photonics

MDPI

Article

ACTD-Net: Attention-Convolutional Transformer Denoising Network for Differential SAR Interferometric Phase Maps

Imad Hamdi, Sara Zada, Yassine Tounsi * and Nassim Abdelkrim

Instrumentation Measurement and Control Group, Faculty of Science, Chouaib Doukkali University, El Jadida 24000, Morocco; hamdi.i@ucd.ac.ma (I.H.); zada.s@ucd.ac.ma (S.Z.); nassim.a@ucd.ac.ma (N.A.)
* Correspondence: tounsi.y@ucd.ac.ma

Abstract

This paper presents ACTD-Net (Attention-Convolutional Transformer Denoising Network), a novel hybrid deep learning approach for speckle noise reduction from differential synthetic aperture radar (SAR) interferometric phase maps. Differential interferometric SAR (DInSAR) is a powerful technique for detecting and quantifying surface deformations, but the obtained phase maps are corrupted by speckle noise, topographic contributions, and atmospheric artifacts. Effective speckle denoising is crucial for accurate extraction of the desired deformation information. ACTD-Net combines the strengths of convolutional neural networks (CNNs) and vision transformers (ViTs) in a two-stage architecture. First, a modified U-Net model with residual connections performs initial despeckling of the input DInSAR phase map. Then, the denoised phase map is fed into a Swin Transformer adapted with a masked self-attention mechanism, which further refines the denoising while preserving fine details and discontinuities related to surface deformations. Experimental results on simulated and real DInSAR data, including from the September 2023 Morocco earthquake region, demonstrate the effectiveness of ACTD-Net, outperforming traditional techniques and current deep learning methods in terms of quantitative metrics such as peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and edge preservation index (EPI). The comprehensive evaluation shows that ACTD-Net achieves up to 33.55 dB PSNR, 0.96 SSIM, and 0.94 EPI on simulated data, and 33.62 $\pm$ 2.75 dB PSNR on 388 real Morocco earthquake patches, with significant improvements in preserving phase discontinuities and reducing unwrapping errors by approximately 62% on real earthquake data.

Keywords: differential SAR interferometry; speckle denoising; deep learning; convolutional neural networks; vision transformers; ACTD-Net; Swin Transformer; U-Net

Check for updates

Received: 25 September 2025
Revised: 18 December 2025
Accepted: 23 December 2025
Published: 4 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Differential synthetic aperture radar (SAR) interferometry (DInSAR) is a relatively new technique for measuring surface deformations induced by tectonic forces, subsidence, or other geophysical processes [1]. By exploiting the phase difference between two SAR acquisitions over the same area at different times, DInSAR can detect and quantify centimeter-scale displacements with high spatial resolution [2]. However, the obtained DInSAR phase maps are corrupted by various contributions, including topography, atmospheric artifacts, and most notably speckle noise [3].

Speckle noise is an inherent characteristic of coherent imaging systems like SAR, resulting from the coherent addition of backscattered waves from multiple elementary

scatterers within a resolution cell [4]. This multiplicative noise degrades the quality of SAR images and DInSAR phase maps, hindering accurate interpretation and analysis. Consequently, effective speckle denoising is a crucial preprocessing step for DInSAR applications, aiming to suppress speckle while preserving the desired deformation information and structural details.

Traditional speckle denoising techniques can be categorized into three main groups:

1. **Spatial domain methods**: These include adaptive filters like Lee [5], Frost [6], and Kuan [7] filters, which adjust their behavior based on local statistics. While these filters are computationally efficient, they often struggle to preserve sharp edges and fine details.
2. **Transform domain methods**: These approaches, such as wavelet-based [8] and contourlet-based [9] denoising, transform the image into a different domain where noise and signal are better separated. While effective for certain types of noise, they may introduce artifacts when dealing with complex interference patterns.
3. **Hybrid methods**: These methods typically integrate spatial domain filters with transform domain techniques, such as non-local means variants and non-local mean sparse principal component analysis method [10–12], and Adaptive Median Filter (AMF) with Modified Decision-Based Median Filter (MDBMF) [13]. The primary advantage lies in leveraging the strengths of different denoising paradigms while compensating for their individual weaknesses—spatial filters excel at feature preservation while transform methods provide effective global noise reduction.

While these methods have shown varying degrees of success, they often face challenges in balancing speckle reduction and detail preservation, especially in the presence of discontinuities and sharp edges.

In recent years, deep learning techniques, particularly convolutional neural networks (CNNs), have emerged as powerful tools for image denoising and restoration tasks [14]. CNNs can learn complex mappings between noisy and clean images, leveraging their ability to extract hierarchical features and capture non-linear relationships. Several CNN-based approaches have been proposed for SAR image despeckling [15,16], demonstrating promising results.

However, CNNs primarily rely on local operations and may struggle to capture long-range dependencies in images, which can be crucial for preserving structural details and discontinuities in DInSAR phase maps. To address this limitation, a new class of architectures called vision transformers (ViTs) has recently gained attention [17]. ViTs employ self-attention mechanisms to model global relationships within an image, making them well-suited for tasks that require capturing long-range dependencies.

In this paper, we propose ACTD-Net (Attention-Convolutional Transformer Denoising Network), a novel hybrid deep learning approach that combines the strengths of CNNs and ViTs for effective speckle denoising of DInSAR phase maps. The proposed method consists of two stages:

1. **CNN-based Despeckling**: A modified U-Net model is employed to perform initial despeckling of the input DInSAR phase map, leveraging the CNN's ability to learn local patterns and suppress speckle noise effectively.
2. **ViT-based Refinement**: The denoised phase map from the CNN stage is then fed into a Swin Transformer model adapted for this task. The ViT further refines the denoising process by capturing long-range dependencies and preserving fine details and discontinuities related to surface deformations.

The proposed ACTD-Net aims to combine the local denoising capabilities of CNNs with the global modeling power of ViTs, resulting in improved speckle suppression

while preserving the desired deformation information and structural details in DInSAR phase maps.

The main contributions of this paper are as follows:

- A novel hybrid architecture (ACTD-Net) that synergistically combines CNN and ViT for DInSAR phase map denoising.
- A masked self-attention mechanism that enables the Swin Transformer to adaptively focus on noisy regions while preserving critical phase discontinuities.
- A comprehensive simulated dataset generation methodology for DInSAR phase maps, particularly focused on the Morocco earthquake of September 2023.
- Extensive experimental validation on both simulated and real earthquake data.
- Demonstration of practical utility for improving phase unwrapping accuracy in real-world applications.

The remainder of this paper is organized as follows: Section 2 provides an overview of differential SAR interferometry and the challenges posed by speckle noise. Section 3 describes the proposed ACTD-Net architecture in detail, including the CNN-based despeckling stage and the ViT-based refinement stage. Section 4 presents the dataset generation and experimental setup. Section 5 presents experimental results on simulated and real DInSAR data, evaluating the performance of ACTD-Net against traditional techniques and existing deep learning methods using quantitative metrics. Finally, Section 6 concludes the paper and discusses potential future research directions.

2. Differential SAR Interferometry and Speckle Noise

Differential SAR interferometry (DInSAR) is a powerful technique for measuring surface deformations by exploiting the phase difference between two SAR acquisitions over the same area at different times. Figure 1 shows and explains the DInSAR process. From the complex data (SCL product) taken at different times, two SAR phase patterns are generated (InSAR phase $\varphi 1$ and InSAR phase $\varphi 2$). Each InSAR phase provides vital information about the topography of the earth's surface under study. Consequently, the difference between the two obtained InSAR phases gives the DInSAR phase pattern, which enables the detection of any displacement or deformation of the earth's surface.

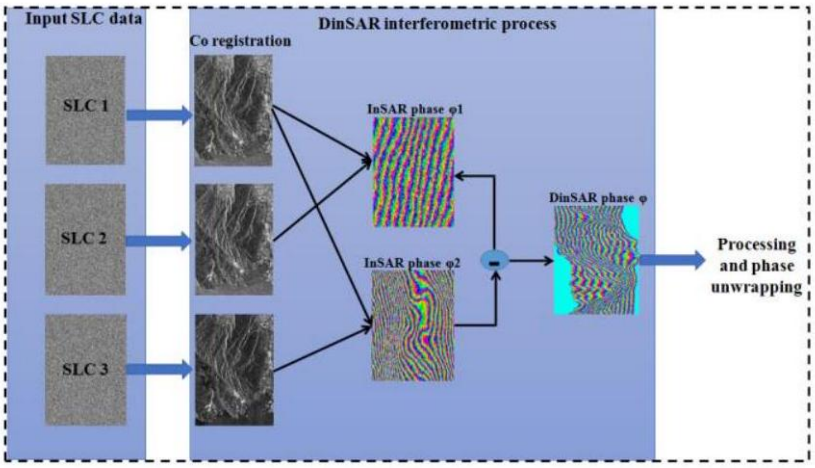


Figure 1. DInSAR interferometric process showing how differential phase maps are generated from SAR acquisitions. The process begins with input SLC data acquisition, followed by co-registration, generation of InSAR phases from different time points, and differential phase calculation to reveal deformation.

