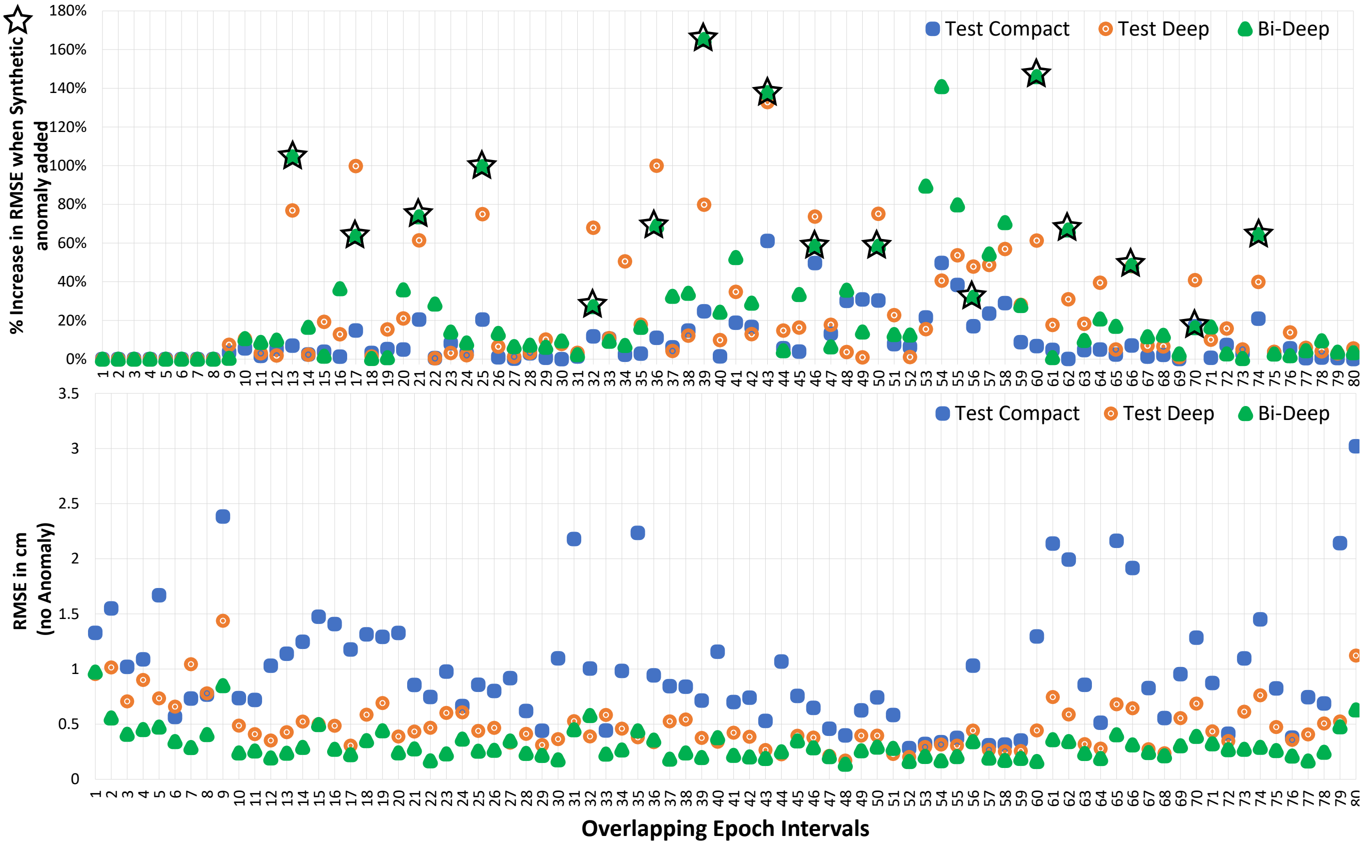


Fig. 6. (Bottom) The plot shows the comparison of RMSE i.e. the reconstruction error all models. (TOP) The plot shows the percentage increase in the error when synthetic anomaly (represented by black star) is added in the same testing sequence. These plot indicates that the Bi-Deep has lowest reconstruction error (bottom) and when an anomaly is added, it has a highest percentage increase in reconstruction error (top) that is required to detect them. The x-axis covers total 10 sequences each containing 8 epoch intervals, and last 4 of every sequence overlaps with first 4 of next sequence.

