Detection of tectonic and volcanic deformation as anomalies in InSAR: deep-learning tailored to differential data

Anza Shakeel[1], Richard John Walters[1], Susanna K Ebmeier[2], and Noura Al Moubayed[1]

[1]Durham University
[2]University of Leeds

December 7, 2022

Abstract

There are now more interferograms being generated from global satellite radar datasets than can be assessed by hand. The reliable, automatic detection of true displacement from these data is therefore critical, both for monitoring deformation related to geohazards and understanding solid earth processes. We discuss improvements to an unsupervised, event agnostic method for automatically detecting deformation in unwrapped interferograms. We use an anomaly detection framework that recognises any deformation as "anomalies" by learning the 'typical' spatio-temporal pattern of atmospheric and other noise in sequences of interferograms. Here, we present developments to our prototype model, ALADDIn (Autoencoder-LSTM based Anomaly Detector of Deformation in InSAR) using (1) a self-attention training technique to exploit redundancy in interferogram networks to capture the temporal structure of signals and (2) the addition of synthetic data for training. We evaluate the impact of these developments using two geophysical scenarios: (1) the detection of the same M_w 5.7 earthquake used to test our original model (20.03.2019, south-west Turkey), (2) the persistent uplift of Domoyu volcano (17.05.2017 to 14.12.2018, Argentina). We make a quantitative evaluation of the performance of our method using synthetic test data and find that for peak displacements exceeding a few cm and of length-scale greater than a few hundred metres, overall detection accuracy is 80 to 90%.

Detection of tectonic and volcanic deformation as anomalies in InSAR: deep-learning tailored to differential data

Anza Shakeel[1], Richard J Walters[1], Susanna K Ebmeier[4], Noura Al Moubayed[2]

[1]COMET, Department of Earth Sciences, Durham University, UK
[2]Department of Computer Sciences, Durham University, UK
[3]Department of Earth Sciences, Durham University, UK
[4]School of Earth and Environment, University of Leeds, UK

Key Points:

- We present a novel, unsupervised deep learning architecture tailored to differential InSAR to flag anomalous deformation
- Customised, iterative training system based on temporal self-attention improves model accuracy, but training just on synthetic data does not
- Highest accuracy is for signals with peak line-of-sight displacement of a few cm and of a length scale bigger than a few hundred meters.

Corresponding author: Anza Shakeel, `anza.shakeel@durham.ac.uk`

Abstract

There are now more interferograms being generated from global satellite radar datasets than can be assessed by hand. The reliable, automatic detection of true displacement from these data is therefore critical, both for monitoring deformation related to geohazards and understanding solid earth processes. We discuss improvements to an unsupervised, event-agnostic method for automatically detecting deformation in unwrapped interferograms. We use an anomaly detection framework that recognises any deformation as "anomalies" by learning the 'typical' spatio-temporal pattern of atmospheric and other noise in sequences of interferograms. Here, we present developments to our prototype model, ALADDIn (Autoencoder-LSTM based Anomaly Detector of Deformation in InSAR) using (1) a self-attention training technique to exploit redundancy in interferogram networks to capture the temporal structure of signals and (2) the addition of synthetic data for training. We evaluate the impact of these developments using two geophysical scenarios: (1) the detection of the same M_w 5.7 earthquake used to test our original model (20.03.2019, south-west Turkey), (2) the persistent uplift of Domoyu volcano (17.05.2017 to 14.12.2018, Argentina). We make a quantitative evaluation of the performance of our method using synthetic test data and find that for peak displacements exceeding a few cm and of length-scale greater than a few hundred metres, overall detection accuracy is 80 to 90%.

1 Introduction

The abundance of routinely acquired Synthetic Aperture Radar (SAR) imagery from missions such as the European Space Agency's Sentinel-1 (and anticipated for the NASA-ISRO SAR Mission - NISAR) has led to a surge in deep-learning-based approaches for the detection of deformation (Anantrasirichai et al., 2018; Gaddes et al., 2019; Rouet-Leduc et al., 2021). These efforts are critical for optimising the usefulness of large Interferometric SAR (InSAR) datasets for monitoring deformation, given that the high volumes and rates of data (1000-2000 images/day, $\approx$ 10 TB/day) prevents systematic manual analysis. Detecting deformation in InSAR datasets is critical for monitoring geohazards (especially volcanoes (Ebmeier et al., 2018), slow landslides (Bekaert et al., 2020) and anthropogenic deformation (Semple et al., 2017) and for our understanding of broader tectonic processes (Elliott et al., 2020)). Deep learning approaches also have the potential to transform the emphasis of scientific research, allowing the automated discovery of signals in uniformly analysed regional or global datasets rather than studies focused on locations where deformation is already well known.

Deep learning has been widely applied to the field of remote sensing, (e.g., (Sharma et al., 2020; Ren et al., 2021; Shakeel et al., 2019), mostly to satellite datasets that comprise time sequences of images. In contrast, InSAR has a unique spatio-temporal structure, as interferograms provide information about changes between two dates (differential data). However, the majority of applications of deep learning to InSAR so far have used 2D spatial patterns of phase in individual interferograms (Anantrasirichai et al., 2018). This has allowed the application of off-the-shelf models like AlexNet (Krizhevsky et al., 2012) or U-Net (Ronneberger et al., 2015), e.g., modified by (Chen et al., 2022) for semantic segmentation of active landslides. A disadvantage of this is that off-the-shelf methods are rigid in terms of input size (e.g., the input size of VGG (Simonyan & Zisserman, 2015) is $224 \times 224 \times 3$, where 224 is the size of an image in X and Y dimensions and 3 represents the R(red), G(green) and B(blue) channels of a digital image) because they are built on existing models that were initially trained on RGB images. An alternative approach is to use time-series derived from interferograms to obtain time sequences of images for input (Gaddes et al., 2019), although this requires an additional processing step that also has the potential to introduce errors. Most deep learning approaches applied so far for InSAR data are supervised and tailored to detect specific deformation events (Anantrasirichai et al., 2018; Sun et al., 2020; Rouet-Leduc et al., 2021). These approaches are either trained on event-specific labelled interferograms (Anantrasirichai et al., 2018) or on synthetic data specifically designed for a

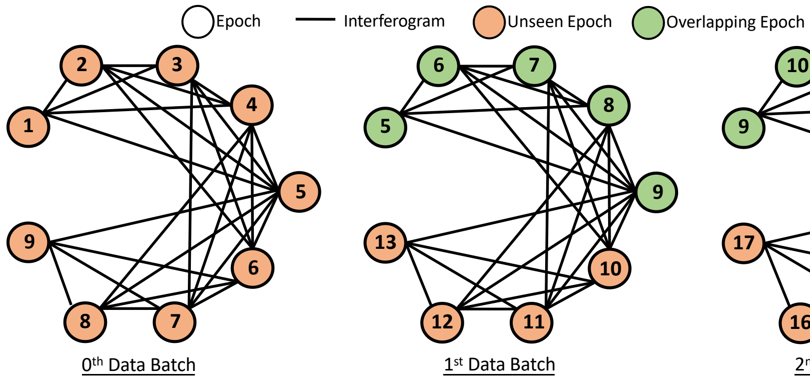

Figure 1. Cartoon of the dataset structure where each line represents an interferogram made by subtracting corresponding epochs (single date images) represented by circles. We consider a combination of 9 epochs (Time t^1 to t^9) that constructs 26 interferograms, to be a single 'data batch'. Moving along in time with a temporal overlap, the next data batch consists of 9 epochs but from t^5 to t^{13} and so on. For illustration purposes, only three consecutive data batches are shown here. The interferograms (input) constructed from the relevant epochs (output) are shown in orange when they are first passed to the model and green when they have been passed twice due to the overlapping nature of input data. The data network constructed within each batch is such that epochs from t^1 to t^5 are connected to the following four epochs but for the last 4 epochs from t^6 to t^9, each epoch connects only with the following available epochs in the data batch. For example at epoch t^8, it is used to construct interferograms with t^9 only, these connections can be visualized by the 'lines' in each data batch. Only one line is going forward from t^8 to t^9.

particular task, for example, to detect volcanic deformation (Anantrasirichai et al., 2019), landslides (Zhang et al., 2022; Chen et al., 2022), anthropogenic signals (Radman et al., 2021; Anantrasirichai et al., 2020) or tectonic deformation (Rouet-Leduc et al., 2021).

Here, we present the development of an alternative approach based on anomaly detection and tailored specifically to the differential structure of InSAR data, where individual images (interferograms) represent the difference in phase between two temporally separated SAR images. InSAR data are very different to the video time-series commonly used to develop machine learning analysis methods, where a single image, on the other hand, records information at a specific moment in time (for example surveillance video time-series (Nawaratne et al., 2019)). A particular challenge presented by InSAR data is the often very low signal-to-noise ratio (SNR), because the contribution of deformation to the phase in an individual interferogram may be an order of magnitude lower than contributions from changes in atmospheric properties. We aim to develop an approach that is event-agnostic, sensitive to both low-rate and transient deformation, and insensitive to errors associated with higher level InSAR processing (e.g., time-series smoothing, fading signal in time series constructed from short-timespan interferograms).

Our previous prototype work (ALADDIn: Autoencoder-LSTM based Anomaly Detector of Deformation in InSAR (Shakeel et al., 2022)), was trained on sequences of unwrapped interferograms from northern Turkey. We use a fully convolutional network (FCN) (Long et al., 2015) that comprises a CNN-LSTM-based encoder and decoder, separated by a neural network. Although the model is capable of detecting deformation as an anomaly, we observed a lack of temporal dependency in some results, and qualitative analysis showed that estimations of deformation varied more than expected for independent estimations of the same epoch. For ALADDIn, a group of 9 epochs (making 26 interferograms, see Figure

1), makes up a single 'data batch', and is treated independently from the next batch. A comparison of epochs that were estimated by sequential batches (green circles in Figure 1) showed that ALADDIn sometimes estimated different spatial patterns of phase for the same epoch due to poor perception of the temporal connection between batches.

In this study, we present important improvements to the ALADDIn approach comprising (1) an improved training method that makes use of the redundancy in interferogram networks to incorporate information about the temporal structure of signals from multiple data batches and (2) the addition of synthetic data for training. We take a transfer learning (Torrey & Shavlik, 2010) approach by re-purposing the pre-trained model from ALADDIn (Shakeel et al., 2022) with our new model for longer interferogram sequences. In addition, we assess the performance of our method using three different scenarios. We evaluate and compare our models on a synthetic test dataset consisting of multiple variations of magnitude and wavelength of an anomaly representing deformation (figure 5). We use the magnitude 5.7 earthquake from southwestern Turkey previously used as a test for ALADDIn (Shakeel et al., 2022) to illustrate the impact of our method improvements (figure 7). We then assess the impact of our choice of 'patch' size on anomalous deformation retrieved while exploring the potential for reproducing long-lived variations in displacement rate using a volcanic test case from Domuyo volcano, Argentina (figure 8).

