

Coherent pixel selection using a dual-channel 1-D CNN for time series InSAR analysis

Y. Zhang^{a,*}, J. Wei^{a,*}, M. Duan^a, Y. Kang^b, Q. He^{a,c}, H. Wu^a, Z. Lu^d

^a Chinese Academy of Surveying and Mapping, Beijing 100036, China

^b School of Geomatics, Liaoning Technical University, Fuxin, Liaoning 123000, China

^c School of Environment and Spatial Informatics, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China

^d Southern Methodist University, Dallas, TX 75275, USA

ARTICLE INFO

Keywords:

Coherent pixel selection
Deep learning
One-dimensional convolution neural network
(1-D CNN)
Time series InSAR

ABSTRACT

Coherent pixel (CP) selection is an important step in the processing chain of time series InSAR analysis. In this research, we propose a light deep learning framework, i.e., a dual-channel one-Dimensional Convolution Neural Network (1-D CNN) to select CPs. The 1-D CNN has simple input: SAR amplitude and interferogram coherence, and can be trained with CP samples generated by traditional thresholding method. In an experiment based on Sentinel-1 temporal images in Tianjin, China, the 1-D CNN substantially outperforms the thresholding method and the StaMPS method in terms of the amount and the quality of selected CPs. Additionally, a new measure is proposed to quantify CP quality, which is very useful when other reference data is unavailable. The proposed 1-D CNN framework on CP selection is reliable and fast, and of great significance in developing automatic time-series InSAR processing system.

1. Introduction

Time series InSAR (TS-InSAR) techniques, including persistent scatterer InSAR (PS-InSAR) (Ferretti et al., 2000, 2001) and small baseline subset (SBAS) InSAR (Berardino et al., 2002), have been widely used to monitor ground deformation. Coherent pixels (CPs) related to permanent scatterer (PS) or distributed scatterer (DS) (Ferretti et al., 2011) exhibiting consistent scattering response over time are the basis for time series InSAR analysis. As an important step in TS-InSAR processing chain, CP selection aims to extract true coherent pixels as dense as possible, over which geophysical parameters such as deformation, DEM error and atmosphere, can be retrieved accurately.

A PS pixel, usually containing dominant scatterers within the resolution cell, is characterized by high coherence values in all interferograms and stable and high amplitude values in the time series SAR images (Ferretti et al., 2001). On the other side, a DS pixel, which doesn't contain dominant scatterers within the resolution cell, still has moderate coherence in some interferograms. Therefore, CPs can be detected by simply setting thresholds to specific parameters related to radar reflectivity, such as the amplitude dispersion (Ferretti et al., 2001), and/or parameters related to coherence behavior, such as the mean coherence of available interferograms (Mora et al., 2003) or the

coherence of each interferogram (Berardino et al., 2002). Wang et al. (2018) proposed a novel coherence-based method for CP selection. In this method, a pixel is recognized as a CP if the interferograms in which the coherence of the pixel is above a threshold forms a connected network involving all the SAR images. In comparison with the methods of checking the mean coherence or each interferogram coherence, this method is more relevant to the reliability of deformation retrieval, because the coefficient matrix in this method is full rank, thus deformation phase can be estimated under the least square sense, without taking the risk of solving a under-determined system.

Apart from these straightforward parameters, the StaMPS (Stanford Method for PS) (Hooper et al., 2004) uses a more complicated parameter, the temporal coherence, to set threshold values for CP selection. The temporal Coherence represents a measure of the variation of phase noise over the used interferograms.

$$\gamma_T = \frac{1}{M} \left| \sum_{i=1}^M e^{j\phi_{n,i}} \right| \quad (1)$$

Where M is the number of used interferograms, and $\phi_{n,i}$ denotes the noise phase due to variability in scattering, thermal noise, coregistration errors and uncertainty in the position of the phase center (Hooper et al.,

* Corresponding authors.

E-mail addresses: yhzhang@casm.ac.cn (Y. Zhang), weijj@casm.ac.cn (J. Wei).

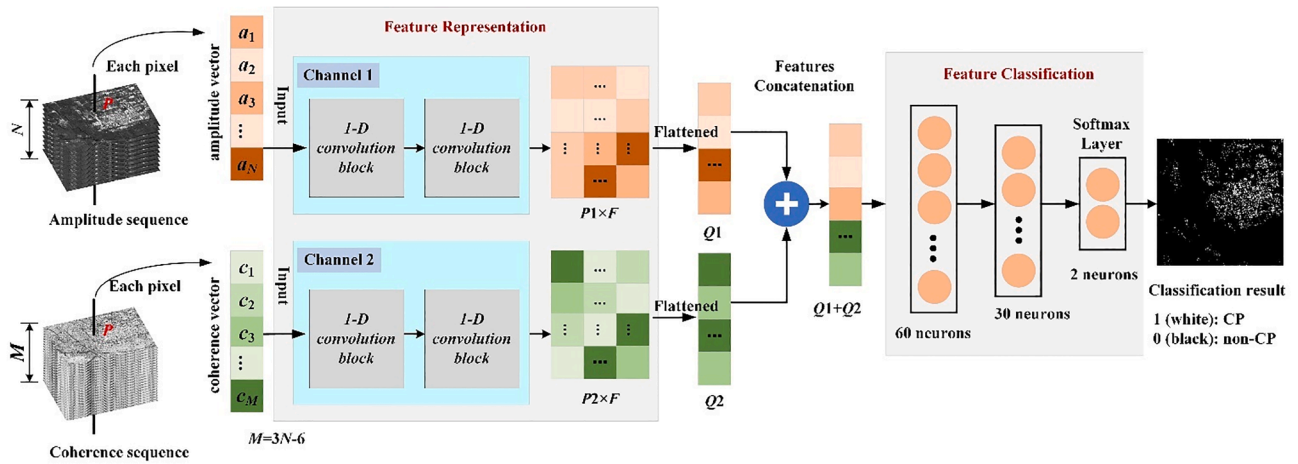


Fig. 1. Diagram of the dual-channel 1-D CNN framework.

2007) of a pixel in the i th interferogram. For a CP, the noise should be small and presents a random behavior (Zhao and Mallorqui, 2019), thus a large value of the temporal coherence is expected. However, in order to obtain a reliable estimate on noise phase, the deterministic phase components, including phases related to deformation, atmospheric artifacts, DEM error and orbit uncertainty, must be estimated and subtracted from the interferometric phase. Therefore, CP selection with StaMPS involves large computation cost.