TABLE II
MEAN RECONSTRUCTION ERROR OF INTERFEROGRAMS, *EP*-INTERVALS

	Mean RMSE IFG						Mean RMSE Ep-Intervals					
	Normal	Synth-Anomaly	% Increase (Synth)	Real-Anomaly	Estimated Anomaly Removed	% Increase (Real)	Normal	Synth-Anomaly	% Increase (Synth)	Real-Anomaly	Estimated Anomaly Removed	% Increase (Real)
Test Compact	1.05	1.15	15.18	1.59	1.53	9.95	0.99	1.03	9.48	1.40	1.38	2.46
Test Deep	0.53	0.66	**33.7**	0.87	0.82	**18.8**	0.48	0.56	23.49	0.75	0.73	**7.67**
Bi-Deep (ALADDIn)	**0.40**	**0.47**	28.73	**0.79**	**0.77**	10.35	**0.30**	**0.36**	**29.42**	**0.66**	**0.65**	6.66

contain synthetic anomalies (red lines) are significantly different from semivariograms corresponding to 'normal' epoch intervals (blue lines). A semivariogram is calculated for each residual epoch interval, and the root-mean-squared-error is computed between each semivariogram and all others in the entire set of residual epoch intervals across all sequences. The threshold used to separate the anomalous values varies per study and is not fixed a-priori, as the spatial structure of the background noise will vary depending on the dataset, resulting in varying semivariance values. But in each case the key assumption is that deformation events are rare. The threshold is computed using the inter-quantile range of the average error values.

2) *Density Based Spatial Clustering (DBSCAN):* The second detection operation we perform is density based clustering (DBSCAN) [39] of the residual epoch intervals. Under normal circumstances the residual epoch intervals are expected to be similar with values near to zero (as they are accurately reconstructed by the model, e.g. see figure 4 bottom row), but in case of an anomaly or multiple anomalies within a sequence, there must be an interval containing the spatial

structure of that anomaly. So, the goal is to separate all normal intervals in one cluster and anomalies in other clusters without any prior knowledge of the instances of anomalies in a sequence. This algorithm performs clustering based on the density of data points and has the advantage for this unsupervised problem of not requiring a-priori specification of the number of clusters. As we have an overlapping sequence of residual epoch intervals, so each time-interval occurs at least twice. We use the prior knowledge of this overlap in epoch intervals to set the minimum points in a cluster to be two. Due to the varying nature of data-points, we compute search radius (epsilon) for the algorithm separately for each sequence, by sorting and plotting the distance to the nearest n points for each point. Epoch intervals are classed as anomalous if they do not fall within the predominant cluster (Figure 5c).

Finally the classified anomalies from the semivariogram and clustering analysis of epoch interval time series are combined using an AND operation in order to reduce the number of false positives. DBSCAN clustering is prone to false positives due to its sensitivity to the distance metric, and for our synthetic test set we reduce the number of false positive from 12 when using DBSCAN to just 4 when combining it with the semi-variogram analysis. Table III shows that combining DBSCAN and semivariogram analysis gives high overall accuracy, as the false positives from DBSCAN are mitigated by incorporation of the semivariogram analysis.

IV. Results and Analysis

We evaluate all our models with three testing scenarios, i.e. a normal test sequence with no deformation (see fig 3b), the same normal test sequences but with synthetic anomalies added, and test data that contains a real earthquake (see figure 3c). The purpose of normal test case is to show the comparison of reconstruction error between an anomaly that occurred within a normal sequence (as demonstrated in figure 6). All of these testing sequences are from a different location than the training data. Due to the fact that no ground truth is available for the epoch responses, the extent of our model's accuracy can be judged by the accuracy of interferogram reconstruction. The reconstruction error used for analysis is root-mean-squared-error $RMSE$ between input IFG i.e. ground truth and the reconstructed IFG. An accurate model should result in a reconstruction error near to or equal to zero when a normal test sequence is passed, where as the error should increase by a large fraction when a synthetic deformation is added in that same normal test sequence. The detecting mechanism rely on the output (*EP* responses) of all three models, the overall accuracy of detection depends on TP (true positives), TN (true negatives), FP (false positives) and FN (false negatives), where positive refer to anomolous class and negative refers to normal class. We independently analyse the test results of all models first, then plug in the detection mechanism on top of it and independently investigate the overall accuracy of detection by splitting up the semi-variogram and DBSCAN mechanism and also by merging them together in an AND and OR combination.