2 Methodology

We aspire to provide a method for learning from very large, unlabelled InSAR datasets without the need for manual interpretation. The process of labelling interfeorgrams to act as training data is labour intensive, potentially subjective and requires a priori choices about the characteristics of deformation considered interesting. In principle, more diverse data results in more accurate outputs for deep learning methods (Marcus, 2018), but for unlabelled datasets, this relies on the model architectures being intelligent enough to focus on useful information as there is no 'target' or ground truth (set of actual input interferogram as shown in figure 3(b), referred to as 'GT') available. Our solution is to approach the analysis of large, unlabeled InSAR datasets as an anomaly detection problem, where anomalies are any phenomena that differ from the dataset's "normal" spatio-temporal patterns. For InSAR, we consider 'normal' phenomena to arise from any contributions to phase not caused by changes to the Earth's surface. These are generally dominated by atmospheric phase contributions, but may also include errors in estimations of satellite orbitals and 'nuisance' signals associated with processing such as unwrapping errors, e.g, (Emardson et al., 2003; Simons & Rosen, 2007).

Autoencoders (Baldi, 2012) and fully convolutional networks (FCN) (Long et al., 2015), are types of network architectures commonly deployed to perform unsupervised tasks (Bengio et al., 2012). The input and output of such models are identical, so the models learn the underlying distribution of the data and represent them in the form of low-dimensional feature embedding. These embedding act as a bridge between an encoder and a decoder (the main components of an autoencoder), that encrypts and de-crypts useful information about multiple attributes of the data. For the task of anomaly detection, these models are trained on 'normal' data so that they learn the distribution of 'normality' (Gong et al., 2019). After training, when these models are tested on anomalous data, they predict the output with high reconstruction loss as they are unable to accurately reconstruct the anomaly (as they are rare and never seen by the model). Different combinations of layers can be added to these architectures to meet the objectives of the task and to suit the particular data properties. The wide applications of autoencoders for anomaly detection include (Zhao et al., 2017; Gong et al., 2019).

Both ALADDIn (Shakeel et al., 2022) and the developments presented here take an anomaly-detection approach based on the use of autoencoders and thus avoid both time-consuming pixel-wise labelling for training data and are agnostic in terms of deformation detected. By using networks of interferograms as inputs we treat spatial and temporal patterns co-dependently and do not rely on the derivation of time series from interferograms that could introduce further artefacts. We exploit the fact that 'normal' signals associated with individual SAR acquisition dates ('epochs') contribute to related interferograms with a temporal pattern that is quite distinct from deformation, which appears as 'anomalous'. Our model is trained to predict background epoch time-series from the noise in a redundant network of interferograms. Because deformation has a distinct temporal structure, we can therefore separate it from the predicted baseline signals.

2.1 Dataset details

We use Sentinel-1 InSAR data for training and testing our model. Our input data are networks of unwrapped interferograms in radar coordinates generated automatically by the COMET LiCSAR processing system (Lazecký et al., 2020) [1]. This system constructs interferograms with the 4 shortest possible timespans both forwards and backwards from each epoch (as illustrated in the figure 1). For a satellite repeat time of 6 days, this results in each epoch contributing to 8 interferograms (6,12,18,24 days). We re-purpose the model pre-trained for ALADDIn here, using a transfer learning approach. ALADDIn was trained on data from Turkey (Shakeel et al., 2022) and here, we use the same training dataset (LiCSAR Frame name: $014A\_04939\_131313$, data spanning from the year 2017 to 2019) that was first passed through ALADDIn and its predictions are used to initiate the training of our new temporal self-attention model. This dataset was selected on the basis that they were not expected to contain any known examples of deformation, but were dominated by atmospheric signals.

The frame is divided into cubes of size $256 \times 256 \times 26$ pixels, covering an area of $\approx$ $20.5\ km \times 20.5\ km$. This is done in order to manage the model's complexity and the memory needed to train a large number of parameters. Also, instead of passing the entire time series to each training iteration, a set of 26 interferograms (abbreviated to IFGM) that cover 9 epochs (abbreviated to *EP*) are passed, with a 50% spatial overlap (in both N-S and E-W directions). This set is called a data batch and is shown in the figure 1, where circles represent *EP* and the lines connecting each circle i.e. *EP* represents the IFGMs. The 9 *EP* long temporal sliding window moves with a stride of 4, ensuring a $> 50\%$ temporal overlap between subsequent input sequences. Each *EP* in the sequence is connected to all subsequent and preceding *EP* by the 26 IFGM, up to a maximum distance of 4 forward and backward in time. For instance, the central *EP* is linked to all other *EP* by 8 IFGM, but all other *EP* in the data batch are linked with fewer than 8, with the first and last *EP* in the batch being linked with a maximum of 4 IFGM (as illustrated in the figure 1, where only four lines/IFGMs can be seen linking *EP* at t^1 and t^9).

2.2 New network architecture using temporal self-attention

We designed the network architecture of this training system to exploit redundancy in the input interferogram network, that is, that information about each epoch may appear in multiple interferograms (figure 1).Despitef a high detection rate (91.25% overall performance accuracy on a synthetic deformation test case) ALADDIn (Shakeel et al., 2022), produced a different set of solutions for the five *EP* that overlap between data batches (figure 1 green circles represent overlapping *EP*s in the data network). We, therefore, aim to design a system that predicts realistically similar spatio-temporal patterns for *EP*s in the overlap between data batches.

[1] The unwrapped radar-coordinate data format with which the interferograms were saved until the year 2019, is no longer saved. However, the exact data format on which this model is trained can be reconstructed from the LiCSAR intermediate products that are preserved

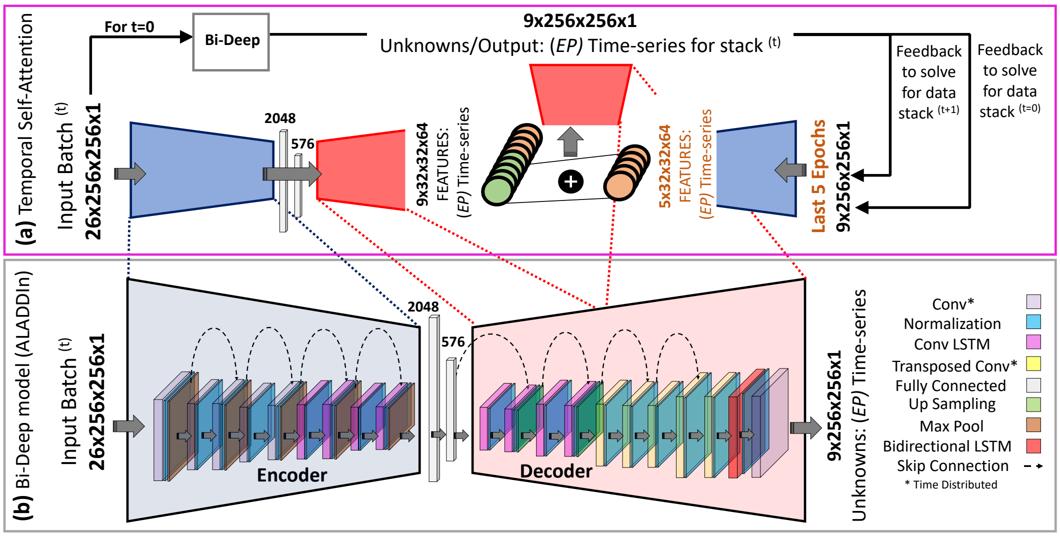

Figure 2. Illustration of network architecture and self-attention training scheme. (b) Bi-Deep model from AlADDIn pipeline (Shakeel et al., 2022) is used for pre-trained weights and to initiate predictions for self-attention. The layers used to create the architecture are: time-distributed 2D convolutional, maxpooling, normalization, 2D convolutional LSTM, fully connected layers, transpose convolutions, upsampling and a bi-directional LSTM layer. Skip connections are also in place to merge features. For temporal self-attention in (a) the decoder is disconnected to fuse features from the overlapping epochs from the previous data batch$^{(t-1)}$. The pre-predictions from the previous data batch$^{(t-1)}$ is passed through a mini-encoder consisting of a pair of time-distributed conv and two convolutional LSTMs

The base of the deep learning model is a fully convolutional network, including an autoencoder combined with a neural network as shown in figure 2(b). We use the same number of layers as used in the encoder and decoder of ALADDIn (Shakeel et al., 2022). In fact, instead of initializing the weights of these layers from scratch, transfer learning (incorporation of previously learned knowledge(Torrey & Shavlik, 2010)) is applied and the previously trained weights are utilized to begin training.

The encoder translates the hidden features/distributions of the input data batch, which is a batch of interferograms (26x256x256) (figure 1). The neural network then converts the features that are learned from IFGMs into the form of *EP*s (9x256x256). This converted feature space is then decoded and interpreted to predict the unknown *EP* time-series. This *EP* time-series should be spatially and temporally consistent (in a sequential manner), regardless of the overlapping nature of the data batches. The features learned for every (*data batch*$^{(t-1)}$) batch should therefore facilitate learning for its proceeding (*data batch*$^{(t)}$), especially for overlapping epochs, as they are already been computed in the previous iteration when (*data batch*$^{(t-1)}$) was processed. This form of attentive learning (a mechanism that focuses on specific temporal regions in a sequence to create a representation of it, for example, here have focused specifically on overlapping *EP*s.) is introduced in the decoder part of the model. The continuity of the decoder is interrupted and predictions of the last 5 epochs (*data batch*$^{(t-1)}$) are fed back by passing them through a mini encoder. These features are combined with the first 5 epochs (*data batch*$^{(t)}$). The merged features are passed to the rest of the decoding layer to make refined epoch predictions. In this way, the cyclic nature of deep learning model training and backpropagation is not affected. Because no previous prediction is available in the case of the very first *data batch*$^{(t=0)}$, the 5 epochs are constrained only by features using the Bi-deep model of ALADDIn (Shakeel et al., 2022).

The model consists of a set of convolutional and convolutional Long Short Term Memory (LSTM) with pooling layers in the encoder. The purpose of pooling layers is to down-sample the input, by taking, for example, a minimum, maximum or average value. In this model, we have used maximum pooling (referred to as maxpool), to gather the maximum value of features within a window size. Likewise, transposed convolutions and convolutional LSTM with upsampling layers are combined in the decoder. In between the encoder and decoder, the neural network consists of three 1D fully connected (FC) layers. The 3D convolution layer spans the input spatially in all dimensions, whereas the Long shot-term memory (LSTM) layer (originally 1D) is capable of maintaining memory with the help of learnable 'forget', 'input' and 'output' gates. This, when combined with the convolutional operation, serves for any multi-dimensional input. This layer is tailored for the specific task of learning both spatial and temporal patterns co-dependently. Every convolution layer is followed by a normalization layer and maxpool for downsampling in the encoder. Similarly, in the decoder, every convolutional LSTM is followed by normalization and upsampling layer. Transposed convolutions (often called deconvolutions) perform similar operations but with broadcasting the feature map instead of downsizing. The mini-encoder used for the attention of overlapping 5 epochs consists of 2 pairs of convolutions, convolutional LSTM following normalization and maxpooling layers.