Besides the above-mentioned pixel based methods, another large group of CP selection methods are segment based, which are designed to exploit the phase information of all statistically homogeneous pixels (SHP) in a neighborhood of a DS pixel. The rationale of this group of methods is to retrieve the parameters of interest, such as deformation velocity, elevation, and deformation time series, more accurately by processing the interferometric phase of many homogeneous pixels in a segment, rather than the phase of each DS pixel alone, as the phase of one DS pixel is noisy due to the effect of decorrelations (Ferretti et al., 2011). Examples of this kind of CP selection methods are: the SqueeSAR (Ferretti et al., 2011), the methods proposed by Wang et al. (2012), Jiang et al. (2015) and the CAESAR (Fornaro et al., 2015). This group of CP selection methods typically consist of two consecutive processing steps. The first step is to identify the SHP within a specified neighborhood of a pixel, and a pixel with the number of SHP above a threshold is recognized as a DS pixel. In the second step, the phase of every extracted DS is reconstructed or filtered by means of the phases of its SHP with different algorithms, such as phase triangulation (Ferretti et al., 2011), the eigenvalue decomposition of covariance matrix (Fornaro et al., 2015) or the temporal coherence estimation (Zhao and Mallorqui, 2019). Thus, the original phase noise in DS pixels can be largely mitigated and the parameters of interest can be retrieved more reliably from the filtered phases. Consequently, this group of CP selection methods is also computationally expensive.

As a summary, the CP selection methods by thresholding straightforward parameters are computationally efficient. But these methods are very sensitive to the threshold values, and determining appropriate threshold values might be difficult and subjective, especially when the scene coverage is large. A strict threshold value tends to discard true CPs, thus lowers the measurement density, while a loose threshold value can result in large commission errors. For those complicated CP selection methods, including the StaMPS and the DS exploitation methods, determination of threshold values is not a problem as the threshold value is either a derivative, like the threshold value to the temporal coherence to identify a CP in StaMPS, or can adopt an empirical value, like the threshold value to the number of SHP to confirm a DS pixel in SqueeSAR. However high computation cost remains a problem for those methods. For example, the time required by CP selection with StaMPS

may vary from hours to days depending on number of candidate pixels and the size of the dataset as reported by Tiwari et al. (2020). Moreover, high computation cost could become unacceptable in specific circumstances such as emergency responses and regular monitoring activities. Therefore, reliable and computationally fast methods of CP selection need to be developed in order to meet the challenges imposed by deformation monitoring applications concerning large data volume, wide area and urgent time requirements.

During the last decade, neural networks technology in the name of deep learning, with strong ability in automatically learning high-level features hidden in the training samples, have been successfully used in image classification (Krizhevsky et al., 2012) and object detection (Girshick et al., 2014). CP selection can also be regarded as a task of classifying stable scattering pixels from time series SAR images, thus suitable for deep learning-based solutions. Recently, Tiwari et al. (2020) reported a study on CP selection in TS-InSAR with deep learning methods. In this study, a two-dimensional (2-D) convolution neural network (CNN) architecture and a convolutional long short term memory network (CLSTM) architecture were proposed to learn spatial and spatio-temporal features of CPs respectively. The training dataset used in Tiwari et al. (2020) is composed of around 10,000 image (interferometric phase) blocks of size 100×100 with each pixel labelled as a CP or a non-CP by referring to the CP selection result from StaMPS. However, the adoption of 2-D convolution in Tiwari et al. (2020) is open to discussion. CPs are spatially discrete, and the occurrence of a CP is independent of the existence of other CPs in its neighborhood. In other words, spatial context has little implication for identifying a CP. The finding in Tiwari et al. (2020) that the CLSTM architecture can achieve better accuracy and density than the CNN architecture itself also implies that the temporal features play more important roles than the spatial features in CP identification. Another drawback of using 2-D convolutions is the problem of class imbalance existed in training samples: the majority (i.e., above 95 %) of pixels within a training image block on most occasions are non-CPs (Tiwari et al., 2020). As a result, the network will be trained to tend to label a CP as a non-CP, thus the trained network will increase the false negative rate. Although the class imbalance problem has been tackled in the research by introducing the f1-score as the loss function and giving much higher weight to CP class than non-CP class, these treatments are not entirely effective according to Johnson and Khoshgoftar (2019).

In this research, we propose to use one-dimensional CNN (1-D CNN) for CP selection in TS-InSAR. The 1-D CNN was developed for time series signal processing in 2015 (Kiranyaz et al., 2015). In comparison with 2-D CNN, 1-D CNN has fewer network parameters, thus is computationally more efficient and requires a smaller number of training samples to train the network. Moreover, there is not the problem of class imbalance in

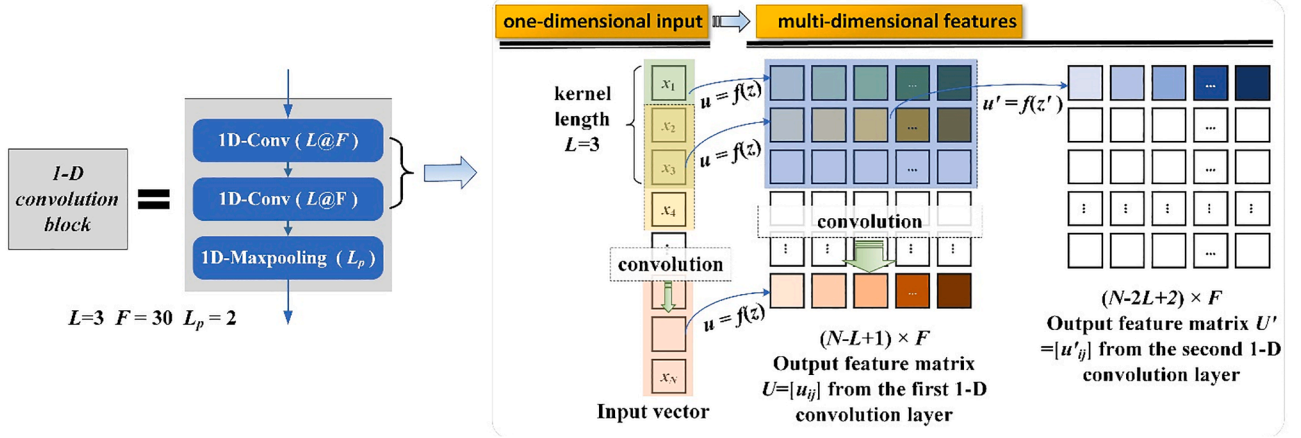


Fig. 2. Diagram of a 1-D convolution block.

training samples, because training samples are pixels, rather than image blocks, and the proportion of positive (CP) and negative (non-CP) samples is under control.

Generally, the information of time series single-polarization SAR images is fully represented by the Hermitian data covariance matrix. The diagonal elements of the covariance matrix are the intensity (square of amplitude) of the time series SAR images, and the off-diagonal elements are related to the complex coherence of all possible interferograms. Therefore, a dual-channel 1-D CNN with one channel processing the SAR amplitude and the other processing the coherence of interferograms is adopted.