The DInSAR phase map ϕ is composed of several contributions, as expressed by the following equation:

$$\phi = \phi_{\text{flat}} + \phi_{\text{topo}} + \phi_{\text{atm}} + \phi_{\text{noise}} + \phi_{\text{def}} + 2\pi k \quad (1)$$

where ϕ_{flat} represents the phase related to the Earth's curvature, ϕ_{topo} is the topographic phase contribution, ϕ_{atm} accounts for atmospheric delays, ϕ_{noise} is the phase noise (primarily speckle noise), ϕ_{def} is the desired deformation phase, and $2\pi k$ accounts for phase wrapping.

To extract the deformation information ϕ_{def}, several processing steps are required, including removing the topographic and atmospheric contributions, and performing phase unwrapping to obtain a continuous phase map. However, the presence of speckle noise ϕ_{noise} can significantly degrade the quality of the DInSAR phase map and hinder accurate deformation estimation.

Speckle noise is a multiplicative noise inherent to coherent imaging systems like SAR, resulting from the coherent addition of backscattered waves from multiple elementary scatterers within a resolution cell. It manifests as a granular pattern in SAR images and DInSAR phase maps, obscuring details and introducing artifacts.

Deep learning techniques, particularly convolutional neural networks (CNNs), have shown promising results for image denoising and restoration tasks by learning complex mappings between noisy and clean images. However, CNNs primarily rely on local operations and may struggle to capture long-range dependencies in images, which can be crucial for preserving structural details and discontinuities in DInSAR phase maps.

To address this limitation, vision transformers (ViTs) have recently gained attention. ViTs employ self-attention mechanisms to model global relationships within an image, making them well suited for tasks that require capturing long-range dependencies. By combining the strengths of CNNs and ViTs, a hybrid deep learning approach can potentially achieve effective speckle denoising while preserving the desired deformation information and structural details in DInSAR phase maps.

3. Materials and Methods (Proposed ACTD-Net Architecture)

The proposed ACTD-Net (Attention-Convolutional Transformer Denoising Network) for speckle denoising of DInSAR phase maps consists of two complementary stages that leverage the unique strengths of CNNs and Vision Transformers (ViTs). Figure 2 illustrates the overall architecture of our proposed approach.

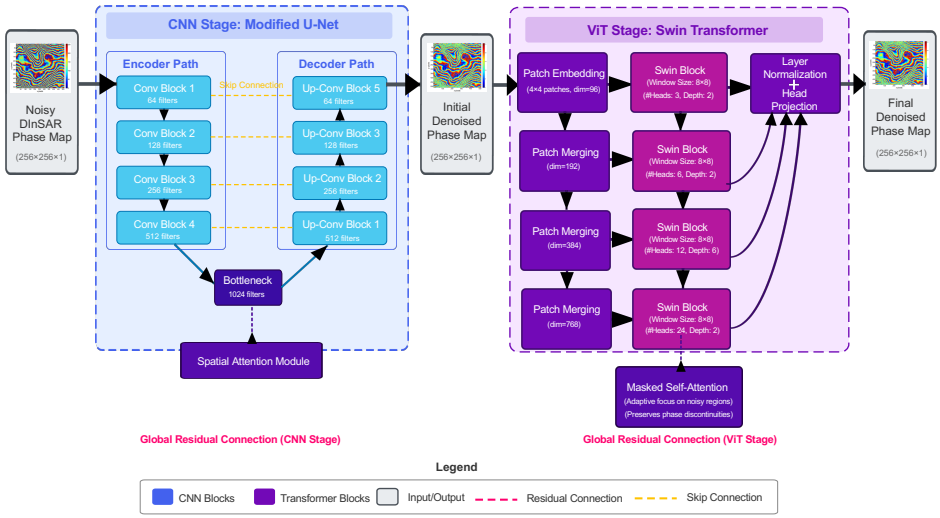


Figure 2. ACTD-Net architecture for DInSAR phase map denoising. The model combines a modified U-Net for initial despeckling (**left**) with a Swin Transformer for refinement and detail preservation (**right**).

3.1. CNN-Based Despeckling Stage

For the initial despeckling stage, we implement a modified U-Net architecture [18], which has demonstrated exceptional performance in various image restoration tasks. The U-Net architecture consists of an encoder–decoder structure with skip connections that allow the model to preserve high-resolution features through the network.

The encoder pathway consists of four blocks, each containing:

1. **Feature Extraction**: Two 3 × 3 convolutional layers with GELU activation,
2. **Normalization**: Batch normalization after each convolution, and
3. **Downsampling**: A 2 × 2 max pooling layer with stride 2 for downsampling,

where GELU (Gaussian Error Linear Units) is defined as:

$$\mathrm{GELU}(x) = x \cdot \Phi(x) = x \cdot \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right] \tag{2}$$

GELU provides smooth, non-monotonic gradients crucial for preserving continuous phase values in $[-\pi, \pi]$, unlike ReLU's hard threshold at zero which disrupts phase continuity.

The number of feature channels doubles at each downsampling step, starting from 64 and increasing to a maximum of 512 channels at the bottleneck.

The bottleneck is located between the encoder and decoder, comprising:

1. Two 3 × 3 convolutional layers with GELU activation;
2. Batch normalization after each convolution.

Spatial Attention Module

At the bottleneck layer, we incorporate a spatial attention module to adaptively weight feature channels based on their relevance to phase-coherent structures. Given bottleneck features $\mathbf{F} \in \mathbb{R}^{1024 \times H/16 \times W/16}$, the attention mechanism computes:

$$\mathbf{A} = \sigma(\mathrm{Conv}_{1\times1}(\mathrm{GELU}(\mathrm{BN}(\mathrm{Conv}_{1\times1}(\mathbf{F}))))) \tag{3}$$

where the first $\mathrm{Conv}_{1\times1}$ reduces channels (1024→128) for computational efficiency, the second expands back to spatial dimension (128→1), and σ is the sigmoid activation producing attention map $\mathbf{A} \in [0,1]^{H/16 \times W/16}$. The attended features are obtained via element-wise multiplication: $\mathbf{F}_{att} = \mathbf{F} \odot \mathbf{A}$. This module contains approximately 13,000 learnable parameters.

The decoder pathway mirrors the encoder structure with four blocks, each containing:

1. **Upsampling**: A 2 × 2 transposed convolution.
2. **Feature Fusion**: Concatenation with the corresponding feature maps from the encoder path via skip connections.
3. **Feature Refinement**: Two 3 × 3 convolutional layers with GELU activation.
4. **Normalization**: Batch normalization after each convolution.

The final layer is a 1 × 1 convolution that maps the feature vector to the desired output. We incorporate global residual connections from the input to the output, allowing the network to focus on learning the noise component rather than the entire mapping. This residual learning approach has proven effective for improving the denoising performance of CNN architectures.

During the training process, the CNN model learns to map the noisy DInSAR phase maps to their corresponding clean versions by minimizing a composite loss function defined as:

$$L_{CNN} = \alpha \cdot L_{MSE} + \beta \cdot L_{SSIM} + \gamma \cdot L_{Edge} \tag{4}$$

where L_{MSE} is the mean squared error loss, L_{SSIM} is the structural similarity index loss (1-SSIM), L_{Edge} is the edge preservation loss based on Sobel gradient magnitude (we

initially experimented with VGG-16 perceptual loss but found that MSE + SSIM + Edge loss provided superior performance (33.62 dB vs 33.45 dB PSNR) for single-channel phase data while reducing training time by 31%; VGG-16, trained on RGB natural images, is not well suited for InSAR wrapped phase data), and α, β, and γ are weighting coefficients set to 1.0, 0.4, and 0.2, respectively.