The neurons used in the fully connected layers are 2048 and 576 (see figure 2). These number of neurons are of immense importance, as they are used for converting the feature maps from 26 interferograms (at the encoder side) to 9 epochs (at the decoder side). For example, the size of data after being processed by the encoder is downsized from $26 \times 256 \times 256 \times 1$ to $26 \times 8 \times 8 \times 1$ then the decoder should receive an input of $9 \times 8 \times 8 \times 1$. Hence, the neurons in the 2nd FC layer are computed by multiplying the dimensions $9 \times 8 \times 8 \times 1 = 576$. So, it can be unrolled back into the dimension $= 9 \times 8 \times 8 \times 1$ and used by the multi-dimensional layers in the decoder. Tanh activation functions are used for every layer. This function ranges from [-1 to 1] which is ideal for this model, as negative values are equally important as positive values.

Furthermore, skip connections are added for feature reusability and to avoid the problem of vanishing gradients (Hochreiter, 1998), where the weights (calculated by each layer in a 'deep' model) gradually decrease to zero and backpropagation fails. In deep convolutional models, this problem often occurs and hinders learning. Skip connections are represented by dotted lines in figure 2(b), their purpose is to re-use the output of layers and feed to deeper layers, to merge the information and process - adding more features helps to dodge the vanishing gradient problem. Finally, a bidirectional convolutional LSTM layer is added, that spans the output both forwards and backwards and combines features to refine the predictions. Two loss functions are used to constrain the model:

$$\begin{aligned} Loss_{IFG} &= \sum_{i=1}^{n} (Output_{IFGM} - Input_{IFGM})^2 \\ Loss_{EP} &= \sum_{i=1}^{n} (5_{epochs^t} - 5_{epochs^{t-1}})^2 \\ Loss &= Loss_{IFG} + Loss_{EP} \end{aligned} \tag{1}$$

where n is the number of interferograms in the case of $Loss_{IFG}$, it refers to the loss computed between reconstructed interferograms by the model and the input interferograms which is also the ground truth. $Loss_{EP}$ is the difference between the current predictions of overlapping epochs and the previous ones, so here n is 5, and the accumulated loss is then backpropagated. With TensorFlow (Abadi et al., 2015) as the backend, the model is trained using Keras (Chollet et al., 2015), a deep learning API. While training, the batch size was set to 1 due to the size of the images in memory being so large. A model can learn features through gradual changes to a loss as opposed to abrupt fluctuations when the learning rate is lower. We therefore selected a learning rate for the Adam optimizer (Kingma & Ba, 2014) of 0.00001.

2.3 Framework for Anomaly Detection

When our model is tested against anomalous data, we expect to find high spatial residuals between our input data and our predicted signals and/or a high overall reconstruction error. Due to the event-agnostic nature of our approach, we do not rely on identifying specific patterns in the reconstruction error. Because we have no preconceptions about the spatial size, intensity and temporal structure of an anomaly, we design a novel detection framework to pick up all kinds of anomalies with a minimum possible rate of false positives. The prediction for every input data batch is refined based on the estimations of findings from previous data batches. We take full advantage of this capability and introduce "*shuttling*" during test time. "*Shuttling*" as its name suggests, completes multiple passes in forward and backward directions for all of the data batches, as shown in figure 3(a), here for illustration purposes the data contains three batches only. During the backward pass, the data is flipped spatial as well because now the pre-EP image serves as the post-EP image to compute the interferogram which is: $post_{EP} - pre_{EP}$. The residuals (RES) shown in figure 3(b) reveal that reconstruction is more accurate after shuttling is implemented and the residuals are near zero after the second forward pass. Input IFGMs that have spatial data gaps or pixels with missing values ('NaN') are a data error and one such IFGM is shown in figure 3(b) black boxes. If 'NaN' is passed through the model, it will propagate through the model due to backpropagation and diminish the learning to a 'NaN'. To avoid this, we identify these missing values and replace them with a zero. This introduces box-like patterns of zeros in the input, but as we are enforcing the model to learn both temporal and spatial patterns. This helps the model to predict even when the input is zero (as shown in the black boxes of the figure 3 (b) 'PRED'). Even though the model makes estimates for missing pixels, these box-like patterns are passed in the 'RES' through subtraction of PRED and GT.

Once the data are shuttled completely (terminated at the third pass, when no improvement results is observed), we expect anomalies to appear as a residue in residual (RES) IFGM as shown in6. An interferogram captures the changes that occurred between two dates, so potentially spans multiple EPs. An anomaly will therefore always show up in several interferograms for our data structure. We first reduce the residuals to a set of N_{EI_r} "residual epoch intervals" that are mutually exclusive in order to precisely detect the temporal window of an anomaly. Since an EP interval is a different image that spans two subsequent EP, so, N_{EI_r} is one less than the total number of EP. As N_{EI_r} are computed from the entire set of residual (RES) interferograms rather than just one residual image, these intervals are more noise-resistant than the shortest spanning set of residual "daisy-chain" interferograms. We execute a linear least squares inversion on a pixel-by-pixel basis of our N_{IFGM_r} residuals in order to estimate this set of N_{EI_r} residual epoch intervals (based on the SBAS approach (Berardino et al., 2002)):

$$d_{IFGM_r} = G.m, \tag{2}$$

where d_{IFGM} is an array of size $N_{IFGM_r} \times 1$ containing all N_{IFGM_r} residual interferograms' pixel values, m is also an array of size $N_{EI_r} \times 1$ containing values of residual EP intervals that we wish to solve for, and G is the design matrix of size $N_{IFGM_r} \times N_{EI_r}$ for this system of equations, which only contains 1s and 0s. For a set of six residual interferograms ($IFGM_{12_r}$, $IFGM_{13_r}$, $IFGM_{14_r}$, $IFGM_{23_r}$, $IFGM_{24_r}$, $IFGM_{34_r}$) that are created from four epochs (EP_{1_r}, EP_{2_r}, EP_{3_r}, EP_{4_r}), matrix G is displayed as an example in Eq 3. This system will output three epoch intervals (EI_{12_r}, EI_{23_r}, EI_{34_r}) based on the residuals.

$$\begin{pmatrix} IFGM_{12_r} \\ IFGM_{13_r} \\ IFGM_{14_r} \\ IFGM_{23_r} \\ IFGM_{24_r} \\ IFGM_{34_r} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} EI_{12_r} \\ EI_{23_r} \\ EI_{34_r} \end{pmatrix} \tag{3}$$

Instead of using all residuals of the overlapping epochs, we only use the latest ones predicted by the model, which should be the most reliable. Also, we perform the linear least square

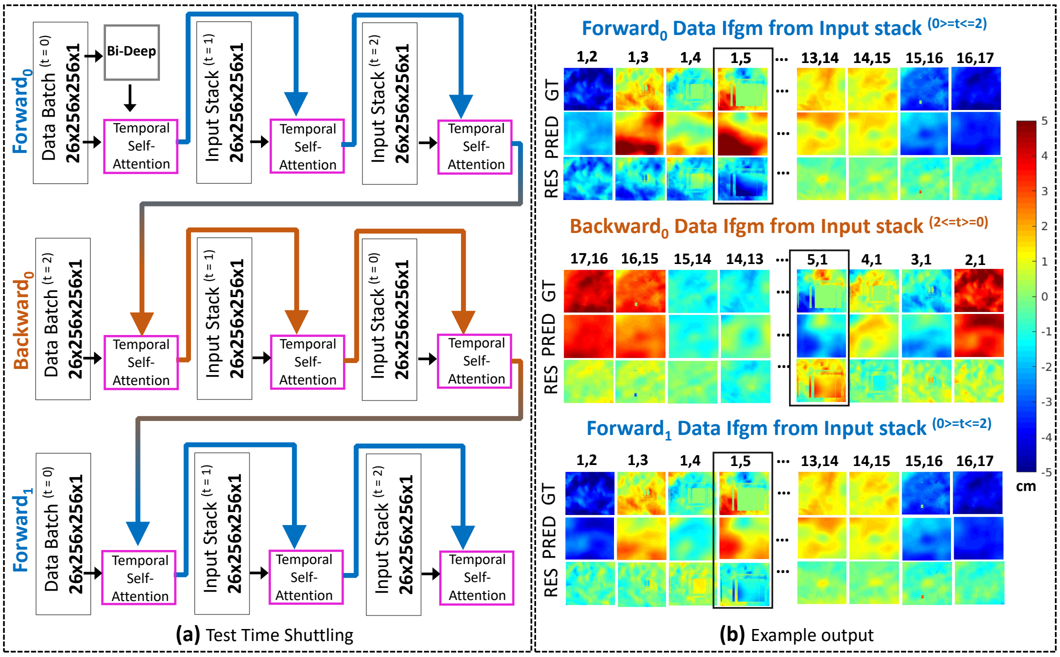

Figure 3. (a) Temporal self attentive training and test-time shuttling procedure is illustrated here. Features of overlapping *EP* (green circles in figure 1) from *data batch*$^{t-1}$ are fed in the model (pink box) for every *batch*^{t} as we move across time from *batch*$^{t=0}$ to *batch*$^{t=2}$. While testing the same procedure is repeated for backward interferograms (both in space and time). The last IFGMs of forward pass (e.g., 15,16 and 16,17) are now the first IFGMs of the backward pass (e.g., 17,16 and 16 and 16,15). This process is called shuttling, it is repeated for another forward pass and so on until no change in the output is observed (only three passes are shown here for illustration purposes). (b) Shows an example output from of the process, where 'GT' refers to ground truth IFGMs which is the input to the model (presented as lines in the figure 1), 'PRED' is predicted IFGMs (output of the model) and 'RES' are the residual IFGMs computed by subtracting PRED from GT to measure what is missed by the model. Shuttling helps to achieve model predictions that are close to the input 'GT'. In comparison of 'RES' of $forward_0$ (top row) and with 'RES' of $forward_1$ (bottom row), it is clear that the 'RES' is near zero. Black boxes enclosing IFGM 1,5 in all shuttling iterations display an example of spatial data gaps or pixels with missing or 'NaN' values, that are replaced with a zero before passing through the model.

inversion for both, forward progressing data and backward progression data, to create two independent sets of residuals *EP* intervals for detection. The presence of spatial anomalies is then automatically identified in these intervals using two complementary analysis techniques: density-based clustering (DBSCAN) (Kriegel et al., 2011) and semivariogram analysis (Wackernagel, 2013).

The residual epoch intervals are predicted to have values close to zero in the absence of anomalous deformation because the model will accurately reconstruct them (e.g. see figure 3(b)), but in the event of an anomaly or multiple anomalies within a sequence, the spatial structure of that anomaly will be visible in at least one epoch residual. Our goal is thus to separate 'normal' intervals from the anomalous ones, without any prior knowledge of where anomalies anomalies appear in a sequence. We use a clustering algorithm (DBSCAN) (Kriegel et al., 2011) that does not require an a-priori specification of the number of clusters to locate anomalies. To ensure we detect all anomalies, we use both, forward and backward independent sets of RES *EP* intervals and perform DBSCAN combining them, establishing two as the minimum number of points in a cluster.