The remainder of this paper is structured as follows. The proposed dual-channel 1-D CNN is detailed in section 2. The section 3 is dedicated to an experiment with Sentinel-1A (S-1A) images. CP extraction results from the proposed 1-D CNN, StaMPS, and thresholding methods are analyzed. Also, a parameter used to measure the quality of CPs is introduced. CPs from the three methods are compared in terms of density, the quality measure, accuracy of the derived deformation and the computation time. Finally, discussion and conclusion are presented in section 4.

2. Proposed neural network framework

The proposed network consists of a feature representation component and a feature classification component (Fig. 1). The feature representation component has two channels responsible for extracting high level features from the input amplitude and the coherence vectors respectively. The methodology points of the proposed network can be presented schematically as follows:

- (1) For each pixel in a stack of co-registered SAR images acquired at N time points, there are a corresponding amplitude vector, i.e., $A = [a_1, a_2, a_3, \dots, a_N]$ and a coherence vector, i.e., $C = [c_1, c_2, c_3, \dots, c_M]$, which form the input vectors of the pixel.
- (2) The input amplitude vector and coherence vector are processed by the two channels in the feature representation component respectively. In each channel, a network formed by cascading two 1-D convolution blocks is used to extract high-level temporal features of the input vector, which are output into a feature matrix.
- (3) The extracted feature matrix output from each of the two channels is reorganized (flattened) as a feature vector. Then the two vectors are concatenated as one feature vector.
- (4) The concatenated feature vector is fed into the feature classification component, i.e., the 3-layer multi-layer perceptron (MLP), to classify the pixel as a CP or a non-CP.

2.1. 1-D convolution block

A 1-D convolution block is the fundamental unit of the feature representation component. Every 1-D convolution block has two 1-D convolution layers (Fig. 2), each consisting of a few convolution kernels to extract multi-dimensional features from the input data, followed by a max-pooling layer to eliminate the redundancy in the extracted features.

In the first 1-D convolution layer, for a given kernel with length L , a N -sized input vector, which is either the SAR amplitude vector or the coherence vector, will generate a $(N-L+1)$ -sized output vector if the stride equals to 1. The convolution operation can be expressed as:

$$u = f(z) = f(W_1 \cdot X + b) = f\left(\sum_{i=1}^L w_{1,i} \cdot x_i + b\right) \quad (2)$$

where X denotes the input vector, $w_{1,i}$ and b denote the weight and the bias specific to the kernel, $f(\bullet)$ is the ReLU activation function.

Given F kernels, F different feature vectors will be obtained. These feature vectors are then stacked as column vectors into a feature matrix U with $N-L+1$ rows by F columns (see Fig. 2). In the second 1-D convolution layer, the 1-D convolution process is conducted similarly, and can be formulated as:

$$u' = f(z') = f(W_2 \cdot U + b) = f\left(\sum_{i=1}^{N-L+1} \sum_{j=1}^F w_{2,i,j} \cdot u_{i,j} + b\right) \quad (3)$$

where W_2 denotes the weight matrix, and $(W_2 \cdot U)$ stands for the inner product of two matrixes. The output feature matrix U' from the second 1-D convolution will have $N-2L+2$ rows and F columns.

After that, a non-overlapping 1-D max pooling with a kernel length of 2 is applied to each column vector of the output feature matrix to extract the more prominent from every two adjacent features:

$$v = \max\{u'(2i, j), u'(2i+1, j)\}, i \in [1, P], j \in [1, F] \quad (4)$$

where $P = \lfloor (N-2L+2)/2 \rfloor$, and $\lfloor \cdot \rfloor$ means taking the integer. The max pooling operation aims to eliminate the redundancy in the feature vector by reducing its dimensionality.

2.2. MLP

The MLP is composed of three fully connected layers having 60, 30 and 2 neurons respectively. Similar to the convolution, the operation associated with every neuron can be expressed as:

$$z = f(V \cdot X + b) \quad (5)$$

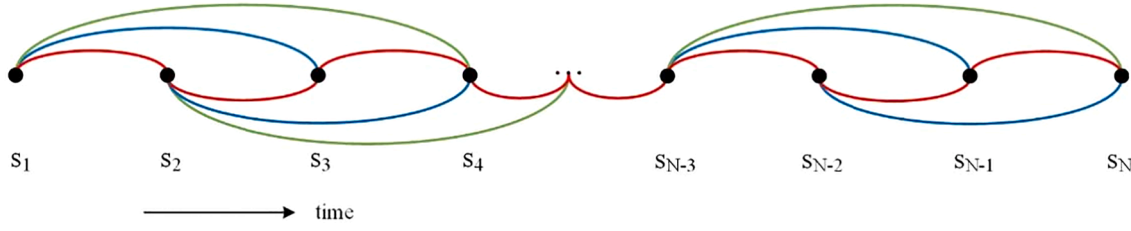


Fig. 3. Illustration of the temporal baseline configuration of interferogram generation rule. Every image is connected to three consecutive images to form interferograms.