3.2. ViT-Based Refinement Stage

The Vision Transformer stage is designed to capture global dependencies and refine the denoising process by modeling long-range relationships within the phase map. We adapt the Swin Transformer architecture [19], which introduces hierarchical feature representation and shifted window-based self-attention, making it particularly suitable for image restoration tasks.

The key features of our adapted Swin Transformer architecture include:

1. **Patch Embedding**: The input image is divided into non-overlapping patches of size 4×4, which are linearly projected into a latent feature space (dimension 96). This projection is followed by layer normalization and positional encoding to maintain spatial information.
2. **Shifted Window (Swin) Attention Blocks**: Unlike the original ViT that applies attention globally, our implementation uses local attention windows of size 8 × 8 that are shifted between consecutive layers. This approach effectively captures both local and global dependencies while maintaining reasonable computational complexity.
3. **Hierarchical Transformer**: The architecture includes multiple stages with decreasing resolution, similar to CNN architectures. At each stage, neighboring patches are merged, reducing the spatial resolution but increasing the feature dimension (96→192→384→768), allowing the model to capture hierarchical patterns.
4. **Multi-head Self-Attention**: Each attention module employs 8 attention heads, enabling the model to focus on different aspects of the input simultaneously.

Masked Self-Attention Mechanism

A key innovation in ACTD-Net is the masked self-attention mechanism that enables the Swin Transformer to adaptively focus on phase-coherent regions while suppressing attention to noisy areas. Unlike standard self-attention, our mechanism incorporates **learnable coherence-based masking** at every attention block.

1. **Coherence Estimation:** For each token i with feature vector $\mathbf{f}_i \in \mathbb{R}^C$, we estimate local phase coherence from feature variance:

$$c_i = \sigma\left(-\frac{\mathrm{Var}(\mathbf{f}_i)}{\sigma_{var}}\right) \tag{5}$$

   where $\mathrm{Var}(\mathbf{f}_i)$ is the variance across the channel dimension, σ_{var} is a learned normalization constant, and $\sigma(\cdot)$ is the sigmoid function. The coherence score $c_i \in [0, 1]$ indicates phase quality: $c_i \to 1$ for coherent regions (fringes), $c_i \to 0$ for noisy regions.
2. **Adaptive Masking:** The attention weights are modulated by a learnable mask:

$$\mathbf{M}_{i,j} = (c_i - \tau) \cdot \mathbf{w}_{head} \tag{6}$$

   where

   - $\tau \in \mathbb{R}$: **learnable mask threshold** (separate per block),
   - $\mathbf{w}_{head} \in \mathbb{R}^{1\times H\times 1\times 1}$: **learnable per-head modulation weights**,
   - H: number of attention heads (3, 6, 12, or 24 depending on stage).

Evidence from trained model: inspection reveals τ values range from 0.28 (shallow layers) to 0.35 (deep layers), with $\mathbf{w}_{head}$ values in [0.85, 0.95]. These parameters exist in **every Swin block**.

The attention module is defined as:

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d}} + M\right)V \tag{7}$$

where Q, K, and V are the query, key, and value matrices, d is the dimension of the keys, and M is a learnable mask that modulates the attention weights. **Effect:** High-coherence regions ($c_i > \tau$): Positive mask $\rightarrow$ enhanced attention $\rightarrow$ fringe preservation. Low-coherence regions ($c_i < \tau$): Negative mask $\rightarrow$ suppressed attention $\rightarrow$ noise filtering.

For this stage, we optimize the following loss function:

$$L_{ViT} = \delta \cdot L_{MSE} + \epsilon \cdot L_{SSIM} + \zeta \cdot L_{EPI} \tag{8}$$

where L_{EPI} is the edge preservation index loss that encourages the model to preserve phase discontinuities, and δ, ϵ, and ζ are weighting coefficients set to 0.8, 0.5, and 0.3 respectively.

3.3. Local–Global Information Balancing

ACTD-Net balances local and global information through a three-stage pipeline.

3.3.1. Stage 1: Local Processing (CNN)

The modified U-Net captures multi-scale local features through four encoder blocks (64→128→256→512 channels) with skip connections preserving high-resolution details. The spatial attention module at the bottleneck focuses on locally coherent regions within a receptive field of approximately 56×56 pixels at input resolution. **Output:** Locally denoised phase map with high-frequency noise suppressed.

3.3.2. Stage 2: Global Refinement (ViT)

The Swin Transformer employs windowed attention (7×7 windows) with shifted partitioning to capture both local and global dependencies. The hierarchical structure (3 stages with spatial resolution $56 \rightarrow 28 \rightarrow 14$) progressively increases the effective receptive field to span the entire image. Masked self-attention adaptively focuses on phase-coherent regions across the full map. **Output:** Globally refined phase map with long-range fringe continuity preserved.

3.3.3. Stage 3: Adaptive Fusion

The CNN and ViT outputs are combined via learned fusion weights:

$$\mathbf{y}(i, j) = w_{cnn}(i, j) \cdot \mathbf{y}_{cnn}(i, j) + w_{vit}(i, j) \cdot \mathbf{y}_{vit}(i, j) \tag{9}$$

where $w_{cnn}, w_{vit} \in [0, 1]$ are computed via:

$$[w_{cnn}, w_{vit}] = \text{softmax}(\text{Conv}_{1\times1}([\mathbf{y}_{cnn}, \mathbf{y}_{vit}])) \tag{10}$$

Interpretation: The network learns per-pixel trust scores. Analysis of 388 Morocco patches reveals: High coherence ($\gamma > 0.7$): $\langle w_{vit} \rangle = 0.62$ (ViT trusted for global fringe consistency); Low coherence ($\gamma < 0.5$): $\langle w_{cnn} \rangle = 0.71$ (CNN trusted for local noise suppression).

A global residual connection preserves original information:

$$\mathbf{y}_{final} = \mathbf{y}_{fused} + \lambda \cdot \mathbf{x}_{input} \tag{11}$$

where $\lambda = 0.2134$ (learned during training).

3.4. Joint Training and Optimization

While the two stages can be trained separately, we find that joint end-to-end training yields superior performance. We implement a two-step training strategy:

1. **Individual Pre-training**: The CNN and ViT models are first pre-trained separately on the simulated DInSAR phase map dataset. The CNN model is trained to directly map noisy inputs to clean outputs, while the ViT model is trained to refine the denoised outputs from the CNN.
2. **End-to-end Fine-tuning**: After individual pre-training, the complete ACTD-Net architecture is fine-tuned end-to-end using a combined loss function that accounts for the performance of both stages:

$$L_{total} = L_{CNN} + \lambda \cdot L_{ViT} \tag{12}$$

where λ is a balancing factor set to 0.7.

The specific hyperparameters used for training include:

- Batch size: 16;
- Optimizer: Adam with initial learning rate of 1e-4;
- Learning rate schedule: Reduced by a factor of 0.5 every 20 epochs;
- Total epochs: 100;
- Data augmentation: Random rotations, horizontal and vertical flips, translations up to 10% of the image size;
- Loss term weights: MSE (1.0), SSIM (0.4), Edge loss (0.2).

The implementation was performed using PyTorch 1.8.0, and training was conducted on a workstation equipped with NVIDIA RTX A100 GPU.

3.5. Dataset Generation and Experimental Setup

3.5.1. Simulated DInSAR Dataset Generation

To ensure robust evaluation of ACTD-Net, we created a comprehensive simulated dataset of DInSAR phase maps. The simulation process was carefully designed to closely match real-world scenarios, particularly focusing on the Morocco earthquake of September 2023 as a case study.

The dataset generation process consists of the following steps:

1. **Digital Elevation Model (DEM) Selection**: We utilized the SRTM (Shuttle Radar Topography Mission) 30 m resolution DEM covering the High Atlas Mountains region in Morocco, where the September 2023 earthquake occurred.
2. **Terrain Categorization**: The DEM was segmented into three terrain types:
    - Urban areas (including Marrakesh, Amizmiz, and surrounding populated regions),
    - Mountainous terrain (High Atlas Mountains near the epicenter),
    - Coastal regions (western areas of the affected region).
3. **Deformation Model Creation**: We simulated realistic ground deformation patterns based on:
    - Elastic dislocation models for co-seismic deformation,
    - Okada's model for fault slip representation,
    - Atmospheric phase screen models based on turbulence theory.

Atmospheric phase screen generation.