To distinguish between larger areas that are normal and anomalies with specific spatial frequencies (such as deformation that is distributed over small regions), we take into account the spatial variability measured by the semivariogram. By clustering or just computing bulk differences between the actual and reconstructed images (e.g. by a Mean Squared Error), these localised but significant changes are less likely to be found. We anticipate that residual epoch intervals (constructed using residual interferograms i.e. ground truth - prediction) that contain no anomalies will all have a similar spatial structure and, consequently, similar empirical semivariograms, whereas epoch intervals that do contain anomalies will have semivariograms that significantly deviate from this typical structure. For each residual epoch interval a semivariogram is computed, and the root-mean-squared-error between each semivariogram and all others in the complete set of residual epoch intervals across all sequences is calculated. In order to minimise the number of false positives, the classified anomalies from the semivariogram and clustering analysis of epoch interval time series are combined using the AND operation. The variables and parameters used for both these analysis method are same as ALADDIn (Shakeel et al., 2022).

3 Training with Synthetic Data

The anomaly detection model is built by understanding the continuous background atmospheric noise (normality). So when an anomalous event (earthquake, volcano, etc) occurs, the model detects it with high error - as the model fails to understand it due to its anomalous nature. The data used to train our model is real InSAR data from a region of Turkey, that do not contain any anomalous activity, but does contain data errors like unwrapping errors that introduce anomaly like patterns in the IFGMs. Deep learning models have been presented in the past to pick unwrapping errors (Zhou et al., 2021; Sica et al., 2020; Wang et al., 2021), ALADDIn (Shakeel et al., 2022) also detects these as an anomaly. The spatial data gaps and missing values for pixels also introduce artifacts in the data. In an attempt to further improve the detection accuracy of our method, a synthetic training data set is designed based on realistic background atmospheric noise. Synthetic datasets are commonly used to train deep learning models, as employing it overcomes the problem of data imbalance (Anantrasirichai et al., 2019). We test the impact of using synthetic training data on the performance of our model. We expected that the addition of synthetic training data should reduce any impact of unwrapping and other processing-related errors. We construct our synthetic interferograms by first generating synthetic epoch images from which to build them. Our simple synthetic 'normal' (non-deforming) data set is made up from phase (ϕ) contributions from a planar ramp (ϕ_{ramp}, representing residual errors in estimation of satellite orbits), stratified troposphere (ϕ_{strat_atm}) and turbulent troposphere (ϕ_{turb_atm}, described in terms of maximum phase variance, $maxvar$ and characteristic length scale exponent, α) , similar to (Ebmeier, 2016) and described as:

$$\begin{aligned}
\phi_{ramp} &= aX + bY + c \\
\phi_{strat_atm} &= kH \\
\phi_{turb_atm} &= \sqrt{maxvar} * \exp(-r * \alpha) \\
\phi_{EP} &= \phi_{ramp} + \phi_{strat_atm} + \phi_{turb_atm}
\end{aligned} \tag{4}$$

where X and Y are pixel locations and H is the elevation from digital elevation model (DEM). The appropriate parameters are estimated using linear least square inversions (e.g., eq 2), where matrix d_IFG is the interferogram patch, reshaped as 65536×1 array, G is the design matrix (size: 65536×4) containing horizontal pixel location (X), vertical pixel location (Y), elevation (H) and ones (for constant c) and m is the desired output of 4 parameters (a,b,c,k). These values are computed for each patch location for all available time acquisitions of the Turkey data frame used for training. We estimate the parameters ($maxvar$,α) for

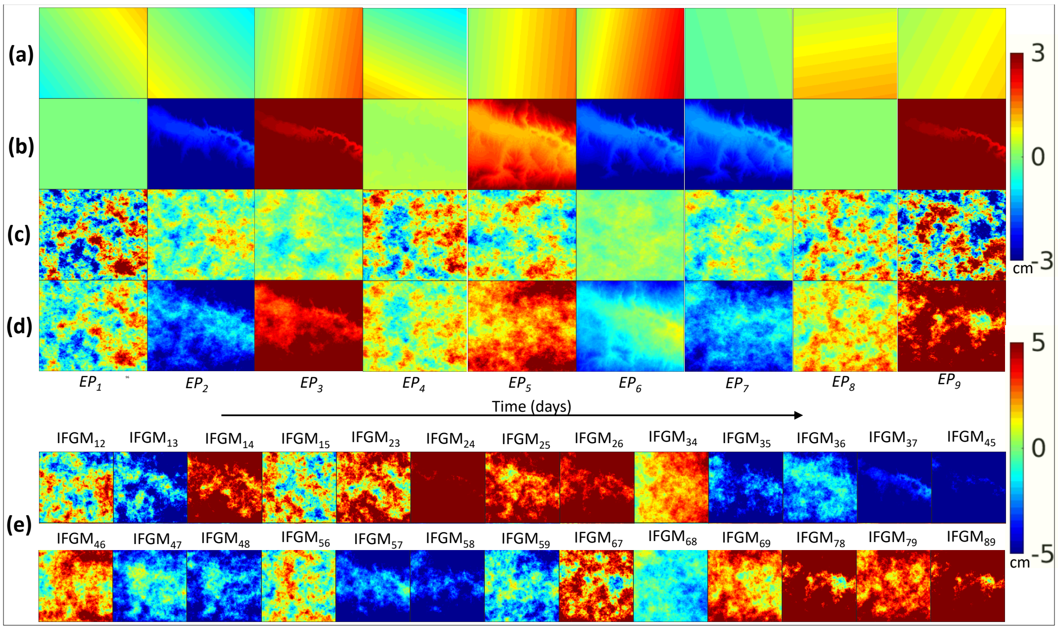

Figure 4. Visualization of the synthetic training data. (a) Planar ramp ($Component_A$), (b) Stratified tropospheric ($Component_B$), (c) turbulent tropospheric ($Component_C$), (d) synthetic 9 epoch generated by aggregating all components (eq 4). (e) 26 synthetic interferograms, following the data structure presented in figure 1 are made using the generated epochs shown in (d). The line-of-sight displacement is measured in cm on a scale of -5 to 5 cm for (d) and (e), where as (a), (b) and (c) are measured on a scale of -3 to 3 cm.

(ϕ_{turb_atm}) using the residual interferogram after removal of $\phi_{strat\_atm} + \phi_{ramp}$:

$$R = d_{IFG} - G.m, \tag{5}$$

To generate the synthetic EP, a, b, k, c, maxvar and α are drawn randomly from a distribution of each of these parameters with mean = 0 and standard_deviation estimated their distribution in the training dataset. Because the variables are computed using interferograms, their sigma values are divided by $\sqrt{2}$ so that it can be used to draw a distribution for EP images.

We use the same network architecture described in Section 2.2, but synthetic data replaces the real data for training of both the Bi-Deep model of ALADDIn and temporal self-attention model. The one main difference when using synthetic training data is in the estimation of the loss function. Previously, we have been using the predicted epochs to reconstruct interferograms to compute loss function. Now, we use the aggregated loss of interferograms and epochs, to take advantage of having synthetic 'ground truth' of EP themselves.

4 Results and Analysis

We evaluate the performance of our improved models on the basis of (1) temporal consistency in overlapping data batches, (2) the range of anomalies detectable in terms of spatial and temporal scale, (3) the models ability to process large areas using sliding spatial and temporal windows and (4) its ability to detect a range of deformation types. We make a quantitative assessment of the detection capabilities of our model using very simplified synthetic deformation (Section 4.1, Figure 5). Our first real-data test case, the 20th March

2019 Southern Turkey magnitude 5.7 earthquake, was also selected to allow us to assess the impact of the developments presented here using a relatively high SNR deformation pattern (Section 4.2, Figure 7 . The second, more challenging test, involves relatively high rate deformation (> 10 cm/yr) at Domuyo volcano, Agentina, that is nevertheless not apparent in individual shorter timespan interferograms due to high magnitude atmospheric noise (Section 4.3, Figure 8).

4.1 Tests on synthetic data

We select which model to apply to real datasets, by using synthetic tests to evaluate the performance of all four model variants: (1) ALADDIn trained on real data (the prototype model presented by (Shakeel et al., 2022), (2) Temporal self attention model trained on real data, (3) ALADDIn trained on synthetic data and (4) Temporal self attention model trained on synthetic data.

We design a synthetic test to compute the accuracy and assess the capacity of our models in terms of the wavelength and magnitude of the deformation signal that can be accurately reconstructed from our model. For simplicity, we use a 2D Gaussian spatial displacement pattern with varying magnitude and wavelength, and add it at eight random instances in a test patch location of Turkey data frame (LiCSAR frame $014A\_04939\_131313$ i.e. the southern section of our training frame reserved for evaluation purposes) that has never been seen by the model during training. The variance of this dataset ranges from 0.1 to 22 with a mean value of 2.4. Noise is dominated by atmospheric phase contributions with typical wavelengths of 10's km. Our synthetic deformation anomaly takes the form:

$$Z(x, y) = A. \exp\left(-(x^2 + y^2)/r\right) \tag{6}$$

where r is the exponential length-scale or wavelength, that varies from 10 m to 12 km. A is the scaling parameter that is directly proportional to the magnitude or peak-value that varies from 1 cm to 11 cm (almost doubling each time to cover maximum range with fewer test variations). The spatial coordinates x and y are relative to the location of Gaussian peak. This means that the SNR for this test dataset range from 0.0003 to 70 (1 cm peak displacement) to 0.0000003 to 1.7 (11 cm peak displacement). The magnitudes shown on the y-axes of Figure 5a reflect peak displacement, which means that the average displacement values are in practise much lower. The detection thresholds indicated by the pink boxes on Figure 5a are therefore very conservative and we are in practise are likely to be able to detect lower average magnitudes of deformation, depending on SNR and interferogram network redundancy (see Section 4.3).