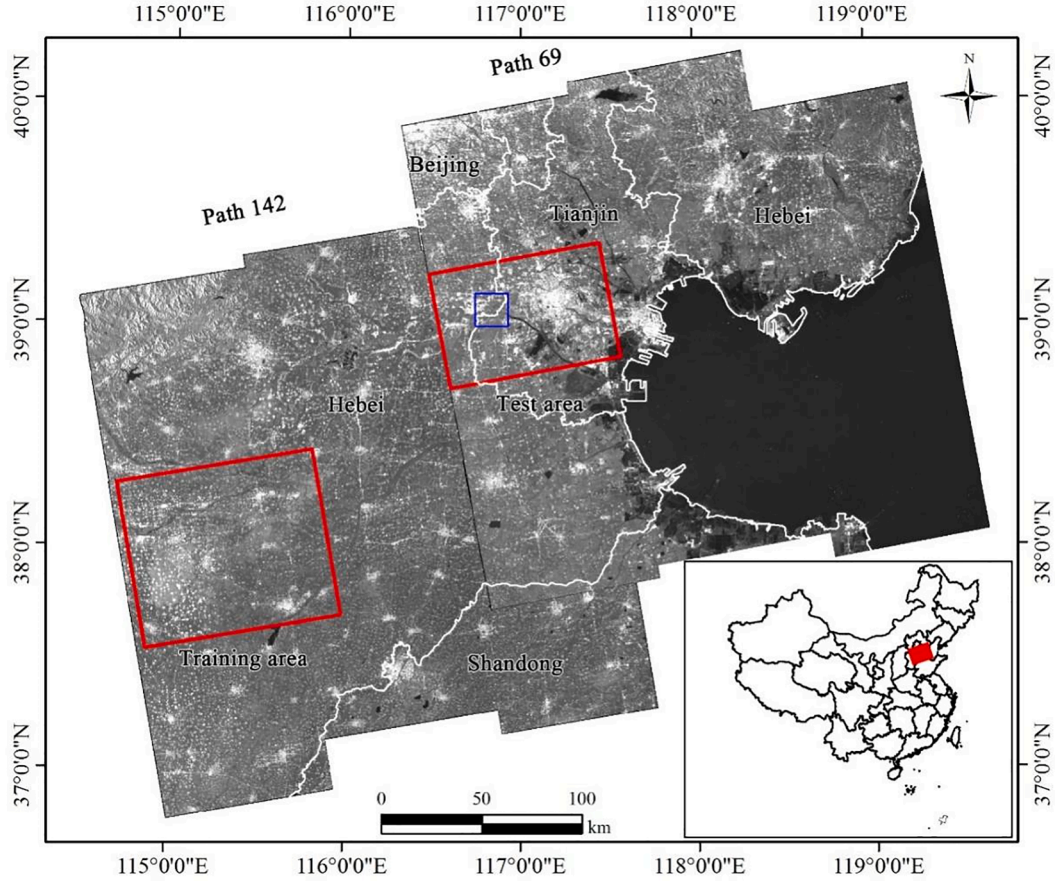


Fig. 4. Geocoded mean amplitude from two neighboring S-1A stacks over Tianjin, China with the inset showing its location in China. The training area and the test area are delineated by red rectangles.

where \mathbf{V} , \mathbf{X} and b denote the weight vector, the combined feature vector and the bias respectively. The activation function $f(\cdot)$ for the first two layers is the ReLU function, and is the SoftMax function for the third layer which calculates the probability of the input pixel being a CP and a non-CP.

2.3. Determination of main hyperparameters

The number of convolution kernels F and the length of each kernel L are two important hyperparameters. The number of convolution kernels corresponds to the dimensionality of the feature vector extracted by the convolution operation and the length of the kernel tells how many consecutive elements within the input vector are used to explore the temporal features. The values of F and L are determined through trial in the training stage. Especially, the number of convolution kernels F and the length of each kernel L have been tested by taking the values of {10, 20, 30, 40, 50}, and {2, 3, 4, 5} respectively. It turns out that the CP

selection result with $F = 30$, and $L = 3$ has best accuracy. This finding is explainable by the characteristics of the input data. The dimensionality of the extracted feature vector is related to the number of SAR images used. 29 acquisitions of S-1A within the year of 2018 are used for CP selection in the experiment. Therefore, 30 is an appropriate value for the hyperparameter F . On the other side, the value of kernel length L is related to the way the interferograms are formed. How to form the interferograms is an important step in time series InSAR analysis. In order to improve the generalization level of the proposed network, the interferograms are formed according to a fixed temporal baseline rule (Fig. 3): every SAR image is connected to three temporally closest images to form interferograms. That is, $3 \times N-6$ interferograms will be formed from N temporal SAR images. Therefore, the temporal features of the input coherence vector could be well investigated based on the coherence values of every 3 consecutive interferograms.

Table 1

The number of CPs selected by the three methods.

Number of CPs	Thresholding	StaMPS	1-D CNN
Thresholding	354,666	169,050	354,522
StaMPS		385,538	207,224
1-D CNN			452,040

3. Experiment and analysis

The test site is in Tianjin, China, and covered by two neighboring tracks (path 142, path 69) of S-1A Interferometric Wide-swath (IW) mode SAR images (Fig. 4). 29 images are collected from path 142 from 3 January to 29 December of 2018 with two acquisitions missing on 16 March and 30 October. 30 images are collected from path 69 from 10 January to 24 December of 2018 with an exact 12-day interval. Since the training data and the test data are from different paths, we remove the image acquired on March 11 from the stack of path 69, so that the number of SAR images over each track is 29.

According to the small temporal baseline configuration (Fig. 3), 81 interferograms are formed from 29 S-1A images over both tracks. 2×8 multi-looking is employed in the interferogram generation to suppress the speckle noise. The amplitude data is generated through three steps: 1) the imaginary and the real parts of the original single look complex data is calibrated respectively, and the 1-look amplitude is generated; 2) averaging the 1-look amplitudes in a 2×8 window to obtain the multi-looked amplitude; 3) each of the multi-looked amplitude is normalized by dividing the mean of the 29 amplitudes. The training area and the test area contain 3000×3000 and 2088×2707 multi-looked pixels respectively (Fig. 4.) The reason for choosing a image subset of 3000×3000 pixels as the training area is that this size of coverage is big enough to generate adequate number of training samples for the network, as the proposed dual-channel 1-D CNN has a slight structure with the amount of network parameters around 58,000. As for the test area, there is no limitation to the size of test area. In fact, the whole scene coverage of the

track 69 can be used as the test area. The reason for choosing the 2088×2707 sized test area is that the leveling measurements used for validation of InSAR-derived deformation results are all within this area. Further, it will take longer time for the StaMPS to complete CP selection if a test area with larger size is adopted.

Having a large number of training samples in advance is a prerequisite for deep learning approaches. Usually generating the training samples is difficult and time-consuming. This is especially true when we try to distinguish a lot of CPs from SAR images through visual interpretation. Here we extract CP training samples from the training area with the thresholding method. Threshold values to the two parameters: mean coherence of the 81 interferograms and mean amplitude of the 29 S-1A images, are manually determined by trial. The final thresholding rule for extracting CP training samples are:

$$(\bar{\gamma}_s > 0.8) \text{ or } [(\bar{\gamma}_s > 0.71) \text{ and } (\bar{A} > 1.1)] \quad (6)$$

where $\bar{\gamma}_s$, \bar{A} are the mean coherence and the mean amplitude respectively. This thresholding rule is intended to extract pixels with very high coherence or medium-high coherence plus medium-high amplitude.

Non-CP samples are selected by thresholding the same parameters inversely. Eventually, 277,389 CP samples and 737,906 non-CP samples are extracted. For every training sample, there is a corresponding vector composed of an amplitude sequence (29 elements), a coherence sequence (81 elements), and a label indicating the sample being a CP (label value = 1) or a non-CP (label value = 0).

The 1,015,295 samples are divided into two parts: 70 % of the CP and non-CP samples used for network training and the rest for validation. The 710,707 training samples are input into the proposed network with a mini batch of 10,000 samples per epoch. The network biases are initialized as zero, and the weights are initialized by the He normal initializer (He et al., 2015). The dropout rate is set to 0.5. At the last stage, the cross-entropy loss function and the Adam optimizer (Kingma and Ba, 2015) with learning rate of 0.001 are adopted to update the network biases and weights.