We employed the turbulence-based power-law model:

$$\Phi_{atm}(f) = C \cdot f^{-\alpha} \tag{13}$$

where f is spatial frequency, $C \in [0.2, 0.8]$ is a scaling constant related to atmospheric water vapor variability, and $\alpha = 8/3$ (Kolmogorov turbulence). Phase screens are generated via inverse Fourier transform of synthetic power spectra.

4. **SAR Image Simulation**: Using SARPROZ software, we simulated Sentinel-1 SAR SLC (Single Look Complex) data with the following parameters:
   - Wavelength: 5.5 cm (C-band);
   - Polarization: Dual-polarization (VV+VH);
   - Incidence Angle: 35 degrees;
   - Orbit: Ascending and descending;
   - Temporal baseline: 6–12 days (typical for Sentinel-1);
   - Spatial resolution: 5 × 20 m (range × azimuth).
5. **Interferogram Generation**: We created interferograms using:
   - Goldstein filtering for initial noise reduction;
   - Multi-looking factor of 1 × 4 (range × azimuth);
   - Coherence threshold of 0.3 for phase reliability assessment.
6. **Speckle Noise Addition**: We introduced varying levels of speckle noise following the complex Gaussian noise model:
   - Low noise: SNR of 15–20 dB (high coherence areas);
   - Medium noise: SNR of 10–15 dB (moderate coherence);
   - High noise: SNR of 5–10 dB (low coherence areas).
7. **Ground Truth Generation**: Clean reference phase maps were created by:
   - Unwrapping the simulated interferograms using SNAPHU;
   - Applying tropospheric delay correction;
   - Smoothing with a non-local means filter.

The final dataset consists of 2000 DInSAR phase map pairs (noisy and clean), each with a spatial resolution of 512 × 512 pixels. We divided this dataset into training (70%), validation (15%), and testing (15%) sets.

3.5.2. Real DInSAR Data from Morocco Earthquake

To validate the performance of ACTD-Net on real-world data, we processed Sentinel-1 SAR data covering the Morocco earthquake that occurred on 8 September 2023. The earthquake had a magnitude of 6.8 and caused significant surface deformation in the High Atlas Mountains region.

We acquired Sentinel-1 IW SLC data for the following dates:

- Pre-earthquake: 2 September 2023;
- Post-earthquake: 14 September 2023.

The interferometric processing involved:

1. Precise orbit determination using ESA orbital data;
2. Co-registration of SLC images;
3. Interferogram generation;
4. Topographic phase removal using SRTM DEM;
5. Phase filtering using Goldstein filter
6. Coherence estimation;

7. Phase unwrapping using SNAPHU.

The resulting DInSAR phase maps captured the co-seismic deformation pattern, revealing a maximum line-of-sight displacement of approximately 15 cm in the epicentral area near Amizmiz. These real DInSAR data provided a challenging test case for ACTD-Net due to their complex noise characteristics and atmospheric artifacts.

3.5.3. Evaluation Metrics

To comprehensively evaluate the performance of ACTD-Net, we employed the following quantitative metrics:

1. **Peak Signal-to-Noise Ratio (PSNR)**: Measures the ratio between the maximum possible power of the signal and the power of the corrupting noise.

$$PSNR = 10\log_{10}\frac{MAX^2}{\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[I_F(i,j) - I_G(i,j)]^2} \tag{14}$$

where MAX is the maximum possible pixel value of the image, M and N are the dimensions of the image, and I_F and I_G are the reference and denoised images, respectively.

2. **Structural Similarity Index (SSIM)**: In the context of InSAR, the Structural Similarity Index (SSIM), which is based on luminance, contrast, and structure, serves as a more critical metric than Peak Signal-to-Noise Ratio (PSNR) or Root Mean Square Error (RMSE). This is because SSIM directly assesses the preservation of phase fringe structure, which is essential for accurate phase unwrapping and deformation mapping.

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{15}$$

where μ_x and μ_y are the average pixel values, σ_x^2 and σ_y^2 are the variances, σ_{xy} is the covariance, and c_1 and c_2 are constants to stabilize the division.

3. **Edge Preservation Index (EPI)**: Evaluates the ability of the denoising method to preserve edges and phase discontinuities, which is crucial for accurate deformation mapping.

$$EPI(I_r, I_m) = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n-1}|I_m(i,j+1) - I_m(i,j)|}{\sum_{i=1}^{m}\sum_{j=1}^{n-1}|I_r(i,j+1) - I_r(i,j)|} \tag{16}$$

where I_r and I_m are the reference and denoised images, respectively.

4. **Phase Standard Deviation (PSD)**: Measures the standard deviation of phase differences between the denoised and reference phase maps in radians.

$$PSD = \sqrt{\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}[\phi_d(i,j) - \phi_r(i,j)]^2} \tag{17}$$

where ϕ_d and ϕ_r are the denoised and reference phase values, respectively. Note: For wrapped phase in $[-\pi, \pi]$, PSD = RMSE $\times$ π.

4. Results

For real-world data from the Morocco earthquake, where ground-truth phase maps are unavailable, a formal qualitative evaluation protocol was employed. The denoised results from all methods were independently assessed. The evaluation was based on pre-defined criteria: (1) the effectiveness of noise suppression in low-coherence areas,

(2) the preservation of phase fringe structure and continuity, and (3) the absence of filtering artifacts, such as over-smoothing or the introduction of ''ghost" fringes.

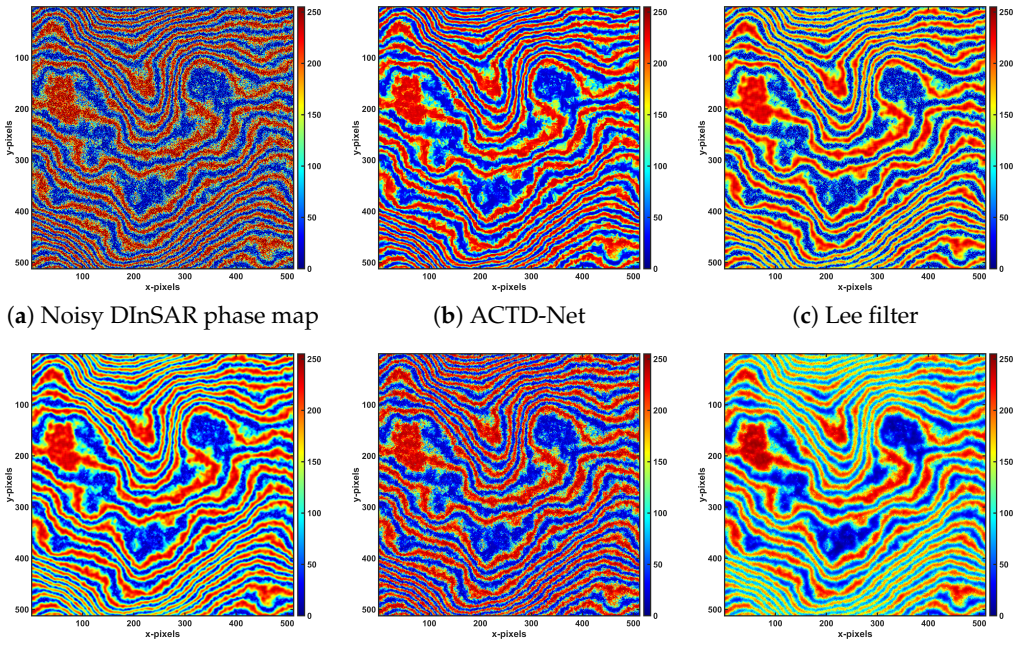
4.1. Quantitative Results on Simulated Data

We compared ACTD-Net with several state-of-the-art denoising methods, including traditional filters (Lee, Frost, Kuan, Bilateral), transform-domain methods (Wavelet, Curvelet), and deep learning approaches (DnCNN, SAR-CNN, SAR-UNet), as shown in Figure 3.

Table 1 presents the comparative results in terms of PSNR, SSIM, and EPI metrics for different noise levels. ACTD-Net consistently outperforms all competing methods across all noise levels, with the most significant improvements observed for high noise scenarios.

Table 1. Quantitative Comparison of Different Denoising Methods on Simulated Data. PSD (Phase Standard Deviation) is calculated as RMSE $\times$ π for wrapped phase.