A. Synthetic test case of Gaussian deformation signal

Ground truth for real-world deformation signals in InSAR data is rarely available, so in order to assess the accuracy of our framework we first simulate a simple deformation anomaly (figure 5g) that has the structure of a 2D Gaussian in space and is effectively instantaneous in time with respect to the temporal-frequency of the data (i.e. the deformation event takes place completely in the time-period between two epochs, which in this dataset is 6 days). The spatial structure of this signal is given by:

$$Z(x,y) = A.\exp\left(-(x^2+y^2)/r\right) \tag{5}$$

where the exponential length-scale r = 10.5 km, the amplitude A = 4.34 cm and x and y are spatial coordinates relative to the location of Gaussian peak. This is an ideal test signal as the amplitude and spatial size of this structure is similar to that of noise in the data, as shown in Figure 4 and 5g. To enable robust assessment of our detection accuracy and minimise the impact of the natural variability of noise throughout our dataset on this assessment, we create a synthetic test case with the same anomaly added to our data at multiple instances in time. Therefore the synthetic anomaly is added at 8 different time instances in the real interferogram dataset for a patch location that is separated from the train and validation data. Each interval occurs twice due to the overlap of successive sequences, so in total there are 16 anomalous synthetic deformation structures in the test case. This patch features in a total 10 sequences spanning July 2017 to April 2018, and contains 260 interferograms and 80 epoch intervals, out of which 16 are anomalous and 64 are normal. The reconstruction error is plotted in figure 6(bottom), showing minimum error recorded by Bi-Deep model, although an elevation in error can be seen when in the same normal sequence the synthetic anomaly is added (figure 6 (top)). A comparison of mean error values for all scenarios and all models (Compact, Deep and Bi-Deep) can be seen in table II. In order to further analyse the difference in error, percentage increase in RMSE is also computed between the 'real' anomolous test set and the cleaned set (when estimated structure of anomaly is removed). The model predicts the time and location of these anomalies with an overall accuracy of 91.25% (presented in table III) and a true positive rate of 81.25%. The full confusion matrix is shown in figure 5 (f). To compute the confusion matrix and the accuracy score, each overlapping interval is treated as an individual anomaly. figure 4 also shows the results from the 10th and final sequence, which is the worst constrained sequence in the dataset as there is no overlapping sequence available for the last 4 epochs. Despite this, the images in figure 4 show how the anomaly is still accurately detected in the residual images. Figure 4 (purple box) also demonstrates how our method can accurately estimate the spatial structure of interferograms even when the ground truth images contain missing data (purple box figure 4 first and third row).

B. Earthquake test case

Finally, we also test our model's ability to detect a real Magnitude 5.7 earthquake that occurred in south west Turkey

TABLE III
OVERALL ACCURACY OF MODELS WITH ALL TESTING SCENARIOS.

	Semivariogram Analysis			Clustering Analysis			**V AND C**			V OR C		
	Normal	Synth-Anomaly	Real	Normal	Synth-Anomaly	Real	Normal	Synth-Anomaly	Real	Normal	Synth-Anomaly	Real
Test Compact	0.825	0.76	0.78	0.45	0.57	0.60	0.91	0.78	0.92	0.36	0.55	0.46
Test Deep	0.77	**0.95**	0.82	0.40	0.42	0.58	0.81	0.83	0.82	0.36	0.53	0.58
Bi-Deep (AladdIn)	**0.88**	0.91	**0.83**	**0.96**	**0.90**	**0.73**	**0.96**	**0.91**	**0.85**	**0.88**	**0.90**	**0.71**

