Temporal Variation-Guided Self-Supervised PolSAR Despeckling Network

Shaowei Shi[1], Liupeng Lin[3], Jie Li[1,2,*], Qiangqiang Yuan[1], Huanfeng Shen[3]

[1] School of Geodesy and Geomatics, Wuhan University, Wuhan, China
[2] Hubei Luojia Laboratory, Wuhan, China
[3] School of Resource and Environmental Sciences, Wuhan University, Wuhan, China
[*] Corresponding author, E-mail address: jli89@sgg.whu.edu.cn

Keywords: PolSAR, self-supervised, despeckling, change detection, polarization decomposition.

Abstract

Deep learning–based polarimetric SAR (PolSAR) despeckling faces two main challenges: the scarcity of clean reference images and the difficulty of preserving structural details while suppressing noise. To address these issues, we propose a temporally guided self-supervised network (TGSD-Net), which generates pseudo training pairs from consecutive noisy observations and leverages a change detection–based prior to exploit temporal redundancy and enhance robustness to land-cover changes. TGSD-Net further integrates model and input feature refinements, including auxiliary polarimetric decomposition parameters and a spatiotemporal information fusion module (STIFM) based on a U-Net backbone, to improve temporal and scattering feature representations. The network is specifically designed to robustly handle multi-temporal SAR acquisitions and heterogeneous land-cover types, maintaining consistent scattering structures across different scenes. Extensive experiments on real PolSAR datasets demonstrate that TGSD-Net effectively balances noise suppression with detail preservation. Quantitative metrics, including the equivalent number of looks (ENL) and edge preservation degree (EPD), confirm its superior despeckling performance. Polarimetric decomposition analyses further verify that the network preserves the physical scattering characteristics of PolSAR images.

1. Introduction

Polarimetric synthetic aperture radar (PolSAR) provides all-weather and all-day imaging capabilities, making it an essential tool in various applications such as geological exploration, environmental monitoring, urban mapping, and post-disaster assessment. By recording the amplitude and phase responses of different polarization channels, PolSAR captures detailed scattering information that enables accurate characterization of surface and structural properties. The spatial resolution typically ranges from several meters to tens of meters depending on the radar system. This capability is essential for both scientific analysis and practical applications. However, the presence of speckle noise severely degrades image quality and hinders the accurate interpretation of scattering mechanisms. This multiplicative noise arises from coherent interference among scatterers and exhibits spatially varying characteristics, which complicates despeckling. Conventional filtering methods often struggle to remove speckle without blurring fine structural details or altering polarimetric properties. To address these limitations, deep learning–based approaches have recently been introduced for SAR image despeckling. These methods can generally be divided into supervised and self-supervised frameworks.

Supervised learning relies on clean SAR images as ground truth. Since such clean references are practically unavailable, most supervised methods construct approximate labels. Common strategies include temporal averaging, multi-looking, or mean filtering. Although these strategies can provide reasonable supervision, the resulting networks often exhibit limited performance on real data. They may struggle to generalize across different scenes or acquisition conditions, leading to suboptimal noise suppression and loss of structural details in complex environments.

In response to the difficulty of obtaining clean labels, researchers have explored self-supervised learning as an alternative. Representative works include Speckle2Void (Molini et al., 2021) and SAR2SAR (Dalsasso et al., 2021). Speckle2Void introduces a blind-spot convolutional network that predicts masked pixels from surrounding spatial context, achieving single-image training without reference data. SAR2SAR extends the Noise2Noise (Lehtinen et al., 2018) principle to the SAR domain by utilizing temporally separated noisy image pairs, which approximately meet the statistical independence requirement. Despite these advances, self-supervised methods still face a trade-off between speckle suppression and fine-detail preservation. Their performance may degrade in the presence of temporal scene variations or abrupt changes, highlighting the need for a framework that fully exploits temporal dynamics while maintaining polarimetric fidelity.

To address these challenges, we propose TGSD-Net, a temporally guided self-supervised network that constructs pseudo training pairs from consecutive noisy observations and incorporates a change detection–based prior to exploit temporal redundancy. Auxiliary polarimetric decomposition parameters and a spatiotemporal fusion mechanism further enhance structure-aware despeckling, enabling effective noise suppression while preserving fine structural and scattering features. The main contributions of this work are threefold: (1) a self-supervised PolSAR despeckling network that leverages polarimetric decomposition and multi-temporal redundancy to enhance denoising performance; (2) a change-detection prior to improve the network’s robustness against dynamically changing regions; and (3) a multi-task joint loss function to collaboratively optimize the despeckling and change-detection networks, ensuring temporal consistency and structural fidelity. These elements collectively ensure that the network can handle heterogeneous land-cover types with high fidelity.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-477-2026 | © Author(s) 2026. CC BY 4.0 License.