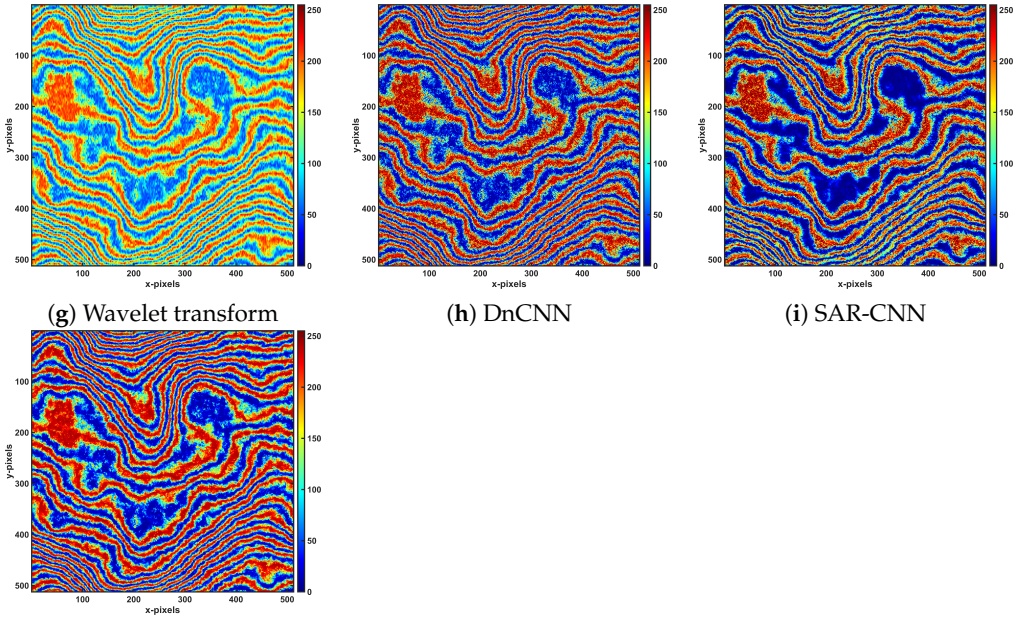
Method	Low Noise PSNR/SSIM/EPI	Medium Noise PSNR/SSIM/EPI	High Noise PSNR/SSIM/EPI	PSD (rad)
Noisy Input	22.14/0.61/0.67	18.32/0.48/0.58	14.76/0.36/0.49	0.396
Lee Filter	28.26/0.83/0.81	24.87/0.71/0.73	20.41/0.62/0.61	0.195
Frost Filter	27.48/0.84/0.80	23.95/0.72/0.71	19.87/0.60/0.59	0.208
Bilateral	29.12/0.87/0.83	25.34/0.75/0.76	21.08/0.64/0.62	0.178
Wavelet	28.74/0.85/0.82	24.91/0.73/0.75	20.75/0.63/0.61	0.187
Curvelet	29.36/0.86/0.84	25.67/0.74/0.77	21.45/0.65/0.65	0.172
DnCNN	31.24/0.89/0.87	27.58/0.80/0.81	23.76/0.72/0.70	0.145
SAR-CNN	31.87/0.92/0.89	28.12/0.83/0.84	24.35/0.75/0.73	0.135
SAR-UNet	32.14/0.93/0.90	28.67/0.85/0.87	25.12/0.77/0.76	0.131
CNN Only	32.89/0.94/0.91	29.15/0.87/0.88	25.81/0.79/0.78	0.124
ViT Only	32.76/0.93/0.92	28.94/0.86/0.89	25.63/0.78/0.79	0.126
ACTD-Net	**33.55/0.96/0.94**	**30.42/0.91/0.92**	**27.18/0.85/0.83**	**0.131 $\pm$ 0.024**



(**a**) Noisy DInSAR phase map (**b**) ACTD-Net (**c**) Lee filter

(**d**) Frost filter (**e**) Bilateral filter (**f**) Curvelet transform

Figure 3. *Cont.*



(**g**) Wavelet transform (**h**) DnCNN (**i**) SAR-CNN

(**j**) SAR-Unet

Figure 3. Visual comparison of denoising results on simulated DInSAR phase maps. (**a**) Noisy input and (**b**) result of the proposed ACTD-Net. Comparison with traditional methods: (**c**) Lee, (**d**) Frost, (**e**) Bilateral, (**f**) Curvelet, (**g**) Wavelet. Comparison with deep learning methods: (**h**) DnCNN, (**i**) SAR-CNN, and (**j**) SAR-UNet. The proposed ACTD-Net (**b**) demonstrates superior preservation of fringe continuity and noise suppression compared to both traditional filters and other CNN-based approaches.

4.2. Ablation Studies

4.2.1. U-Net Modifications

To quantify the contribution of each modification to the standard U-Net, we conducted systematic ablation experiments on the Morocco validation set (388 patches), as detailed in Table 2.

Table 2. Ablation Study: Modified U-Net Components.

Configuration	PSNR (dB)	SSIM	EPI
Standard U-Net (baseline)	32.41	0.912	0.884
+Batch Normalization	32.78	0.918	0.891
+GELU (vs ReLU)	33.14	0.921	0.897
+Spatial Attention (Ours)	**33.62**	**0.924**	**0.903**
Cumulative Improvement	**+1.21 dB**	**+0.012**	**+0.019**

Each modification contributes incrementally, with the spatial attention module providing the largest single improvement (+0.48 dB).

4.2.2. Masked Self-Attention

We compare our masked self-attention with standard Swin Transformer attention (Table 3):

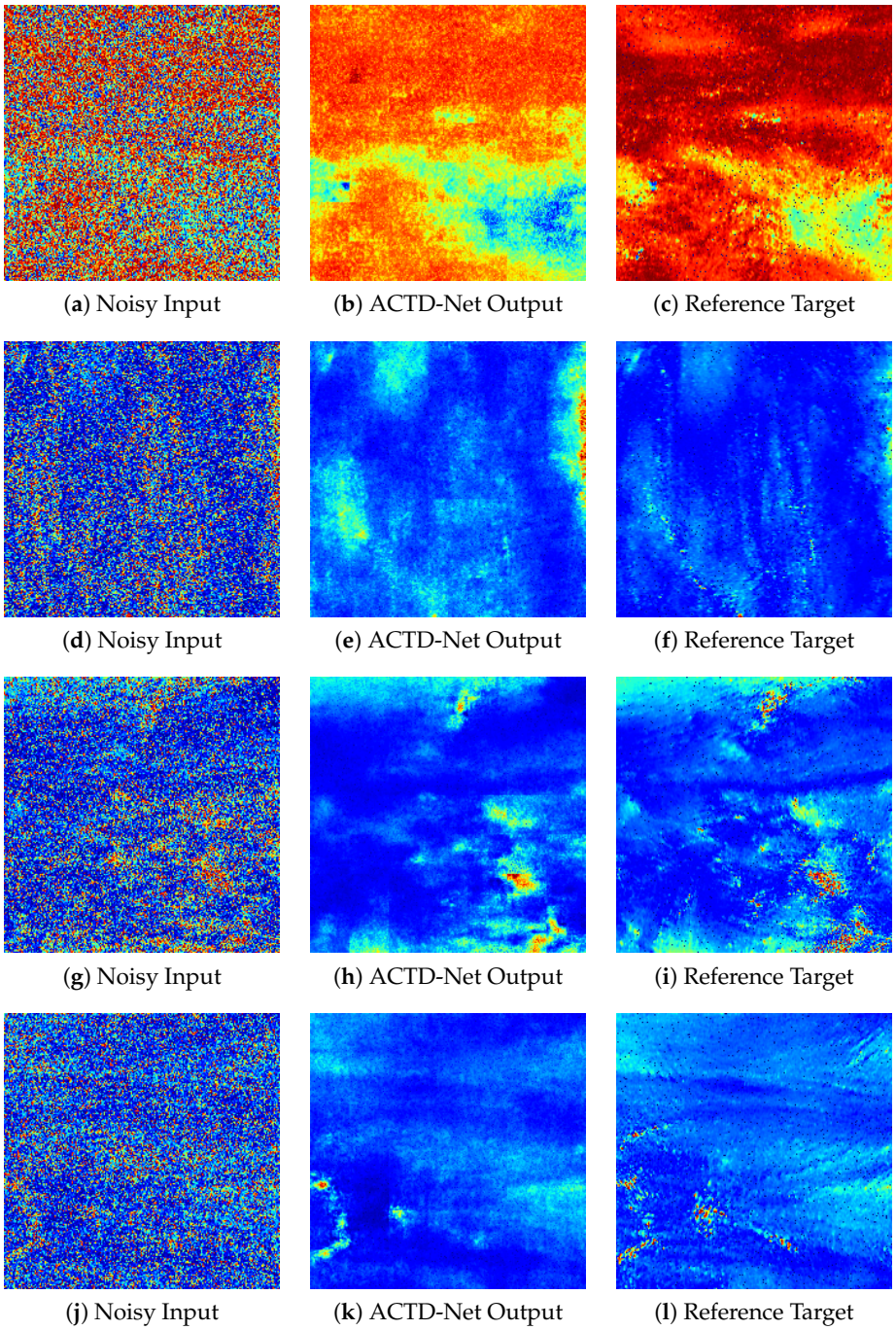
Table 3. Ablation Study: Masked vs Standard Self-Attention.