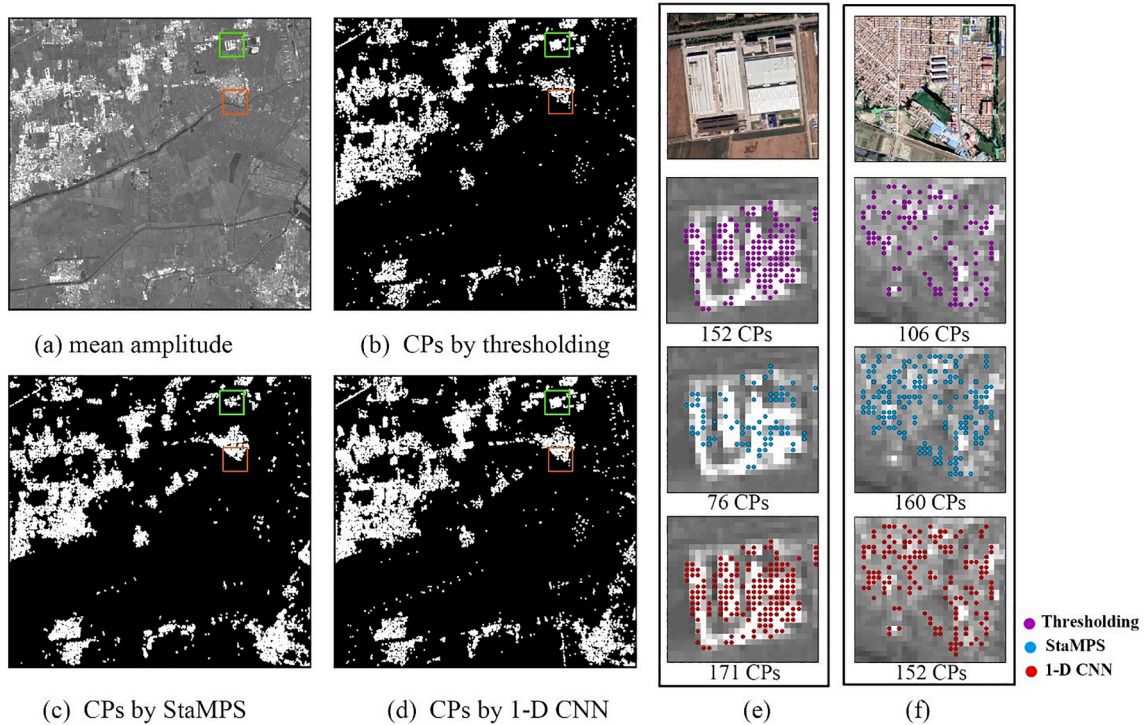


Fig. 5. CP selection result of the three methods over a subset marked by the blue rectangle in Fig. 4. The enlarged views of CPs overlaid on the SAR amplitude over a large building and a rural residential area delineated by the green and the brown rectangles in (a)-(d) are shown in (e) and (f). The top-panel images in (e) and (f) are optical images.

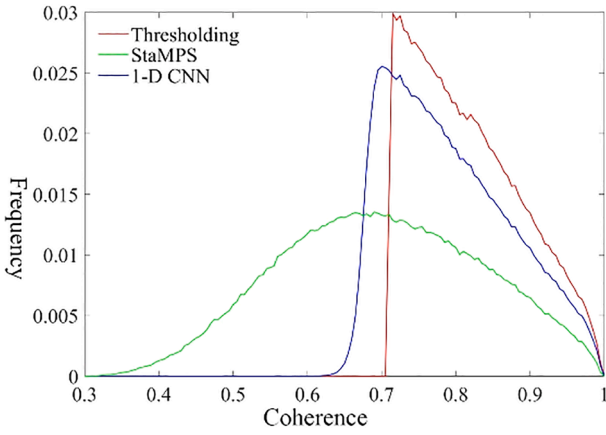


Fig. 6. Histograms of mean coherence of the three sets of CPs.

The trained network is then applied to the test data of path 69, and 452,040 CPs are extracted by the network. As a comparison, the same thresholding rule and the StaMPS method are also applied to the test area. The numbers of CPs selected by the three methods are listed in Table 1 with the diagonal entries recording the number of CPs selected by each of the three methods, and the off-diagonal entries recording the number of common CPs selected by two of the three methods. The CP selection results over a subset of the test area (marked by the blue rectangle in Fig. 4) are shown in Fig. 5. The enlarged views of CPs selected by the three methods over a large building and a rural residential area are shown in Fig. 5(e) and (f). More CPs are extracted by the thresholding and the 1-D CNN methods over the building than the StaMPS (Fig. 5(e)), while more CPs are selected by the StaMPS and the 1-D CNN than the thresholding over the rural residential area (Fig. 5(f)). This will be discussed in more detail in next sub-section.

3.1. CP density

From Table 1, the proposed 1-D CNN has selected the largest number of CPs, e.g., 27 % and 17 % more CPs than the thresholding and the StaMPS methods respectively. Especially, since the 1-D CNN is trained with training samples generated by the thresholding method, the trained network has selected nearly all the CPs (354,522 out of 354,666, or 99.96 %) extracted by the thresholding method over the test area. This demonstrates that the 1-D CNN has learned well the features represented by the thresholding rules. We also notice that the percentage of common CPs between the StaMPS and the 1-D CNN or the thresholding method is less than 50 %, which means that the StaMPS-selected CPs largely don't coincide with those extracted by the other two methods. This difference is evident from the distribution of the mean coherence as shown in Fig. 6. The threshold value for the mean coherence in the thresholding method is 0.71. The 1-D CNN has learned this feature of high coherence in the time series interferograms. Therefore, the mean coherence of the CPs selected by the 1-D CNN also presents large values, i.e., all above

0.63. While the CP selection in the StaMPS algorithm has nothing to do with the spatial coherence, thus the mean coherence of CPs selected by the StaMPS spans a wide range of [0.33, 0.99] and inherently has a Gaussian distribution according to the central limit theorem. From the enlarged views of Fig. 5 (e) and 5 (f), the thresholding and the 1-D CNN methods have selected much more CPs than StaMPS over the large building where many PS pixels with strong and stable backscattering exist. While over the rural residential area where the SAR backscattering is not very strong but remains relatively stable like DS pixels, StaMPS has selected more CPs than thresholding. This has confirmed one of the advantages of the StaMPS: it is able to achieve denser measurements than amplitude-based PS, especially over non-urban area, through identifying both PS and DS (Osmanoğlu et al., 2016). Nevertheless, the 1-D CNN has selected 66,502, or 17 % more CPs than the StaMPS over the whole test area.

3.2. CP quality

Besides density, another significant concern is the quality of the CPs. Generally, one can evaluate the quality of CPs by comparing the deformation derived from these CPs with ground truth. However, the reference data used in deformation accuracy evaluation, such as leveling or GPS measurements, are often unavailable.