We examine a total of 105 scenarios, each consisting of 10 data batches (spanning from to 28^{th} July 2017 to 18^{th} April 2018), containing 8 synthetic anomalies. The test set is passed through all four models and a comparative analysis is done on a quantitative as well as qualitative basis. Figure 5 (a) shows a heat-plot of true positive rate (TPR also called the Recall). As it is a synthetic dataset, ground truth is known before hand and one-hot encoding (an array of zeros and ones representing whole dataset) of ground truth and predictions are computed, where a label '0' is for normal data and '1' for an anomaly. TPR (Recall) is then plotted for each test sample and for each model. The two network architectures ALADDIn and Temporal self-attention, along with their synthetically trained counterparts, follow similar patterns of performance. Both ALADDIn and Synth-ALADDIn perform weakly with a low TPR (Recall) (<20 % and lower) even for higher magnitudes and wavelengths. The temporal self-attention model and its synthetic counterpart perform better, with higher TPR (Recall) (>60%) for lower magnitudes (2 cm) and higher wavelengths. The mean recall is then plotted in comparison with the mean false positive rate

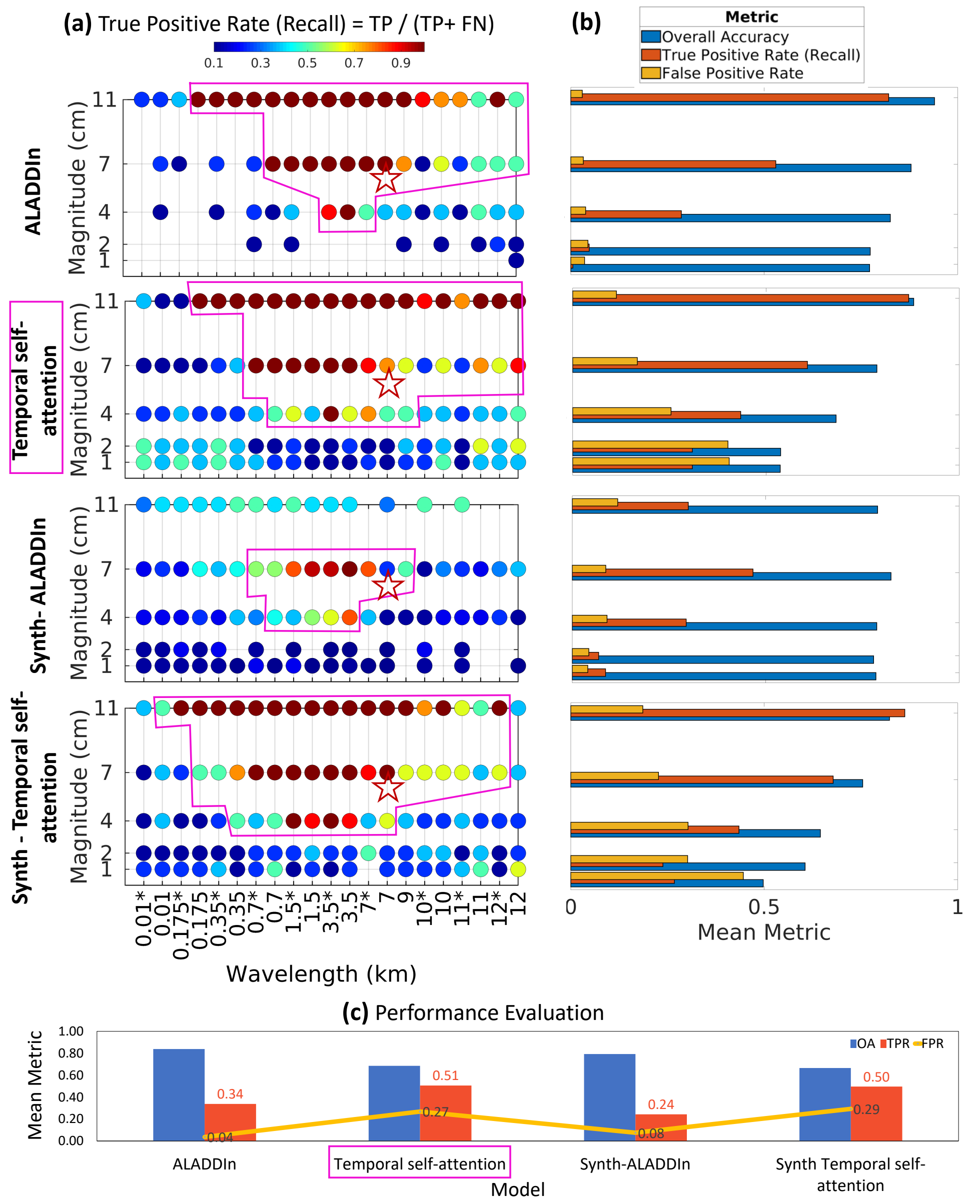

Figure 5. (a) Heat map plot of true positive rates (Recall) for all four model, ALADDIn (Shakeel et al., 2022), Temporal self-attention, synthetic-ALADDIn (trained on synthetic data) and synthetic-Temporal self-attention (trained on synthetic data). SNR ranges from (0.0003 to 70) for 11 cm peak value and (0.0000003 to 1.7) for 1 cm. For 4 cm SNR ranges from (0.00004 to 11). Pink polygon on each of these plots display the region of accurate detection of each model. The x-axis corresponds to wavelengths starting from 10 m to 12 km, '*' represents the same wavelength but with different location (bottom left corner) on the patch. Y-axis corresponds to varying magnitudes from 1 cm to 11 cm. The red star marks the size of the Turkey earthquake shown in Figure X. (b) Bar plots illustrating mean recall (orange), mean false positive rate (yellow) and mean overall accuracy (blue) for all four models. Note that the detectable displacement magnitudes quoted here are peak values in a dataset with a background of variance of 0.6 m, and are therefore very conservative.

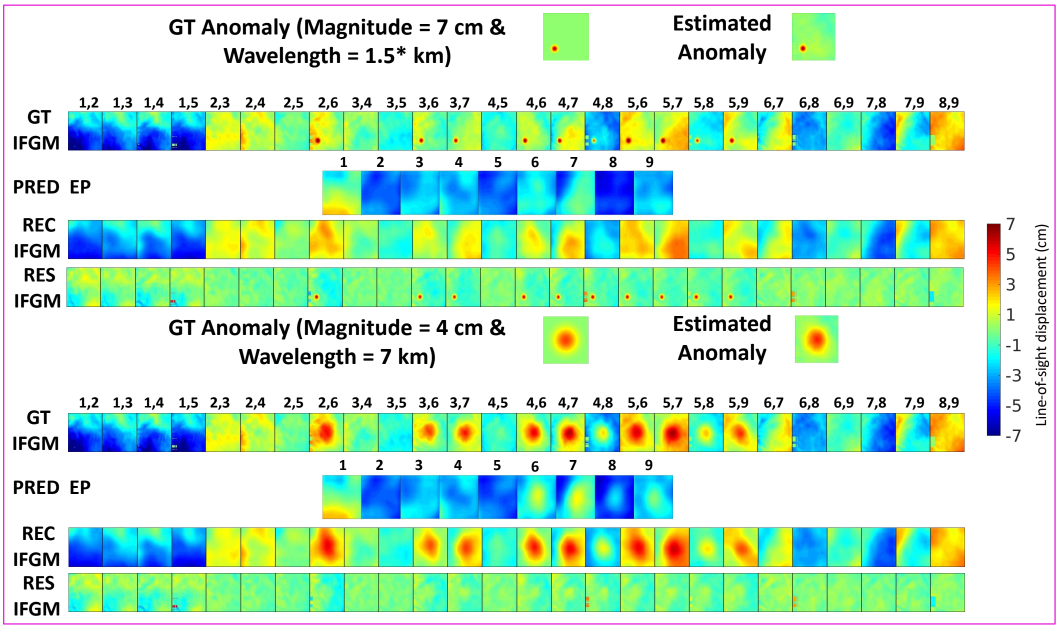

Figure 6. The output of two synthetic data batches for the Temporal self-attention model. The figure shows two examples of synthetic anomalies, (top) magnitude 7 cm and wavelength 1.5 km and (bottom) magnitude 4 cm and wavelength 7km. 'GT IFGM' is ground truth interferogram, 'PRED *EPs*' is the predicted epoch, 'REC IFGM' is the reconstructed interferograms made using PRED *EPs* and 'RES IFGM' is the residual interferogram that carries the anomalous signal missed by the predictions.

(FPR) and mean overall accuracy for each magnitude, as follows::

$$\begin{aligned} TPR(Recall) &= TP/(TP+FN) \\ FPR &= FP/(FP+TN) \\ OA &= (TP+TN)/(TP+TN+FP+FN) \end{aligned} \tag{7}$$

where TP is true positive, FP is false positive, TN is true negative and FN is false negative. Negative here corresponds to the 'normal' data or 0's and positive corresponds to anomalies or 1's. OA is the overall accuracy, TPR and FPR are true and false positive rates respectively. The plots in figure 5 (c) shows that average overall accuracy of each model is greater than 70%, dominated by a high specificity (high TN rate). The FPR increases as we move across the models, largely because the model is fitting to the 'normality' it learned from the training data, leaving greater residuals and resulting in false flags. These FPs includes unwrapping errors and missing data as well as signals due to deformation. The models that do a better job of predicting 'normal' interferogram patterns therefore flag more of the errors in the input data, which are classified as anomalies according to our tests. The temporal self-attention model has proven to be the most accurate one when compared to its synthetic counterpart because the former's FPR is lower (as shown in figure 5 (b)). Figure 6 displays the results of temporal self-attention for two synthetic test scenarios both for low magnitude and greater wavelength (bottom) and for higher magnitude and lower wavelength (top). The model accurately estimates the spatial structure of flagged anomaly as shown in figure 6. The capacity of each model - that is the range of anomaly wavelengths and magnitudes it is capable of flagging - is indicated by the pink polygons on figure 5(a)). The area of polygon is much greater for Temporal Self-attention and its synthetic counterpart than the original ALADDIn model.

Table 1. Fbeta Measure

Models	F1	F0.5	F2
ALADDIn	0.36	**0.39**	0.34
Temporal self-attention	**0.41**	0.37	**0.45**
Synth-ALADDIn	0.26	0.30	0.24
Synth Temporal self-attention	0.37	0.33	0.43

Similar trends can be seen in Table 1, where Fbeta score is considered to balance the affect of minimizing FP (Precision) or minimizing FN (maximizing TP) (Recall). These two factors are combined with a beta values of (1, 0.5 and 2) to compute the Fbeta scores. It is calculated using eq 8. When beta is 1, where recall and precision are given equal importance, Temporal self-attention has the maximum score of 0.41. But when beta is decreased to 0.5, this mean that the measure is focusing more towards minimizing FP then the score of ALADDIn is greatest with 0.39. Whereas, when beta is 2, that is giving more weight to the Recall then again Temporal self-attention has the maximum score of 0.45. The analysis proves that the Temporal self-attention model is best as it outperforms all others models in terms of balancing a high recall. We emphasise the recall in our analysis, because we do not want the model to miss any anomalies. While FP can always be reviewed by human intervention, we prioritise not missing an event that might be anomalous.

$$Fbeta = \frac{((1 + beta^2) \times Precision \times Recall)}{(beta^2 \times Precision + Recall)}, \tag{8}$$

4.2 Real case study (I): 2019 Turkey earthquake

We use a 5.7-magnitude earthquake that took place in southern Turkey on 20^{th} March 2019 (Elliott et al., 2020), and was previously used to test ALADDIn (Shakeel et al., 2022) to evaluate the accuracy of our models. This south-western region of Turkey has experienced major earthquakes in the past (M_w 7.0 in 1914 (Ambraseys, 1988), M_w 6.2 in 1971 (Taymaz & Price, 1992), M_w 6.2 in 1995 (Wright et al., 1999) and M_w 6.6 in year 2017 (Karasözen et al., 2018)). InSAR data that has been analysed by our model, estimates deformation of approximately 4 cm (as reported by (Elliott et al., 2020)), shown in figure 7 (c). The data for this test case is processed from the time-period 18^{th} September 2018 to 10^{th} April 2019. We divide the data into 7 data batches (according to figure 1), comprising 32 epoch intervals in total. This test region is never seen by any of the models during training. The method ALADDIn successfully detected this earthquake and estimated its spatial structure, but due to the overlapping data structure, two variations of the same instantaneous anomaly were retrieved, as shown in figure 7 (a). It can be seen that the Temporal self-attention model constructs one, more accurate, estimate (the spatial structure in the EI_r closely matches with the GT structure captured in the IFGMs) of epoch-interval whereas the two variations estimated by ALADDIn are different in terms of spatial pattern (shown in figure 7 (a)). On comparison of residual intervals EI_r (figure 7 (a)) with the estimated earthquake structure (figure 7 (d)), it can be seen that the proposed method produces an anomaly pattern in the residuals closer to actual structure (as reported by (Elliott et al., 2020)). The 10 anomalous interferograms (shortest 6-day to longest 24-day) spanning the earthquake are all flagged by the model. These are marked red in the plot shown in figure 7 (c) as compared to the rest 'black' (normal) interferograms. The estimated spatial structure of the flagged anomaly by the Temporal Self-attention model is shown in figure 7 (e).