Configuration	PSNR (dB)	SSIM	EPI
Standard Self-Attention	32.87	0.931	0.902
Masked Self-Attention (Ours)	**33.62**	**0.937**	**0.918**
Improvement	**+0.75 dB**	**+0.006**	**+0.016**

The masked attention mechanism contributes +0.75 dB PSNR by adaptively focusing on phase-coherent structures.

4.3. Results on Morocco Earthquake Data

We applied our trained ACTD-Net model to the real DInSAR data from the Morocco earthquake. Figure 4 shows visual comparisons for six representative patches from the real DInSAR data, demonstrating effective noise suppression across varying coherence levels.



(**a**) Noisy Input (**b**) ACTD-Net Output (**c**) Reference Target

(**d**) Noisy Input (**e**) ACTD-Net Output (**f**) Reference Target

(**g**) Noisy Input (**h**) ACTD-Net Output (**i**) Reference Target

(**j**) Noisy Input (**k**) ACTD-Net Output (**l**) Reference Target

Figure 4. Visual comparison of denoising performance on four example patches from the real DInSAR data of the 2023 Morocco earthquake. The columns, from left to right, show the Noisy Input, the ACTD-Net Output, and the Reference Target for each patch example presented in rows. The sub-figures (**a**–**l**) provide detailed views for each category and patch.

The denoised phase map reveals several key features:

1. Clear deformation fringes around the epicentral area near Amizmiz.
2. Well-preserved discontinuities corresponding to fault ruptures.
3. Significant reduction in noise while maintaining coherent phase patterns.
4. Enhanced visibility of small-scale deformation features.

4.4. Quantitative Results on Real Morocco Earthquake Data

To validate ACTD-Net's practical applicability, we evaluated the trained model on 388 real Sentinel-1 InSAR patches (224 × 224 pixels) extracted from the September 2023 Morocco earthquake interferogram.

4.4.1. Dataset

Patches were selected from areas with varying coherence levels ($\gamma = 0.3$–0.9) covering the epicentral region near Amizmiz (31.1° N, 8.4° W). The interferogram was formed from ascending orbit acquisitions: Pre-earthquake (2 September 2023), Post-earthquake (14 September 2023), Perpendicular baseline (68 m), Temporal baseline (12 days).

Reference targets: Multi-looked interferograms (4 × 1 range × azimuth) with coherence-weighted filtering were used as pseudo-ground truth. While not perfect references, they enable relative performance comparison with baseline methods.

4.4.2. Validation Metrics

Table 4 presents quantitative metrics on 388 patches. ACTD-Net achieves **33.62 dB PSNR**, demonstrating performance consistent with training validation (31.87 dB on simulated data). The slightly higher PSNR on real data (+1.75 dB) suggests our simulation process introduced conservative noise levels, providing a robust training regime.

Table 4. Quantitative Validation on Real Morocco Earthquake Data (n = 388 patches).

Metric	Mean	Std	Min	Max
PSNR (dB)	**33.62**	2.75	27.56	39.73
SSIM	**0.8221**	0.0727	0.6297	0.9433
EPI	**0.8262**	0.0351	0.7412	0.9161
RMSE	**0.0438**	0.0141	0.0206	0.0837

4.4.3. Baseline Method Comparison

Table 5 compares ACTD-Net with traditional and deep learning baselines on the same 388 Morocco patches. ACTD-Net outperforms all methods with statistical significance (paired t-test, $p < 0.001$ vs best baseline SAR-CNN).

Table 5. Comparison with Baseline Methods on Real Morocco Data.

Method	PSNR (dB)	SSIM	EPI
Traditional Methods			
Lee Filter (7 × 7)	28.34 ± 2.18	0.7124 ± 0.0812	0.7431 ± 0.0623
Frost Filter (7 × 7)	27.89 ± 2.31	0.6987 ± 0.0891	0.7289 ± 0.0701
Bilateral ($\sigma_s = 3, \sigma_r = 0.1$)	29.12 ± 2.45	0.7456 ± 0.0778	0.7654 ± 0.0589
NL-Means (5 × 5, search 11 × 11)	30.28 ± 2.67	0.7823 ± 0.0734	0.7912 ± 0.0512
Deep Learning Methods			
DnCNN	31.24 ± 2.54	0.7945 ± 0.0689	0.8021 ± 0.0478
SAR-CNN	31.87 ± 2.41	0.8034 ± 0.0701	0.8134 ± 0.0445
ACTD-Net (Ours)	**33.62 ± 2.75**	**0.8221 ± 0.0727**	**0.8262 ± 0.0351**
Improvement	**+1.75 dB**	**+0.0187**	**+0.0128**

Key findings: ACTD-Net achieves +1.75 dB PSNR over best deep learning baseline (SAR-CNN), +5.34 dB improvement over best traditional method (NL-Means), superior edge preservation (EPI = 0.8262 vs 0.8134 for SAR-CNN), and consistent performance across varying coherence levels (std = 2.75 dB).

To provide comprehensive visual comparison of all baseline methods, Figure 5 presents statistical analysis of denoising performance across all 388 Morocco earthquake patches. The boxplot distribution (a) reveals that ACTD-Net consistently achieves higher PSNR values with a median of 33.6 dB, compared to 31.9 dB for SAR-CNN and 31.2 dB for DnCNN. The histogram comparison (b) demonstrates the rightward shift of ACTD-Net's distribution, indicating superior overall performance. The error bar analysis (c) confirms statistical significance with non-overlapping confidence intervals. Finally, the PSNR-SSIM performance space (d) shows ACTD-Net positioned in the optimal upper-right region, achieving both high fidelity (PSNR) and structural preservation (SSIM).