Here we propose to use the multi-image phase coherence defined by Ferretti, et al., (2000) or equivalently the model coherence in Mora, et al., (2003) as an alternative to measure CP quality when reference data are unavailable. The model coherence is defined over an arc connecting two neighboring CPs by.

$$\gamma_{ij} = \frac{1}{L} \left| \sum_{k=1}^L \exp[j(\varphi_k^{ij} - \varphi_{model}^{ij})] \right| \quad (7)$$

where i, j denotes two neighboring CPs, φ_k^{ij} is the wrapped phase difference over the arc in the k th differential interferogram, φ_{model}^{ij} is the model phase associated with deformation and DEM error, and can be express as.

$$\varphi_{model}^{ij} = (\varphi_{def,k}^i - \varphi_{def,k}^j) + aB^k \cdot (h_{err}^i - h_{err}^j) \quad (8)$$

where B^k is the perpendicular baseline of the k th interferogram, h_{err}^i is the DEM error at pixel i , $\varphi_{def,k}^i$ is the deformation phase. In many cases, a linear deformation model is assumed, i.e. $\varphi_{def,k}^i = -4\pi/\lambda \cdot \mathbf{v}^i \cdot \mathbf{t}_k$ with \mathbf{v}^i being the deformation velocity and \mathbf{t}_k being the temporal baseline. A large value of the model coherence, especially close to 1, means that the interferometric phase of the two CPs is less affected by noise and the phase difference is well fitted by the model, thus the retrieved deformation parameters will be reliable. Now we define the model coherence for individual CP by taking the average of model coherence of all arcs connected to the CP by.

$$\gamma_i = \langle \gamma_{ij} \rangle_{\{j|(i,j) \in \Omega\}} \quad (9)$$

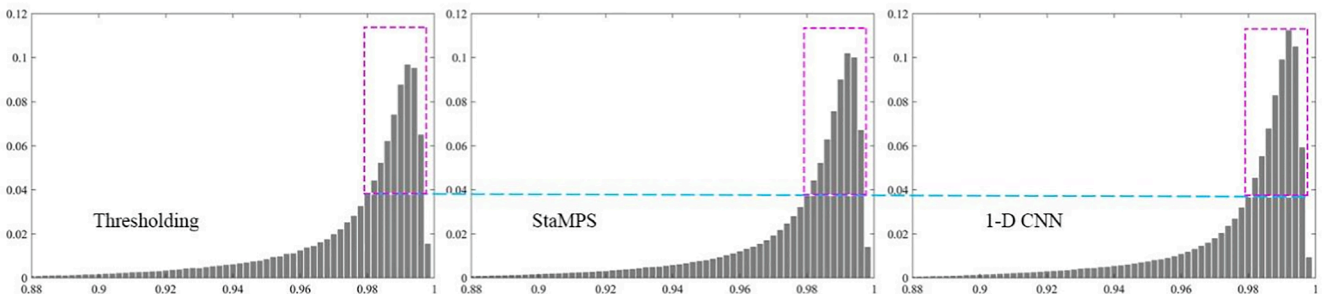


Fig. 7. Histograms of CP's model coherence of the three sets of CPs.

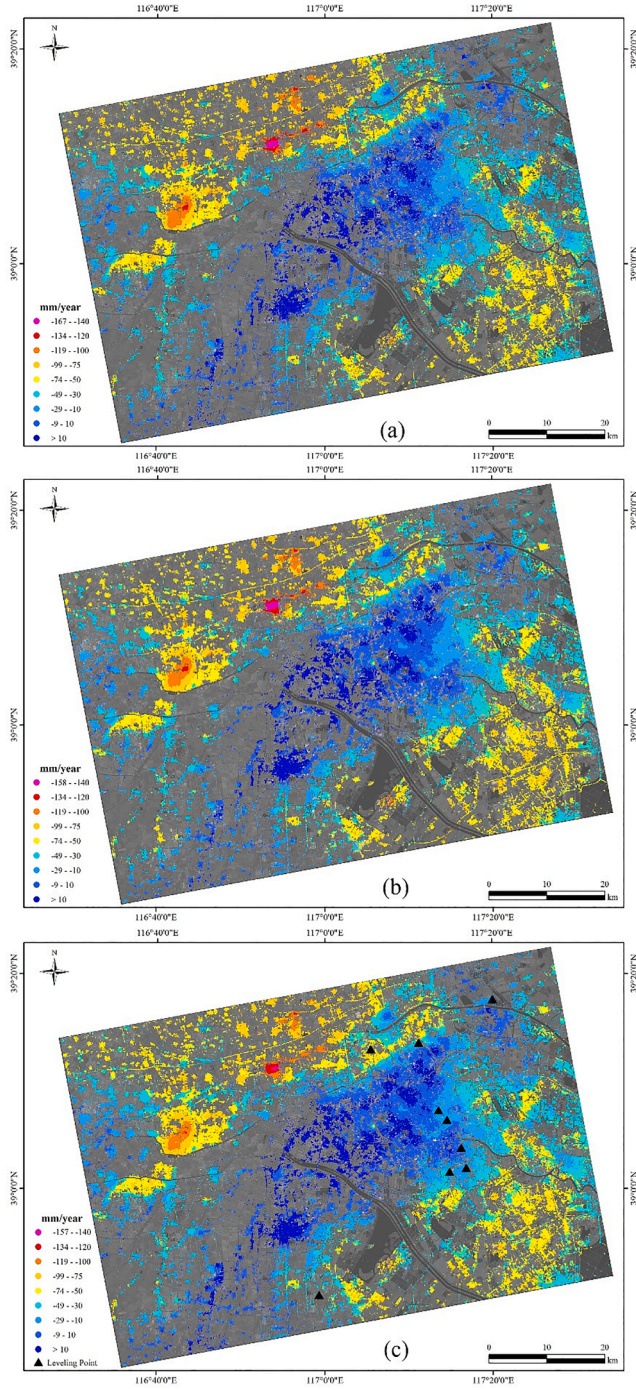


Fig. 8. Subsidence velocity calculated from the CPs extracted by thresholding (a), StaMPS (b) and the 1-D CNN (c), where negative value means ground subsidence, and positive value means ground uplift. The positions of 9 leveling points are marked by black triangles in (c).

where Ω denotes the set of all existing arcs which are usually generated by the Delaunay triangulation of selected CPs, $\langle \cdot \rangle$ means taking the average. The model coherence defined by (9) can be used as a measure of CP's quality.

The histograms of the model coherence of the three sets of CPs are plotted in Fig. 7. There is larger portion of CPs in the 1-D CNN set having very high model coherence, e.g., in the bins of [0.98, 1], than those in the other two sets. Specifically, the ensemble mean of CP's model coherence for the 1-D CNN set, the thresholding set and the StaMPS set are 0.9837, 0.9813 and 0.9810 respectively. Therefore, it can be

concluded that the CPs selected by the 1-D CNN have higher quality than those by the thresholding and the StaMPS methods even though 1-D CNN detects 17–27 % more CPs than the other two methods.