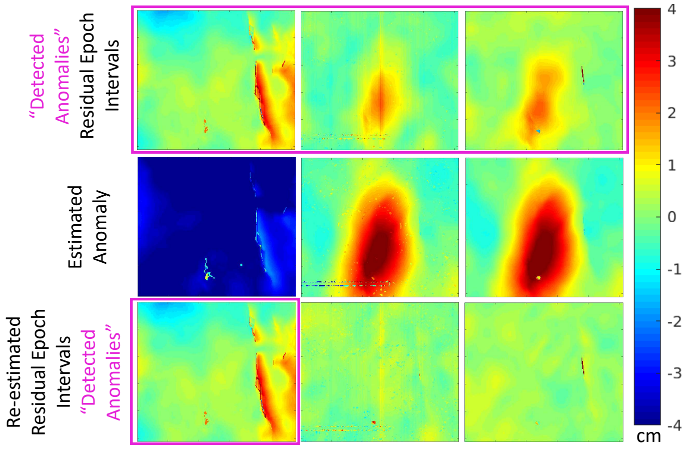


Fig. 7. The figures shows the results of detection and estimation of anomaly when a real earthquake of magnitude 5.7 is tested through the model. Top row shows the residual epoch intervals that are detected as anomolous. 2nd row shows the estimated spatial structures of detected anomalies, which are subtracted from the time series before this 'cleaned' data is then processed again. 3rd row shows the new detection output, after removing previously detected anomalies (pink box) and reprocessing it through the method. Bottom row, pink box, is still identified as an anomaly, which is also unwrapping errors. In contrast, the earthquake intervals are now identified as normal

on 20th March 2019 [40]. Unlike many transient deformation signals of interest, both the time of this event and its location are known, which means we can verify whether our model can correctly identify the earthquake interval as anomalous. Sentinel-1 InSAR data for this test case includes 7 sequences starting from September 2018 to April 2019, and due to the overlap between successive sequences, both the 6th and 7th sequences include the earthquake anomaly.

The 2 realisations of the earthquake interval are accurately detected as anomalous (shown in figure 7 pink box), along with another anomaly which on further inspection is a feature of InSAR data known as an unwrapping error (Figure 7 pink box). These errors occur during the process of converting discrete cycles ('fringes') of $+/-\pi$ phase into continuous values and are particularly significant in areas of phase incoherence associated with steep topography, changes to surface scatterers between satellite image acquisitions or exceptionally high deformation rates. Phase unwrapping errors have magnitudes in multiples of 2π in individual interferograms (potentially several cm apparent displacement) and propagate through time series analysis to hinder the interpretation of tectonic or volcanic deformation. Deep learning approaches to phase unwrapping for InSAR have been proposed by [41], [42], [43] The identification of such errors is valuable in itself, as they often need to be fixed in order to improve a wide range of InSAR-derived products and results. The spatial structures of these three anomalies are then computed by taking the mean of all original data interferograms that contain the anomalous epoch interval (Figure 7 2nd row). In future, for more complex temporal patterns of deformation than those investigated here, where some interferograms may include contributions from more than one anomaly, we can use the same inversion approach as we applied to the interferogram residuals in Equation 3. This would enable us to jointly estimate the spatial structure of each anomaly from the subset of original interferograms that have been identified as containing anomalies. In order to examine our detection results and predicted estimate of anomaly, we remove these estimated anomaly signals from our original interferogram time-series and then re-process our analysis. The re-processed results (Figure 7 3rd row) show that the spatial and temporal patterns of earthquake deformation have been accurately predicted and largely removed because the intervals containing the earthquake are no longer identified as anomalous by the detecting mechanism. In contrast, the unwrapping error persists and is flagged again because it is a data error rather then a natural transient phenomena. In all experiments, the first 4 and last 4 epoch intervals are ignored during the identification of anomalies because they are always poorly estimated, and are always separated by DBSCAN into the negative cluster with epoch intervals that derived from interferograms with large amounts of missing data.