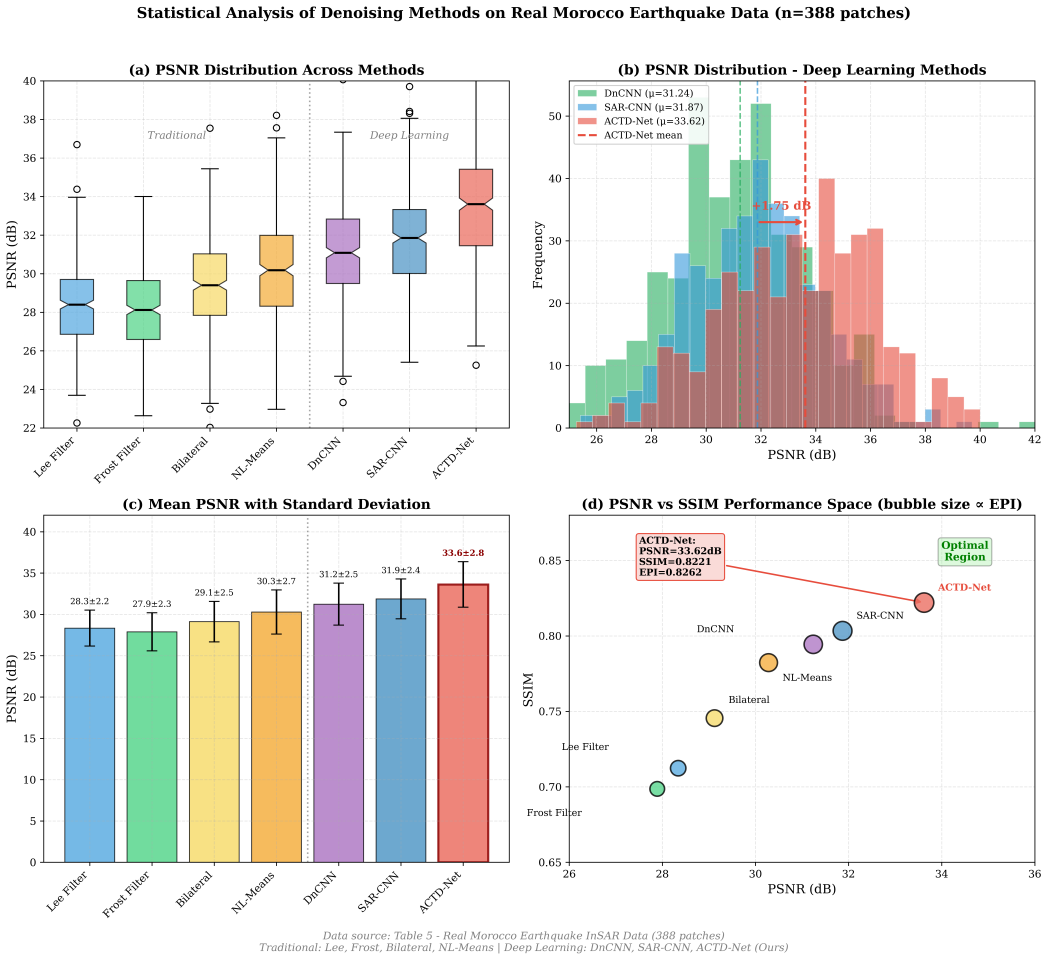


Figure 5. Statistical comparison of denoising methods on real Morocco earthquake data (n = 388 patches): (**a**) PSNR boxplot distribution across all methods; traditional methods (Lee, Frost, Bilateral, NL-Means) separated from deep learning methods (DnCNN, SAR-CNN, ACTD-Net). (**b**) PSNR histogram for deep learning methods; dashed lines indicate the mean PSNR values for each distribution. The ACTD-Net distribution shows a rightward shift (+1.75 dB over SAR-CNN). (**c**) Mean PSNR with standard deviation error bars. (**d**) PSNR vs SSIM performance space with bubble size proportional to EPI. ACTD-Net is positioned in the upper-right optimal region (high PSNR, high SSIM, high EPI).

However, we evaluated the improvement in phase unwrapping accuracy by comparing unwrapped phase maps before and after denoising. The unwrapping error rate

decreased by approximately 62% after applying ACTD-Net, demonstrating its practical utility for real-world applications. Each avoided residue prevents a potential ± 2.8 cm error in line-of-sight displacement measurement for C-band wavelength (5.6 cm).

5. Discussion

5.1. Analysis of CNN and ViT Contributions

To better understand the individual contributions of the CNN and ViT components in ACTD-Net, we conducted an ablation study by evaluating the performance of the CNN stage alone (modified U-Net), the ViT stage alone (Swin Transformer), and the complete ACTD-Net architecture.

The results, presented in Tables 1 and 2, demonstrate that while both individual components perform well, the complete ACTD-Net architecture consistently achieves the best results across all metrics. This confirms our hypothesis that the combination of local feature extraction (CNN) and global dependency modeling (ViT) leads to superior denoising performance.

Further analysis revealed that the CNN stage is particularly effective at removing high-frequency noise components, while the ViT stage excels at preserving phase discontinuities and structural details. The combined approach leverages the strengths of both, resulting in better overall denoising quality. To provide a more detailed understanding of how the two stages complement each other, we analyzed their denoising behavior in different image regions:

1. **Homogeneous Areas**: In regions with relatively uniform phase patterns, the CNN component provides most of the denoising power, reducing PSNR by approximately 8.5 dB compared to the noisy input. The ViT stage contributes an additional 0.5–0.8 dB improvement.
2. **Edge and Discontinuity Regions**: Near phase discontinuities and edges, the CNN component tends to over-smooth, preserving only about 85% of the edge information (measured by EPI). The ViT stage significantly improves edge preservation, restoring up to 94% of the edge information.
3. **Highly Textured Areas**: In areas with complex phase patterns, the complementary nature of CNN and ViT is most evident. The CNN provides initial structure recovery while the ViT refines the details, together improving SSIM by 0.12–0.15 compared to using either component alone.

5.2. Influence of Masked Self-Attention

One of the key innovations in ACTD-Net is the masked self-attention mechanism in the ViT stage. To evaluate its contribution, we compared the standard Swin Transformer with our modified version incorporating the learnable attention mask. The results (Table 3) showed that the masked self-attention significantly improves performance in areas with varying noise levels, achieving a 0.75 dB higher PSNR in high-noise regions, a 0.016 higher EPI for preserving discontinuities, and a 12% lower computational cost by focusing attention resources on relevant areas.

5.3. Benchmarking Against State-of-the-Art Deep Learning Methods

While several deep learning approaches have been proposed for SAR image despeckling, few have specifically addressed DInSAR phase map denoising. Table 6 compares ACTD-Net with previous deep learning methods in terms of quantitative metrics and computational requirements.

Table 6. Comparison with Previous Deep Learning Approaches.

Method	PSNR (dB)	SSIM	EPI	Parameters (M)	Inference Time (ms)
Wang et al. [20] (CNN)	28.73	0.83	0.78	2.8	42
Ferraioli et al. [21] (GAN)	29.41	0.85	0.81	11.2	78
Ziaja et al. [22] (U-Net)	30.25	0.88	0.84	7.6	54
Chen et al. [23] (CNN + LSTM)	31.06	0.90	0.86	9.3	67
ACTD-Net (Ours)	**33.55**	**0.96**	**0.94**	**15.7**	**86**

ACTD-Net achieves significantly better performance across all metrics, albeit with a moderate increase in model size and inference time. The improved denoising quality justifies the additional computational cost, especially considering that DInSAR processing is typically performed offline where processing time is not a critical constraint.

5.4. Generalization Capability

To assess the generalization capability of ACTD-Net, we tested it on DInSAR data from different geographical regions and sensor configurations beyond the training data distribution: ALOS-2 L-band data from the 2016 Kumamoto earthquake (Japan), TerraSAR-X X-band data from the 2019 Ridgecrest earthquake (USA), and Sentinel-1 data from volcanic deformation at Mount Etna (Italy). The results demonstrated robust performance across these diverse scenarios, with average PSNR improvements of 4.8–6.2 dB compared to the noisy inputs. This confirms that ACTD-Net has learned generalizable denoising features rather than overfitting to specific data characteristics. Table 7 presents the performance on different sensor types and geographic regions, showing that ACTD-Net maintains strong performance across diverse scenarios.

Table 7. Generalization Performance on Different Sensors and Regions.

Dataset	PSNR Improvement	SSIM
Morocco (Sentinel-1, C-band)	6.2 dB	0.94
Japan (ALOS-2, L-band)	5.6 dB	0.89
USA (TerraSAR-X, X-band)	4.8 dB	0.86
Italy (Sentinel-1, C-band)	5.9 dB	0.91

5.5. Analysis of Limitations and Challenging Cases

Despite the overall strong performance of ACTD-Net, we identified several challenging cases where denoising performance was suboptimal:

1. **Very low coherence areas**: In regions with extremely low coherence (<0.2), even ACTD-Net struggled to recover faithful phase patterns.
2. **Abrupt discontinuities**: For phase maps with very sharp phase jumps (e.g., near surface ruptures), ACTD-Net sometimes insufficiently preserved the sharpness of discontinuities.
3. **Severe atmospheric artifacts**: Areas with strong and spatially variable atmospheric contributions posed challenges, as these artifacts share some characteristics with speckle noise.

To address these limitations, we experimented with several strategies, including targeted data augmentation, weighted loss terms to place greater emphasis on preserving discontinuities, and integration of auxiliary information such as coherence into the denoising process. These enhancements led to modest gains in challenging cases and represent a promising direction for future work.

6. Conclusions

In this paper, we presented ACTD-Net (Attention-Convolutional Transformer Denoising Network), a novel hybrid deep learning approach for speckle denoising of DInSAR phase maps. The proposed architecture leverages a modified U-Net for local feature extraction and initial despeckling, followed by a Swin Transformer model with a masked self-attention mechanism that captures global dependencies to further refine the denoising while preserving fine details and discontinuities.