3.3. Deformation accuracy validated by leveling measurements

Fortunately, we have collected a few leveling measurements over the test area, which can be used to validate the accuracy of deformation result. We have calculated the deformation velocity using the three sets of CPs with the algorithm and software presented in Zhang et al. (2016) and Zhang et al. (2021). The three results are shown in Fig. 8. To compare leveling measurements with InSAR, the InSAR line of sight (LOS) deformation velocity has been projected into the vertical direction through dividing the LOS velocity by the cosine of SAR viewing angle to obtain a subsidence velocity. Note, only 9 out of the collected leveling points are used for accuracy validation. The remainder are discarded because there are no CPs in the 3×3 window centered at those leveling points. For every used leveling point, a CP closest to the leveling point is located and its subsidence velocity is compared with the leveling measurement. The comparison result is listed in Table 2.

The RMSE of the subsidence velocity derived from 1-D CNN CPs is 6.3 mm/year, better than those from the StaMPS and the thresholding CPs. This result is consistent with the evaluation result of CP's model coherence presented in previous subsection and confirms that the CPs derived by the 1-D CNN has the best quality among the three sets of CPs.

3.4. Computation time

Computation time is of great significance for time series InSAR analysis, especially when the study area is large or there is a strict time limitation. The computation times costed by the three methods over the test area are listed in Table 3. Data processing was conducted on a personal computer equipped with an Intel Xeon Gold 6154 CPU with 3.00 GHz frequency, a 64 GB RAM, and a 24 GB NVIDIA Quadro RTX 6000 GPU. For the thresholding method, the time used by the operator to determine the proper threshold values has not been included. Similarly, for the 1-D CNN, the time used to train the network has also not been included. Obviously, the thresholding method is the fastest. The computation time of StaMPS and the 1-D CNN is 9,692 and 144 s respectively when the CPU is utilized, and it is shortened to 8 s if the 1-D CNN is executed on the GPU. Therefore, the 1-D CNN is computationally much more efficient than StaMPS. The high computation cost is related to the complicated procedure of CP selection in StaMPS which has been briefed in the introduction section. Therefore, one must bear in mind the time cost when applying StaMPS to large areas and/or big data stacks.

4. Discussion and conclusion

A deep learning framework, i.e., the dual-channel 1-D CNN, has been proposed to select CPs for time series InSAR analysis. The proposed network has a light structure characterized by 1-D convolutions. The adoption of 1-D rather than 2-D CNN is consistent with the fact that CPs are spatially discrete, and the occurrence of a CP is independent of the existence of other CPs in its neighborhood. Moreover, in comparison with 2-D convolution, the utilization of 1-D convolution can avoid the problem of class imbalance, require less training samples to train the network, and increase the computation efficiency. The proposed network has simple input, i.e., SAR amplitude data and interferogram coherence data of a pixel, therefore can be easily integrated into any TS-InSAR processing chain.

The effectiveness and efficiency of the 1-D CNN in CP selection has been tested with S-1A time series images. In comparison with the thresholding and the StaMPS methods, the 1-D CNN has selected 27 % and 17 % more CPs respectively. The CP's model coherence has been proposed as a measure to quantify CP quality, especially when other reference data, such as deformation measurements from ground-based

Table 2

The difference between leveling subsidence velocity and InSAR subsidence velocity (unit: mm/year).

Leveling point	Subsidence velocity from leveling	subsidence velocity from Thresholding CPs	Difference with leveling	subsidence velocity from StaMPS CPs	Difference with leveling	subsidence velocity from 1-D CNN CPs	Difference with leveling
C135	-16.0	-28.9	-12.9	-28.8	-12.8	-26.3	-10.3
JC259	-49.0	-59.8	-10.8	-47.4	1.6	-56.9	-7.9
JC1417	-15.0	-11.9	3.1	-25.7	-10.7	-9.1	5.9
B577A	-16.0	-23.6	-7.6	-20.8	-4.8	-21.6	-5.6
B654	-15.0	-16.8	-1.8	-13.3	1.7	-16.1	-1.1
B1860	-10.0	-3.1	6.9	-6.7	3.3	-5.9	4.1
JC1197	-40.0	-36.1	3.9	-35.3	4.7	-34.5	5.5
JC1034	-63.0	-61.9	1.1	-56.0	7.0	-58.3	4.7
JC1600	-26.0	-9.7	16.3	-15.0	11.0	-18.4	7.6
Root mean square error (RMSE)			8.7				6.3

Table 3

Computation time of the three methods.

CP selection method	Computation time (seconds)
Thresholding (on CPU)	0.3
StaMPS (on CPU)	9,692
1-D CNN (on CPU)	144
1-D CNN (on GPU)	8

methods, are unavailable. The CPs selected by the 1-D CNN has shown better quality in terms of CP's model coherence. Moreover, this quality evaluation result has been confirmed through validating the deformation accuracy with leveling measurements. Considering computation time, the 1-D CNN is 67 times faster than the StaMPS method when implemented on a CPU, and the computation efficiency of the 1-D CNN can increase by 17 times if it is run on a GPU. Therefore, it can be concluded that the 1-D CNN method is superior in CP selection through achieving higher density and better quality of the CPs with competitive computation cost.

Since interferogram coherence is one of the two input data, adoption of different methods of coherence estimation will have effects on the performance of the proposed 1-D CNN CP selection method. The classical boxcar method, i.e., 2×8 multi-looking, is adopted in the current implementation of the 1-D CNN. Recently, Mukherjee et al. (2021) proposed a CNN-based generative model (namely GenInSAR) to estimate coherence, which could slightly improve the accuracy of coherence estimation by 4 % and save 50 % of computation time in comparison with boxcar. Therefore, integration of the GenInSAR into the framework is one of the future tasks to improve the 1-D CNN CP selection method further.

Given the increased interest in obtaining timely deformation over large areas, applying high performance computing (HPC) techniques to InSAR processing, such as Casu et al. (2014) and Duan et al. (2020), has received extensive attention in recent years (Imperatore et al., 2021). In order to fully exploit the capability of HPC resources, the InSAR processing methodology needs to be as automatic as possible. However, CP selection is one of the few processing steps still lacking enough automation. In the experiment, the 1-D CNN is trained with samples taken from a data stack, and the trained network is directly applied to another data stack and performs well. In other words, the generalization ability of the proposed 1-D CNN framework has been partly demonstrated. Hence, the automation of CP selection could be improved by adopting the 1-D CNN method. Of course, more tests are required to investigate how large the area a trained 1-D CNN can be successfully applied to. Nevertheless, it can be concluded that the proposed 1-D CNN has great potential to be incorporated into a HPC-based InSAR processing system as a solution to speed up reliable CP selection.

Funding

This work was supported by the National Natural Science Foundation

of China (grant number 41874014); the National Key R&D Program of China (grant number 2018YFB0505400); and the Ministry of Natural Resources of China (grant number A2018).

CRedit authorship contribution statement

Y. Zhang: Conceptualization, Supervision, Methodology, Funding acquisition, Project administration, Writing - original draft, Writing - review & editing. **J. Wei:** Methodology, Software, Data curation. **M. Duan:** Resources, Validation. **Y. Kang:** Formal analysis, Data curation, Investigation, Visualization. **Q. He:** Investigation, Visualization. **H. Wu:** Investigation. **Z. Lu:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Sentinel-1 data are copyrighted to the European Space Agency.