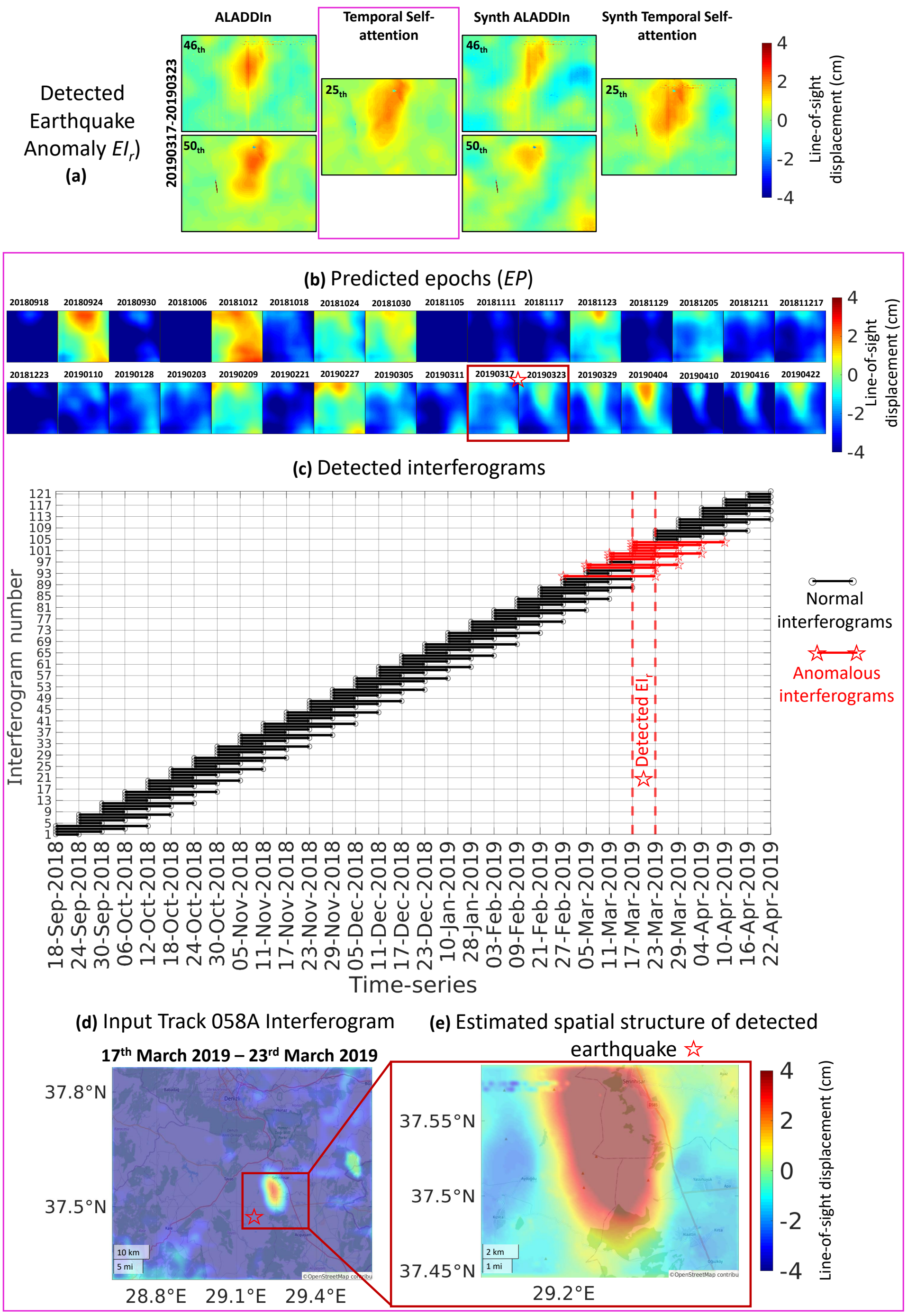

Figure 7. Test results of real earthquake of all four models are shown in this figure. (a) shows the detected residual based epoch intervals (EI_r) of all four models. (b) Shows the predicted epoch (EPs) for the Temporal Self-attention model. The red box enclose the 2 EPs that cover earthquake date (20.03.2019), represented by a red star. (c) Shows the network of interferogram used for testing with detected interferograms marked red, covering the detected earthquake interval (marked with red dotted line). (d) Shows the shortest interferogram capturing the earthquake anomaly. (e) Shows the zoomed in patch of estimated spatial structure of the detected anomaly.

4.3 Real case study (II): Domuyo Volcano, Argentina

We apply our most accurate model - Temporal Self-attention trained on real data - to an additional real test case. We select a well-documented period of uplift at Domuyo volcano, Argentina (Lundgren et al., 2020; Astort et al., 2019; Derauw et al., 2020), because this allows us to examine (1) how our model trained on real InSAR data from northern Turkey performs in a location with completely different atmospheric conditions and topography and (2) how well we can detect persistent rather than transient deformation, and (3) how our model performs when the deformation signal (in this case - 64×40 km) exceeds our patch size (20.5×20.5 km).

Domuyo stratovolcano (4702 m elevation), in northern Patagonia, is thought to be late Pleistocene (but possibly Holocene) age. It has no record of historical activity, but a major hydrothermal field centred southwest of the volcano's flanks has a very high thermal energy release and recent gas-driven explosions, which imply the presence of an active magmatic system (Chiodini et al., 2014; Lundgren et al., 2020). Further evidence for this comes from uplift, which has been attributed to the intrusion of volatile-rich magma at 6.5-7 km depth (Astort et al., 2019), and has occurred in lagged correlation with edifice-wide warming (Lundgren et al., 2020). Domuyo subsided between 2008 and approximately 2013, before entering a phase of uplift in 2014 with a maximum rate of 15 cm/yr. Uplift slowed until early 2021, when the volcano began to subside. We selected a period of relatively high-rate uplift between May 2017 and December 2018 for our method test (Figure 8a). Over this interval deformation was relatively constant, so we expect every interferogram to be flagged.

This case study also provides the opportunity to assess the implications of applying deep learning to automatically processed, noisy InSAR data sets with significant of data gaps. The network design for standard LiCSAR processing relies on short timespan interferograms: 4 forward connections for each epoch, maximum interferogram length of 48 days for 12 day acquisition intervals, as at Domuyo. This means that even for the relatively high rate persistent deformation at Domuyo, displacements in individual interferograms are commonly < 1 cm, well below the level of atmospheric contributions (Figure 8b). Furthermore, the standard LiCSAR network design is not optimised for regions with major seasonal variations in phase coherence (e.g., snow cover). This results in loss of coherence in our test dataset, relative to a network tailored to include only summer-summer interferograms (Lundgren et al., 2020).

The interferograms we analysed also had a minimum 12 day interval as compared to our training data which had a minimum of 6 day gap. Where there were large gaps between use-able interferograms in the automatically processed data, epochs were skipped. The sample interferograms shown in Figure 8 illustrate typical data gaps (circle), low coherence with unwrapping errors (square), low SNR (inverted triangle and square) and an example of a high SNR image where the Domuyo displacements are visible (triangle). Although our training dataset did include regions of poor coherence, it did not include data gaps as shown in Figure 8a (circle), so these are flagged as anomalies, leading to a high FPR.

Our method successfully identifies all 170 interferograms as containing anomalies (all 44 epochs flagged) in our test dataset - as expected for a steady displacement signal. It performs well both for high and low SNR interferograms (Figure 8e), although large data gaps and unwrapping errors result in false positives. These results demonstrate the transferability of training using data from Northern Turkey to a completely different geographic setting, with very different topography, vegetation, patterns and therefore interferogram noise. Our estimation of the accumulated deformation from our automatic detection is very consistent with displacements estimated using conventional analysis methods such as 'stacking' (compare Figure 8 b and d).

These results also demonstrate how our method can be used to process larger regions using a 'sliding window' approach, and stitching the results together to reconstruct the