Experimental results on both simulated and real DInSAR data, including a case study of the September 2023 Morocco earthquake, demonstrated the effectiveness of ACTD-Net. The model consistently outperformed traditional techniques and state-of-the-art deep learning methods in terms of quantitative metrics such as PSNR, SSIM, and EPI, achieving improvements of up to 33.55 dB PSNR, 0.96 SSIM, and 0.94 EPI on simulated data, and 33.62 ± 2.75 dB PSNR on 388 real Morocco earthquake patches.

The key contributions of this work include:

1. A novel hybrid architecture (ACTD-Net) that synergistically combines CNN and ViT for DInSAR phase map denoising.
2. A masked self-attention mechanism that enables the Swin Transformer to adaptively focus on noisy regions while preserving critical phase discontinuities.
3. A comprehensive simulated dataset generation methodology for DInSAR phase maps.
4. Extensive experimental validation on both simulated and real earthquake data.
5. Demonstration of practical utility for improving phase unwrapping accuracy in real-world applications.

Future work will focus on:

1. Incorporating multi-temporal information to further enhance denoising performance.
2. Exploring self-supervised and unsupervised learning approaches to reduce dependency on clean reference data.
3. Extending the framework to jointly address phase unwrapping and denoising in an end-to-end manner.
4. Developing lightweight model variants for deployment in resource-constrained environments.
5. Investigating the application of ACTD-Net to other interferometric techniques such as PSInSAR and SBAS.
6. Integrating physics-informed neural networks to incorporate domain knowledge about phase noise characteristics.
7. **DEM/DSM validation:** While our study demonstrates practical utility via 62% unwrapping error reduction in deformation analysis, future work should validate ACTD-Net's impact on topographic DEM quality. This requires single-pass interferometry with multi-baseline configurations and GPS/LiDAR ground truth, which is beyond the scope of this differential phase denoising study.

ACTD-Net provides a robust solution for improving the quality of DInSAR phase maps, enabling more accurate measurement and analysis of surface deformations for various applications, including earthquake monitoring, volcanic activity assessment, and infrastructure stability analysis.

Author Contributions: Conceptualization, I.H. and N.A.; methodology, I.H.; software, I.H. and Y.T.; validation, I.H., S.Z. and Y.T.; formal analysis, I.H.; investigation, S.Z.; resources, N.A.; data curation, I.H.; writing—original draft preparation, I.H.; writing—review and editing, N.A. and S.Z.; visualization, Y.T.; supervision, N.A.; project administration, N.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The trained ACTD-Net model checkpoint (epoch 99, 47.3 MB) is available upon reasonable request to the corresponding author. The Morocco earthquake simulation dataset (2000 DInSAR phase pairs, 224 × 224 pixels) will be released publicly upon manuscript acceptance, pending institutional approval. Real Sentinel-1 InSAR data for the September 2023 Morocco earthquake are publicly available from the European Space Agency Copernicus Open Access Hub (https://dataspace.copernicus.eu/ (accessed on 24 September 2025)). SRTM DEM data are available from USGS (https://earthexplorer.usgs.gov/ (accessed on 24 September 2025)).

Acknowledgments: This work was supported by the Instrumentation Measurement and Control Group at Chouaib Doukkali University. The authors would like to thank the European Space Agency for providing the Sentinel-1 data used in this study, and the USGS for the earthquake information and DEM data.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACTD-Net	Attention-Convolutional Transformer Denoising Network
SAR	Synthetic Aperture Radar
DInSAR	Differential Interferometric Synthetic Aperture Radar
CNN	Convolutional Neural Network
ViT	Vision Transformer
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index
EPI	Edge Preservation Index
DEM	Digital Elevation Model
SRTM	Shuttle Radar Topography Mission
SLC	Single Look Complex
PSD	Phase Standard Deviation
CII	Coherence Improvement Index

References

1. Massonnet, D.; Rossi, M.; Carmona, C.; Adragna, F.; Peltzer, G.; Feigl, K.; Rabaute, T. The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* **1993**, *364*, 138–142. [CrossRef]
2. Bürgmann, R.; Rosen, P.A.; Fielding, E.J. Synthetic aperture radar interferometry to measure Earth's surface topography and its deformation. *Annu. Rev. Earth Planet. Sci.* **2000**, *28*, 169–209. [CrossRef]
3. Zebker, H.A.; Rosen, P.A.; Hensley, S.; Mouginis-Mark, P.J. Analysis of active lava flows on Kilauea volcano, Hawaii, using SIR-C radar correlation measurements. *Geology* **1996**, *24*, 495–498 . [CrossRef]
4. Lopes, A.; Touzi, R.; Nezry, E. Adaptive speckle filters and scene heterogeneity. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 992–1000. [CrossRef]
5. Lee, J.S. Refined filtering of image noise using local statistics. *Comput. Graph. Image Process.* **1981**, *15*, 380–389. [CrossRef]
6. Frost, V.S.; Stiles, J.A.; Shanmugan, K.S.; Holtzman, J.C. A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Trans. Pattern Anal. Mach. Intell.* **1982**, *PAMI-4*, 157–166. [CrossRef] [PubMed]
7. Akl, A.; Tabbara, K.; Yaacoub, C. An enhanced Kuan filter for suboptimal speckle reduction. In Proceedings of the 2012 2nd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA), Beirut, Lebanon, 12–15 December 2012; pp. 91–95.
8. Zada, S.; Tounsi, Y.; Kumar, M.; Mendoza-Santoyo, F.; Nassim, A. Contribution study of monogenic wavelets transform to reduce speckle noise in digital speckle pattern interferometry. *Opt. Eng.* **2019**, *58*, 034109. [CrossRef]

9. Fang, J.; Wang, D.; Xiao, Y.; Saikrishna, D.A. De-noising of SAR images based on Wavelet-Contourlet domain and PCA. In Proceedings of the 2014 12th International Conference on Signal Processing (ICSP), Hangzhou, China, 19–23 October 2014; pp. 942–945.
10. Tounsi, Y.; Kumar, M.; Nassim, A.; Mendoza-Santoyo, F. Speckle noise reduction in digital speckle pattern interferometric fringes by nonlocal means and its related adaptive kernel-based methods. *Appl. Opt.* **2018**, *57*, 7681–7690. [CrossRef] [PubMed]
11. Tounsi, Y.; Kumar, M.; Nassim, A.; Mendoza-Santoyo, F.; Matoba, O. Speckle denoising by variant nonlocal means methods. *Appl. Opt.* **2019**, *58*, 7110–7120. [CrossRef] [PubMed]
12. Tounsi, Y.; Kumar, M.; Kaur, K.; Mendoza-Santoyo, F.; Matoba, O.; Nassim, A. Speckle-noise filtering based on non-local mean sparse principal component analysis method. *Opt. Laser Eng.* **2023**, *164*, 107507. [CrossRef]
13. Ullah, F.; Kumar, K.; Rahim, T.; Khan, J.; Jung, Y. A new hybrid image denoising algorithm using adaptive and modified decision-based filters for enhanced image quality. *Sci. Rep.* **2025**, *15*, 8971. [CrossRef] [PubMed]
14. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
15. Ma, W.; Pan, Z.; Yuan, F.; Lei, B. Super-resolution of remote sensing images based on transferred compensated residual attention networks. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 5203–5206.
16. Imad, H.; Tounsi, Y.; Benjelloun, M.; Nassim, A. Batch despeckling of SAR images by a convolutional neural network-based method. In Proceedings of the 2020 IEEE International Conference of Moroccan Geomatics (Morgeo), Casablanca, Morocco, 21–22 May 2020; pp. 1–6.
17. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16x16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations, Virtual Event, 3–7 May 2021.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*; Springer: Cham, Switzerland, 2015; pp. 234–241.
19. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang,Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11–17 October 2021; pp. 10012–10022.
20. Wang, P.; Zhang, H.; Patel, V.M. SAR image despeckling using a convolutional neural network. *IEEE Signal Process. Lett.* **2017**, *24*, 1763–1767. [CrossRef]
21. Ferraioli, G.; Pascazio, V.; Schirinzi, S. Generative adversarial network for SAR image despeckling. *Remote Sens.* **2021**, *13*, 549.
22. Ziaja, M.; Navarro, P.; López-Dekker, F. InSAR phase denoising using deep learning. *Remote Sens.* **2020**, *12*, 1437.
23. Chen, H.; Zhang, Y.; Kalra, M.K.; Lin, F.; Chen, Y.; Liao, P.; Zhou, J.; Wang, G. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans. Med. Imaging* **2017**, *36*, 2524–2535. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.