References

- Berardino, P., Fornaro, G., Lanari, R., Sansosti, E., 2002. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans. Geosci. Remote Sens.* 40 (11), 2375–2383. <https://doi.org/10.1109/TGRS.2002.803792>.
- Casu, F., Elefante, S., Imperatore, P., Zinno, I., Manunta, M., De Luca, C., Lanari, R., 2014. SBAS-DInSAR parallel processing for deformation time-series computation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 7 (8), 3285–3296. <https://doi.org/10.1109/JSTARS.2014.2322671>.
- Duan, W., Zhang, H., Wang, C., Tang, Y., 2020. Multi-temporal InSAR parallel processing for sentinel-1 large-scale surface deformation mapping [Online]. Available: Remote Sens. 12, 3749 <https://www.mdpi.com/2072-4292/12/22/3749>.
- Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., Rucci, A., 2011. A new algorithm for processing interferometric data-stacks: SqueeSAR. *IEEE Trans. Geosci. Remote Sens.* 49 (9), 3460–3470. <https://doi.org/10.1109/TGRS.2011.2124465>.
- Ferretti, A., Prati, C., Rocca, F., 2000. Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Trans. Geosci. Remote Sens.* 38 (5), 2202–2212. <https://doi.org/10.1109/36.868878>.
- Ferretti, A., Prati, C., Rocca, F., 2001. Permanent scatterers in SAR interferometry. *IEEE Trans. Geosci. Remote Sens.* 39 (1), 8–20. <https://doi.org/10.1109/36.898661>.
- Fornaro, G., Verde, S., Reale, D., Paucillo, A., 2015. CAESAR: An approach based on covariance matrix decomposition to improve multibaseline-multitemporal interferometric SAR processing. *IEEE Trans. Geosci. Remote Sens.* 53 (4), 2050–2065. <https://doi.org/10.1109/TGRS.2014.2352853>.
- Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587. <https://doi.org/10.1109/CVPR.2014.81>.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: *IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 7–13 Dec. 2015, pp. 1026–1034. doi: 10.1109/ICCV.2015.123.
- Hooper, A., Zebker, H., Segall, P., Kampes, B., 2004. A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. *Geophys. Res. Lett.* 31 (23), 1–5. <https://doi.org/10.1029/2004GL021737>.

- Hooper, A., Segall, P., Zebker, H., 2007. Persistent scatterer interferometric synthetic aperture radar for crustal deformation analysis, with application to Volcán Alcedo, Galápagos. *J. Geophys. Res.* 112, B07407. <https://doi.org/10.1029/2006JB004763>.
- Imperatore, P., Pepe, A., Sansosti, E., 2021. High performance computing in satellite SAR interferometry: a critical Perspective. *Remote Sens.* 2021 (13), 4756. <https://doi.org/10.3390/rs13234756>.
- Jiang, M., Ding, X., Hanssen, R.F., Malhotra, R., Chang, L., 2015. Fast statistically homogeneous pixel selection for covariance matrix estimation for multitemporal InSAR. *IEEE Trans. Geosci. Remote Sens.* 53 (3), 1213–1224. <https://doi.org/10.1109/TGRS.2014.2336237>.
- Johnson, J.M., Khoshgoftar, T.M., 2019. Survey on deep learning with class imbalance. *J. Big Data.* 6 (1) <https://doi.org/10.1186/s40537-019-0192-5>.
- Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, Hilton San Diego, May 7–9, 2015. [Online]. Available: <http://arxiv.org/abs/1412.6980>.
- Kiranyaz, S., Ince, T., Hamila, R., Gabbouj, M., 2015. Convolutional neural networks for patient-specific ECG classification. In: 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 Aug. 2015, pp. 2608–2611. doi: 10.1109/EMBC.2015.7318926.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*. Curran Associates Inc., Lake Tahoe, pp. 1097–1105.
- Mora, O., Mallorqui, J.J., Broquetas, A., 2003. Linear and nonlinear terrain deformation maps from a reduced set of interferometric SAR images. *IEEE Trans. Geosci. Remote Sens.* 41 (10), 2243–2253. <https://doi.org/10.1109/TGRS.2003.814657>.
- Mukherjee, S., Zimmer, A., Sun, X., Ghuman, P., Cheng, I., 2021. An unsupervised generative neural approach for InSAR phase filtering and coherence estimation. *IEEE Geosci. Remote Sens.* 18 (11), 1971–1975. <https://doi.org/10.1109/LGRS.2020.3010504>.
- Osmanoğlu, B., Sunar, F., Wdowinski, S., Cabral-Cano, E., 2016. Time series analysis of InSAR data: methods and trends. *ISPRS J. Photogramm. Remote Sens.* 115, 90–102. <https://doi.org/10.1016/j.isprsjprs.2015.10.003>.
- Tiwari, A., Narayan, A.B., Dikshit, O., 2020. Deep learning networks for selection of measurement pixels in multi-temporal SAR interferometric processing. *ISPRS J. Photogramm. Remote Sens.* 166, 169–182. <https://doi.org/10.1016/j.isprsjprs.2020.06.005>.
- Wang, Y., Zhu, X.X., Bamler, R., 2012. Retrieval of phase history parameters from distributed scatterers in urban areas using very high resolution SAR data. *ISPRS J. Photogramm. Remote Sens.* 73, 89–99. <https://doi.org/10.1016/j.isprsjprs.2012.06.007>.
- Wang, Z., Li, Z., Mills, J., 2018. A new approach to selecting coherent pixels for ground-based SAR deformation monitoring. *ISPRS J. Photogramm. Remote Sens.* 144, 412–422. <https://doi.org/10.1016/j.isprsjprs.2018.08.008>.
- Zhang, Y., Wu, H., Li, M., Kang, Y., Lu, Z., 2021. Investigating ground subsidence and the causes over the whole Jiangsu Province, China using Sentinel-1 SAR data [Online]. Available: *Remote Sens.* 13, 179 <https://www.mdpi.com/2072-4292/13/2/179>.
- Zhang, Y., Wu, H., Kang, Y., Zhu, C., 2016. Ground subsidence in the Beijing-Tianjin-Hebei region from 1992 to 2014 revealed by multiple SAR stacks. *Remote Sens.* 8 (8), 675. <https://doi.org/10.3390/rs8080675>. <https://www.mdpi.com/2072-4292/8/8/675>.
- Zhao, F., Mallorqui, J.J., 2019. A temporal phase coherence estimation algorithm and its application on DInSAR pixel selection. *IEEE Trans. Geosci. Remote Sens.* 57 (11), 8350–8361. <https://doi.org/10.1109/TGRS.2019.2920536>.