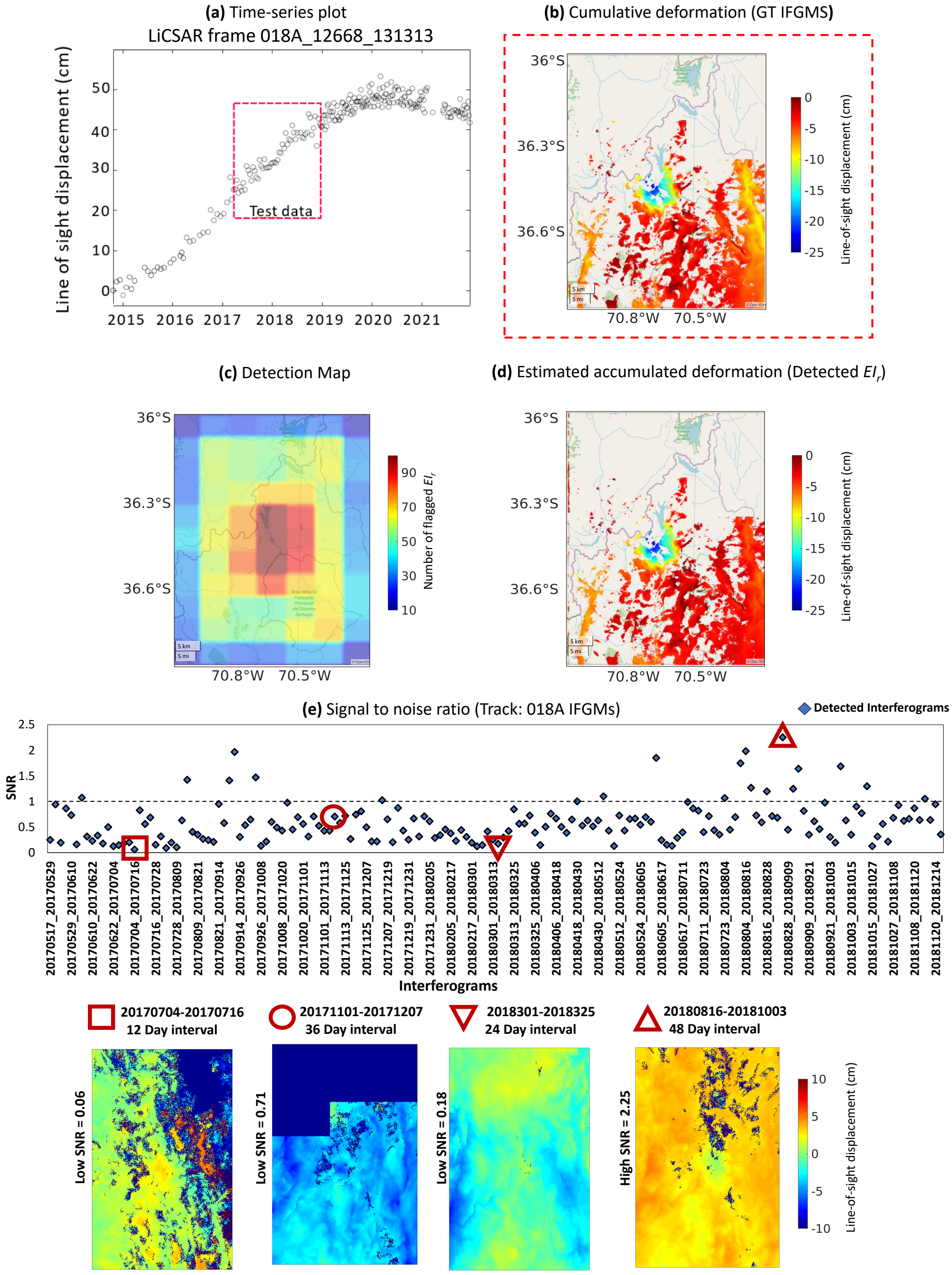

Figure 8. (a) Time series of displacements at Domuyo Volcano, made from LiCSBAS time series algorithm (Morishita et al., 2020) using LiCSAR interferograms (b) Cumulative deformation estimated from a stack of the dispalcement rates for all GT IFGMs, assuming a linear displacement rate. (c) Detection map showing flagged EI_r on a patch by patch basis (spatial and temporal sliding window for processing). (d) Our final estimation of the spatial structure of deformation of all detected epoch intervals. (e) (Top) SNR, as estimated as the ratio of peak displacement to interferogram variance with deforming area masked, for the processed region of interest for all interferograms. (Bottom) Interferograms with lower SNR are displayed in comparison to the one with higher SNR (red triangle). All of which are detected by the model.

deformation pattern. We extract the southwest corner of the automatically processed interferograms centred on Domuyo and with dimensions of $\sim 13000km \times 11000km$ (1048×896 pixels). This region is further divided into 60 overlapping patches of size 256×256 and each patch is independently passed through the model and detection framework. The data spans from 17^{th} May 2017 to 14^{th} December 2018 ($\sim$ 1.5 years), which results in 10 data batches with 44 epoch intervals for each patch location. All the flagged epoch intervals of each patch are visualized on a heat map (detection map shown in figure 8 (c)). While the total number of epoch intervals is 44 the overlapping nature of patches means that each pixel location is covered by at least 4 patches, except the boundaries, hence epoch intervals EI_r are flagged by multiple patches. Patch overlap also mitigates the impact of data gaps and low coherence in the area of interest. A key benefit of using a CNN is translation invariance, which has proved to be beneficial in this case and has resulted in spatially consistent output, despite the fact that the whole area is processed on patch by patch basis in a sliding window scheme and stitched together at the end. This property of our CNN is comes from the spatial overlap in our training dataset. Our final prediction of total displacement spanning May 2017 to December 2018 is a union of all the detections for all patches (Figure 8d).

5 Discussion

While Deep Learning methods cannot replace detailed analysis of specific deformation events (including tailored InSAR processing, atmospheric correction and time series analysis) flagging of anomalies in very large datasets has the potential to be a powerful tool for finding new tectonically significant signals. Traditional methods for analysing InSAR mitigate the impact of low SNR in individual interferograms by methods including stacking (Pritchard & Simons, 2004), construction of time series (Lundgren et al., 2001), (Rouet-Leduc et al., 2021), filtering (Dalaison & Jolivet, 2020), etc. Some of these methods, especially when used in combination, are capable of detecting relatively low magnitude deformation (e.g., <10 mm/yr). However, they commonly rely on deformation signals being either persistent, high magnitude relative to noise, or having an a priori idea of signal pattern (e.g., parametric fitting for expected co-seismic or inter-seismic deformation patterns). An advantage of the anomaly detection methodology proposed here, is that it is similarly successful in the detection of transient events that occur in a single epoch (Section 4.2) and long term, persistent signals (Section 4.3). Here, we discuss (1) the performance of our iterative, temporal self attention scheme and (2) suggest directions for the future developments for anomaly detection in InSAR and (3) describe the potential of our method for application to global datasets.

5.1 Performance of new network architecture

Our new deep learning architecture and iterative training scheme for our model presents significant improvements to the protoype model in our previous work (ALADDIn (Shakeel et al., 2022)). The requirement of temporal consistency in particular improved our ability to accurately reconstruct the spatial structure of flagged displacement signals as assessed by comparison to both individual interferograms (Figure 7d) and stacked data (e.g., Figure 8b). The ability both to flag and reconstruct the spatial structure of deformation as part of a single pipeline is an additional advantage of our anomaly detection process (Section 2.3).

The flagging of deformation and reconstruction of its spatial structure is highly dependent on the quality of input data. For example, data gaps and unwrapping errors resulted in many false positive detections in the Domuyo dataset, as can be seen in the non-zero values for detection in the areas surrounding the Domuyo displacement signal (Figure 8c). This makes the overall accuracy for our method in this location lower than it was in Turkey. However, this is mitigated by the redundancy introduced by our 'sliding' patches (Section 5.2). SNR, estimated as ratio of peak displacement over the volcano to background variance with volcanic areas masked,is also relatively low at Domuyo (Figure 8e). More than 90 %

of the input interferograms have low SNR (as highlighted by a dashed line on Figure 8e), but the model successfully detects the anomaly even with SNR as low as 0.06.

5.2 Potential for global deformation flagging

Although our model is trained only on an ascending InSAR dataset from northern Turkey, it is successful in flagging both the 2019 southern Turkey earthquake and the Domuyo uplift in Argentina. This demonstrates that our approach does not necessarily require tailored training data sets for settings with different atmospheric noise, although it is possible that this might improve model performance. Our anomaly detection approach to InSAR data exploits the redundancy in networks of interferograms to learn the relationship between interferograms and the signals attributable to each epoch (differential data structure), and is therefore not sensitive to the specific spatial and temporal properties of the background noise (atmosphere and errors in orbit estimations).

As shown in Section 4.3, the relatively small patch size that we use in our model is not an obstacle to detection of much longer wavelength deformation. In fact, using the sliding window technique for processing large areas has proven to be advantageous because anomalous intervals can be flagged by multiple patch locations (see Figure 8), or if missed in one patch, can be flagged in others. This could be improved for processing large areas of data by the development of a voting ensemble (machine learning model that combines the predictions from multiple models) instead of taking a simple union for gathering the epoch intervals.

The types of deformation detectable with our method are also strongly dependent on the resolution of the input dataset. LiCSAR processing is optimised for scientific study of very long wavelength, low magnitude interseismic signals, using multi-looking and some filtering to clean up the interferograms at the expense of spatial resolution (Morishita et al., 2020). They are therefore less appropriate for the detection of small-scale displacements such as anthropogenic or some volcanic signals.

5.3 Directions for new development

We found that the addition of synthetic training datasets did not significantly improve the performance of our model. Initially, it was challenging to train similar network architectures on synthetic data. The problem of model over-fitting very early in training time hindered its learning. The network architectures are designed to cater for raw unwrapped interferograms, the phase values of which varied across a wide range (for example from -300 rad to 15 rad). However, the range of values in our synthetic dataset are smaller, and not diverse as compared to raw data. This was a consequence of the estimation of parameters using least square inversion and residual interferograms as explained in Section 3. In addition, noise was not separately simulated and added in the data, resulting in lower data diversity. A better approach may be to incorporate data augmentation techniques (Taylor & Nitschke, 2018), instead of solely depending on synthetic training data. This would increase the number of data samples and also add diversity to the training set. For example, descending frames could be augmented (as it also a variant of data that captures similar region but with different look angle) or synthetic data could be augmented with real data to create one training set.

The key to a diverse and well-fitted model is generally hidden in its training data. A diverse dataset can lead to a model that could be re-used to be applied on wrapped or geocoded data through transfer learning. Once fully trained, the weights of the model can be re-used to initiate training on various data types. For example, we incorporated transfer learning and utilized the weights of ALADDIn (Shakeel et al., 2022) to train a Temporal self-attention model. Similarly, it could be fine-tuned on different data types to create a generalized model. Also, while preparing the data, introducing regions or temporal stamps with extreme atmospheric conditions would be helpful and reduce false positives. For example, it could be advantageous to use interferograms with winter images or snow in them or including regions near sea, river or lake.

The model developed in this study can be used further to build weakly supervised methods (Campanella et al., 2019). Such methods incorporate constrained and indirect sources of supervision to devise labels for large volumes of unlabelled data. The residuals created by the model predictions can be used as weak pixel-wise labels/masks for anomalies. This can lead to supervised anomalous instance segmentation by creating a global labelled dataset.

6 Conclusions

This work addresses two limitations of our previous prototype anomaly-detector for InSAR data: (1) temporal inconsistency and (2) limited patch size. The development of new architecture that incorporates temporal self-attention improves the performance of our model so that it flags deformation of peak magnitude of a few cm and wavelength of a few hundred metres with an overall accuracy of 80-90 %. We demonstrate that our model, although trained on data in northern Turkey is successful in flagging displacements at an Argentinian volcano, and that it is capable of detecting both transient (earthquake) and persistent (volcano) deformation. We have shown that the architecture developed here has important potential for anomaly detection, and believe that fruitful future developments to improve its performance could include development of more diverse, realistic training data, and testing against a wider variety of volcano-tectonic and other deformation sources.

7 Acknowledgments

The Looking into the Continents from Space (LiCS) database and the Centre for Observation and Modelling of Earthquakes, Volcanoes and Tectonics (COMET) were used to build the dataset for this project. AS's PhD was funded by Durham's GCRF Global Challenges Research Fund. SKE is funded by a NERC Independent Research Fellowship (NE/R015546/1). We Thank Andrew Watson and Dr Milan Lazecky for providing immense support regarding the data used in this study. We would like to express our gratitude to Professor Mark Allen for reviewing this study and for his insightful geological commentary.