V. CONCLUSION

In this paper, we have attempted to systematically automate the detection and extraction of transient episodes of crustal deformation applicable to global InSAR datasets, a goal which is valuable for a wide range of solid earth and natural hazard applications. We propose a new, state-of-the-art deep-learning based anomaly detection approach for the automatic identification of transient deformation events (anomalies) in noisy time-series of unwrapped InSAR images, without requiring supervision or labelling of known example events. Our novel workflow learns patterns of the 'normal' non-tectonic signals in the InSAR dataset, leveraging the unique three-dimensional structure of the interferogram stack to estimate the unknown 2D fields that correspond to individual SAR acquisition dates (epochs). Our method automatically flags intervals containing deformation and separates the deformation from the normal background time-series. Our method can successfully identify synthetic deformation signals with peak line-of-sight displacements of 4.3 cm and of length scale 10 km, with high overall accuracy 91.25% and true positive rate 81.25%, and has also been used to successfully identify a Magnitude 5.7 earthquake and unwrapping errors within data from SW Turkey - a geographic region distinct from the location of the training dataset. We plan to further develop this method by incorporating joint analysis of data from multiple overlapping InSAR tracks, undertaking detailed testing on deformation events with varying temporal and spatial signatures, and employing domain adaptation so that the method can be applied to varied global regions beyond the training region.

References

[1] J. Elliott, R. Walters, and T. Wright, "The role of space-based observation in understanding and responding to active tectonics and earthquakes," *Nature communications*, vol. 7, p. 13844, 2016.

[2] M. A. Floyd, R. J. Walters, J. R. Elliott, G. J. Funning, J. L. Svarc, J. R. Murray, A. J. Hooper, Y. Larsen, P. Marinkovic, R. Bürgmann *et al.*, "Spatial variations in fault friction related to lithology from rupture and afterslip of the 2014 south napa, california, earthquake," *Geophysical Research Letters*, vol. 43, no. 13, pp. 6808–6816, 2016.

[3] M. Lazecký, K. Spaans, P. J. González, Y. Maghsoudi, Y. Morishita, F. Albino, J. Elliott, N. Greenall, E. Hatton, A. Hooper, D. Juncu, A. McDougall, R. J. Walters, C. S. Watson, J. R. Weiss, and T. J. Wright, "Licsar: An automatic insar tool for measuring and monitoring tectonic and volcanic activity," *Remote Sensing*, vol. 12, no. 15, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/15/2430

[4] R. Youssef, M. Aniss, and C. Jamal, "Machine learning and deep learning in remote sensing and urban application: A systematic review and meta-analysis," in *Proceedings of the 4th Edition of International Conference on Geo-IT and Water Resources 2020, Geo-IT and Water Resources 2020*, 2020, pp. 1–5.

[5] L. Ma, Y. Liu, X. Zhang, Y. Ye, G. Yin, and B. A. Johnson, "Deep learning in remote sensing applications: A meta-analysis and review," *ISPRS journal of photogrammetry and remote sensing*, vol. 152, pp. 166–177, 2019.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[7] N. Anantrasirichai, J. Biggs, F. Albino, P. Hill, and D. Bull, "Application of machine learning to classification of volcanic deformation in routinely generated insar data," *Journal of Geophysical Research: Solid Earth*, vol. 123, no. 8, pp. 6592–6606, 2018.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.

[9] M. Gaddes, A. Hooper, and M. Bagnardi, "Using machine learning to automatically detect volcanic unrest in a time series of interferograms," *Journal of Geophysical Research: Solid Earth*, vol. 124, no. 11, pp. 12 304–12 322, 2019.

[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[11] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[12] J. Sun, C. Wauthier, K. Stephens, M. Gervais, G. Cervone, P. La Femina, and M. Higgins, "Automatic detection of volcanic surface deformation using deep learning," *Journal of Geophysical Research: Solid Earth*, vol. 125, no. 9, p. e2020JB019840, 2020.

[13] 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, pp. 1–11, 2021.

[14] M. Simons and P. Rosen, "Interferometric synthetic aperture radar geodesy," 2007.

[15] T. Emardson, M. Simons, and F. Webb, "Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation," *Journal of Geophysical Research: Solid Earth*, vol. 108, no. B5, 2003.