8 Open Research

Sentinel-1 interferograms are used for training and evaluations of the models. Unwrapped interferograms, including all of those used in this study, were processed automatically and can be downloaded from the open access LiCSAR hub (Lazecký et al., 2020) `https://comet.nerc.ac.uk/comet-lics-portal/` . The deep learning models (ALADDIn and Temporal self-attention (Anza Shakeel, 2022)) developed and tested in this study are available at GitHub: `https://github.com/AnzaShakeel/Deep-Learning-for-InSAR.git` via DOI: `https://doi.org/10.5281/zenodo.7326911` .The machine learning platforms used to develop the models presented here are Keras (Chollet et al., 2015) with TensorFlow (Abadi et al., 2015) as backend.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems* [Software]. Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org)

Ambraseys, N. N. (1988). Engineering seismology: part ii. *Earthquake engineering & structural dynamics*, *17*(1), 51–105.

Anantrasirichai, N., Biggs, J., Albino, F., & Bull, D. (2019). A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets. *Remote Sensing of Environment*, *230*, 111179.

Anantrasirichai, N., Biggs, J., Albino, F., Hill, P., & Bull, D. (2018). Application of machine learning to classification of volcanic deformation in routinely generated insar data. *Journal of Geophysical Research: Solid Earth*, *123*(8), 6592–6606.

Anantrasirichai, N., Biggs, J., Kelevitz, K., Sadeghi, Z., Wright, T., Thompson, J., … Bull, D. (2020). Detecting ground deformation in the built environment using sparse satellite insar data with a convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(4), 2940–2950.

Anza Shakeel. (2022). *Anzashakeel/deep-learning-for-insar: Temporal self-attention* [Software]. Zenodo. Retrieved from `https://zenodo.org/record/7326911` doi: 10.5281/ZENODO.7326911

Astort, A., Walter, T. R., Ruiz, F., Sagripanti, L., Nacif, A., Acosta, G., & Folguera, A. (2019). Unrest at domuyo volcano, argentina, detected by geophysical and geodetic data and morphometric analysis. *Remote Sensing*, *11*(18), 2175.

Baldi, P. (2012). Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of icml workshop on unsupervised and transfer learning* (pp. 37–49).

Bekaert, D. P., Handwerger, A. L., Agram, P., & Kirschbaum, D. B. (2020). Insar-based detection method for mapping and monitoring slow-moving landslides in remote regions with steep and mountainous terrain: An application to nepal. *Remote Sensing of Environment*, *249*, 111983.

Bengio, Y., Courville, A. C., & Vincent, P. (2012). Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR, abs/1206.5538*, *1*(2665), 2012.

Berardino, P., Fornaro, G., Lanari, R., & Sansosti, E. (2002). A new algorithm for surface deformation monitoring based on small baseline differential sar interferograms. *IEEE Transactions on geoscience and remote sensing*, *40*(11), 2375–2383.

Campanella, G., Hanna, M. G., Geneslaw, L., Miraflor, A., Werneck Krauss Silva, V., Busam, K. J., … Fuchs, T. J. (2019). Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nature medicine*, *25*(8), 1301–1309.

Chen, X., Yao, X., Zhou, Z., Liu, Y., Yao, C., & Ren, K. (2022). Drs-unet: A deep semantic segmentation network for the recognition of active landslides from insar imagery in the three rivers region of the qinghai–tibet plateau. *Remote Sensing*, *14*(8), 1848.

Chiodini, G., Liccioli, C., Vaselli, O., Calabrese, S., Tassi, F., Caliro, S., … D'alessandro, W. (2014). The domuyo volcanic system: an enormous geothermal resource in argentine patagonia. *Journal of volcanology and geothermal research*, *274*, 71–77.

Chollet, F., et al. (2015). *Keras* [Software]. `https://keras.io`.

Dalaison, M., & Jolivet, R. (2020). A kalman filter time series analysis method for insar. *Journal of Geophysical Research: Solid Earth*, *125*(7), e2019JB019150.

Derauw, D., Jaspard, M., Caselli, A., Samsonov, S., et al. (2020). Ongoing automated ground deformation monitoring of domuyo-laguna del maule area (argentina) using sentinel-1 msbas time series: Methodology description and first observations for the period 2015–2020. *Journal of South American Earth Sciences*, *104*, 102850.

Ebmeier, S. (2016). Application of independent component analysis to multitemporal insar data with volcanic case studies. *Journal of Geophysical Research: Solid Earth*, *121*(12), 8970–8986.

Ebmeier, S., Andrews, B., Araya, M., Arnold, D., Biggs, J., Cooper, C., … others (2018). Synthesis of global satellite observations of magmatic and volcanic deformation: implications for volcano monitoring & the lateral extent of magmatic domains. *Journal of Applied Volcanology*, *7*(1), 1–26.

Elliott, J., de Michele, M., & Gupta, H. (2020). Earth observation for crustal tectonics and earthquake hazards. *Surveys in Geophysics*, *41*(6), 1355–1389.

Emardson, T., Simons, M., & Webb, F. (2003). Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation. *Journal* *of Geophysical Research: Solid Earth*, *108*(B5).

Gaddes, M., Hooper, A., & Bagnardi, M. (2019). Using machine learning to automatically detect volcanic unrest in a time series of interferograms. *Journal of Geophysical Research: Solid Earth*, *124*(11), 12304–12322.

Gong, D., Liu, L., Le, V., Saha, B., Mansour, M. R., Venkatesh, S., & Hengel, A. v. d. (2019). Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 1705–1714).

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(02), 107–116.

Karasözen, E., Nissen, E., Büyükakpınar, P., Cambaz, M. D., Kahraman, M., Kalkan Ertan, E., ... Özacar, A. A. (2018). The 2017 july 20 m w 6.6 bodrum–kos earthquake illuminates active faulting in the gulf of gökova, sw turkey. *Geophysical Journal International*, *214*(1), 185–199.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kriegel, H.-P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(3), 231–240.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Lazecký, M., Spaans, K., González, P. J., Maghsoudi, Y., Morishita, Y., Albino, F., ... Wright, T. J. (2020). Licsar: An automatic insar tool for measuring and monitoring tectonic and volcanic activity [Dataset]. *Remote Sensing*, *12*(15). Retrieved from `https://www.mdpi.com/2072-4292/12/15/2430` doi: 10.3390/rs12152430

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 3431–3440).

Lundgren, P., Girona, T., Bato, M. G., Realmuto, V. J., Samsonov, S., Cardona, C., ... Aivazis, M. (2020). The dynamics of large silicic systems from satellite remote sensing observations: the intriguing case of domuyo volcano, argentina. *Scientific reports*, *10*(1), 1–15.

Lundgren, P., Usai, S., Sansosti, E., Lanari, R., Tesauro, M., Fornaro, G., & Berardino, P. (2001). Modeling surface deformation observed with synthetic aperture radar interferometry at campi flegrei caldera (paper 2001jb000194). *JOURNAL OF GEOPHYSICAL RESEARCH-ALL SERIES-*, *106*(9; SECT 2), 19–355.

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

Morishita, Y., Lazecky, M., Wright, T. J., Weiss, J. R., Elliott, J. R., & Hooper, A. (2020). Licsbas: An open-source insar time series analysis package integrated with the licsar automated sentinel-1 insar processor. *Remote Sensing*, *12*(3), 424.

Nawaratne, R., Alahakoon, D., De Silva, D., & Yu, X. (2019). Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Transactions on Industrial Informatics*, *16*(1), 393–402.

Pritchard, M., & Simons, M. (2004). An insar-based survey of volcanic deformation in the central andes. *Geochemistry, Geophysics, Geosystems*, *5*(2).

Radman, A., Akhoondzadeh, M., & Hosseiny, B. (2021). Integrating insar and deep-learning for modeling and predicting subsidence over the adjacent area of lake urmia, iran. *GIScience & Remote Sensing*, *58*(8), 1413–1433.

Ren, X., Li, X., Ren, K., Song, J., Xu, Z., Deng, K., & Wang, X. (2021). Deep learning-based weather prediction: a survey. *Big Data Research*, *23*, 100178.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241).

Rouet-Leduc, B., Jolivet, R., Dalaison, M., Johnson, P. A., & Hulbert, C. (2021). Autonomous extraction of millimeter-scale deformation in insar time series using deep learning. *Nature communications*, *12*(1), 1–11.

Semple, A. G., Pritchard, M. E., & Lohman, R. B. (2017). An incomplete inventory of suspected human-induced surface deformation in north america detected by satellite interferometric synthetic-aperture radar. *Remote Sensing*, *9*(12), 1296.

Shakeel, A., Sultani, W., & Ali, M. (2019). Deep built-structure counting in satellite imagery using attention based re-weighting. *ISPRS journal of photogrammetry and remote sensing*, *151*, 313–321.

Shakeel, A., Walters, R. J., Ebmeier, S. K., & Al Moubayed, N. (2022). Aladdin: Autoencoder-lstm-based anomaly detector of deformation in insar. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, 1–12.

Sharma, M., Dhanaraj, M., Karnam, S., Chachlakis, D. G., Ptucha, R., Markopoulos, P. P., & Saber, E. (2020). Yolors: Object detection in multimodal remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 1497–1508.

Sica, F., Calvanese, F., Scarpa, G., & Rizzoli, P. (2020). A cnn-based coherence-driven approach for insar phase unwrapping. *IEEE Geoscience and Remote Sensing Letters*.

Simons, M., & Rosen, P. (2007). Interferometric synthetic aperture radar geodesy.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Sun, J., Wauthier, C., Stephens, K., Gervais, M., Cervone, G., La Femina, P., & Higgins, M. (2020). Automatic detection of volcanic surface deformation using deep learning. *Journal of Geophysical Research: Solid Earth*, *125*(9), e2020JB019840.

Taylor, L., & Nitschke, G. (2018). Improving deep learning with generic data augmentation. In *2018 ieee symposium series on computational intelligence (ssci)* (pp. 1542–1547).

Taymaz, T., & Price, S. (1992). The 1971 may 12 burdur earthquake sequence, sw turkey: a synthesis of seismological and geological observations. *Geophysical Journal International*, *108*(2), 589–603.

Torrey, L., & Shavlik, J. (2010). Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* (pp. 242–264). IGI global.

Wackernagel, H. (2013). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media.

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

Wright, T., Parsons, B., Jackson, J., Haynes, M., Fielding, E., England, P., & Clarke, P. (1999). Source parameters of the 1 october 1995 dinar (turkey) earthquake from sar interferometry and seismic bodywave modelling. *Earth and Planetary Science Letters*, *172*(1-2), 23–37.

Zhang, T., Zhang, W., Cao, D., Yi, Y., & Wu, X. (2022). A new deep learning neural network model for the identification of insar anomalous deformation areas. *Remote Sensing*, *14*(11), 2690.

Zhao, Y., Deng, B., Shen, C., Liu, Y., Lu, H., & Hua, X.-S. (2017). Spatio-temporal autoencoder for video anomaly detection. In *Proceedings of the 25th acm international conference on multimedia* (pp. 1933–1941).

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