[16] Z. Zhu, X. Wang, S. Bai, C. Yao, and X. Bai, "Deep learning representation using autoencoder for 3d shape retrieval," *Neurocomputing*, vol. 204, pp. 41–50, 2016.

[17] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of machine learning research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[18] M. I. Sameen and B. Pradhan, "A novel road segmentation technique from orthophotos using deep convolutional autoencoders," *Korean J. Remote Sens*, vol. 33, no. 4, pp. 423–436, 2017.

[19] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, "Learning deep representations of appearance and motion for anomalous event detection," *arXiv preprint arXiv:1510.01553*, 2015.

[20] W. Lu, Y. Cheng, C. Xiao, S. Chang, S. Huang, B. Liang, and T. Huang, "Unsupervised sequential outlier detection with deep architectures," *IEEE transactions on image processing*, vol. 26, no. 9, pp. 4321–4330, 2017.

[21] E. Marchi, F. Vesperini, F. Weninger, F. Eyben, S. Squartini, and B. Schuller, "Non-linear prediction with lstm recurrent neural networks for acoustic novelty detection," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–7.

[22] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla, "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, 2019, pp. 1409–1416.

[23] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Lstm-based encoder-decoder for multi-sensor anomaly detection," *arXiv preprint arXiv:1607.00148*, 2016.

[24] W. Luo, W. Liu, and S. Gao, "Remembering history with convolutional lstm for anomaly detection," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 439–444.

[25] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with lstm," 1999.

[26] Y. Miao, M. Gowayyed, and F. Metze, "Eesen: End-to-end speech recognition using deep rnn models and wfst-based decoding," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*. IEEE, 2015, pp. 167–174.

[27] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[28] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, and A. Yuille, "Deep captioning with multimodal recurrent neural networks (m-rnn)," *arXiv preprint arXiv:1412.6632*, 2014.

[29] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 6, no. 02, pp. 107–116, 1998.

[30] A. Shah, J. B. Lamare, T. N. Anh, and A. Hauptmann, "Accident forecasting in cctv traffic camera videos," *arXiv preprint arXiv:1809.05782*, 2018.

[31] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.

[32] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," in *Advances in neural information processing systems*, 2015, pp. 802–810.

[33] G. Ning, Z. Zhang, C. Huang, X. Ren, H. Wang, C. Cai, and Z. He, "Spatially supervised recurrent convolutional neural networks for visual object tracking," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2017, pp. 1–4.

[34] B. L. Kalman and S. C. Kwasny, "Why tanh: choosing a sigmoidal function," in *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, vol. 4. IEEE, 1992, pp. 578–581.

[35] Z. Li, T. Wright, A. Hooper, P. Crippa, P. Gonzalez, R. Walters, J. Elliott, S. Ebmeier, E. Hatton, and B. Parsons, "Towards insar everywhere, all the time, with sentinel-1." *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 41, 2016.

[36] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[37] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, “A new algorithm for surface deformation monitoring based on small baseline differential sar interferograms,” *IEEE Transactions on geoscience and remote sensing*, vol. 40, no. 11, pp. 2375–2383, 2002.

[38] H. Wackernagel, *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media, 2013.

[39] H.-P. Kriegel, P. Kröger, J. Sander, and A. Zimek, “Density-based clustering,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 3, pp. 231–240, 2011.

[40] J. Elliott, M. de Michele, and H. Gupta, “Earth observation for crustal tectonics and earthquake hazards,” *Surveys in Geophysics*, vol. 41, no. 6, pp. 1355–1389, 2020.

[41] L. Zhou, H. Yu, Y. Lan, S. Gong, and M. Xing, “Canet: An unsupervised deep convolutional neural network for efficient cluster-analysis-based multibaseline insar phase unwrapping,” *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[42] F. Sica, F. Calvanese, G. Scarpa, and P. Rizzoli, “A cnn-based coherence-driven approach for insar phase unwrapping,” *IEEE Geoscience and Remote Sensing Letters*, 2020.

[43] H. Wang, J. Hu, H. Fu, C. Wang, and Z. Wang, “A novel quality-guided two-dimensional insar phase unwrapping method via gaunet,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7840–7856, 2021.