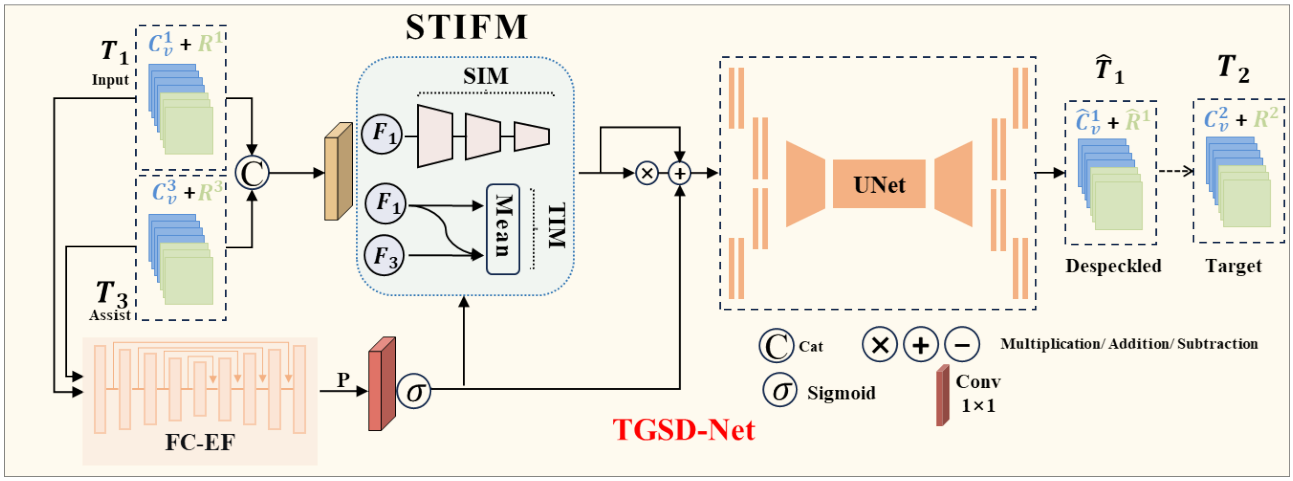

Figure 1. The framework of the proposed method.

The remainder of the paper is organized as follows. Section II details the TGSD-Net architecture, Section III presents experimental evaluations and comparisons, and Section IV concludes with discussions and potential future directions.

2. Method

2.1 Data Organization

For each pixel in the PolSAR image, the polarimetric information is represented by its covariance matrix C, derived from the Pauli scattering vector. In the case of dual-polarization Sentinel-1 data (VV and VH channels), the upper-triangular elements of C are rearranged into a real-valued feature vector:

$$C_v = [C_{11}, \Re(C_{12}), \Im(C_{12}), C_{22}]^T \in \mathbb{R}^{4\times 1}, \quad (1)$$

where C_{11} and C_{22} denote the power of the VV and VH channels, respectively, and C_{12} represents their complex correlation. This vectorized form allows the network to process the covariance representation in a real-valued domain while preserving the polarimetric structure.

To further enhance scattering representation, three additional Raney polarimetric decomposition parameters are computed for each pixel: surface scattering (Odd), double-bounce scattering (Dbl), and volume scattering (Rnd). These parameters provide complementary physical insights into the scattering mechanism, helping the network distinguish structural variations from speckle noise and improving the interpretability of despeckled results.

Inspired by the ”polarization–polarization training strategy” for self-supervised PolSAR learning (Li et al., 2024), we adopt the vectorized covariance components C_v together with three Raney decomposition parameters (Odd, Dbl, Rnd) as inputs to the network, guiding self-supervised despeckling training while fully leveraging polarimetric information to enhance scattering feature representation.

2.2 Framework of the Despeckling Network

The proposed Temporal Variation-Guided Self-Supervised Despeckling Network (TGSD-Net) aims to effectively suppress speckle noise in multi-temporal PolSAR images while maintaining structural integrity and polarimetric consistency. To exploit redundancy among consecutive acquisitions, TGSD-Net constructs pseudo training pairs from adjacent temporal observations and integrates a change-detection prior to emphasize temporally stable regions. As illustrated in Fig. 1, the entire framework consists of a primary despeckling network and an auxiliary change detection branch (Daudt et al., 2018), which are jointly optimized through a unified loss function during training.

2.2.1 Primary Despeckling Network (TGSD-Net)

The primary network adopts a U-Net backbone to effectively capture both spatial and scattering characteristics of PolSAR data. To further exploit spatiotemporal redundancy, a Spatiotemporal Information Fusion Module (STIFM) is integrated to fuse multi-scale spatial features and temporal correlations between consecutive acquisitions, enabling stability estimation and feature enhancement.

The network input comprises three consecutive acquisitions T_1, T_2, and T_3, where T_1 serves as the noisy input, T_2 acts as the pseudo reference for self-supervised learning, and T_3 provides additional spatiotemporal context. Each temporal acquisition contains four covariance matrix components $[C_{11}, \Re(C_{12}), \Im(C_{12}), C_{22}]$ and three Raney decomposition parameters (Odd, Dbl, and Rnd).

During training, the SAR data from T_1 and T_3 are first processed by the STIFM to extract and fuse spatiotemporal features, capturing both spatial coherence and temporal dependency. To mitigate feature misalignment caused by temporal variations, the prior generated by the auxiliary change detection network is incorporated to constrain the cross-attention mechanism within STIFM and to recalibrate the fused features, thereby enhancing temporal consistency and robustness. The recalibrated features are subsequently fed into the U-Net backbone for multi-scale feature extraction and despeckling reconstruction. The network is trained by minimizing the mean squared error between the despeckled output and the pseudo reference T_2, enabling TGSD-Net to suppress speckle noise while preserving polarimetric and structural consistency.

Finally, the despeckled output $\hat{T}_1$ includes four covariance components and three corresponding Raney decomposition

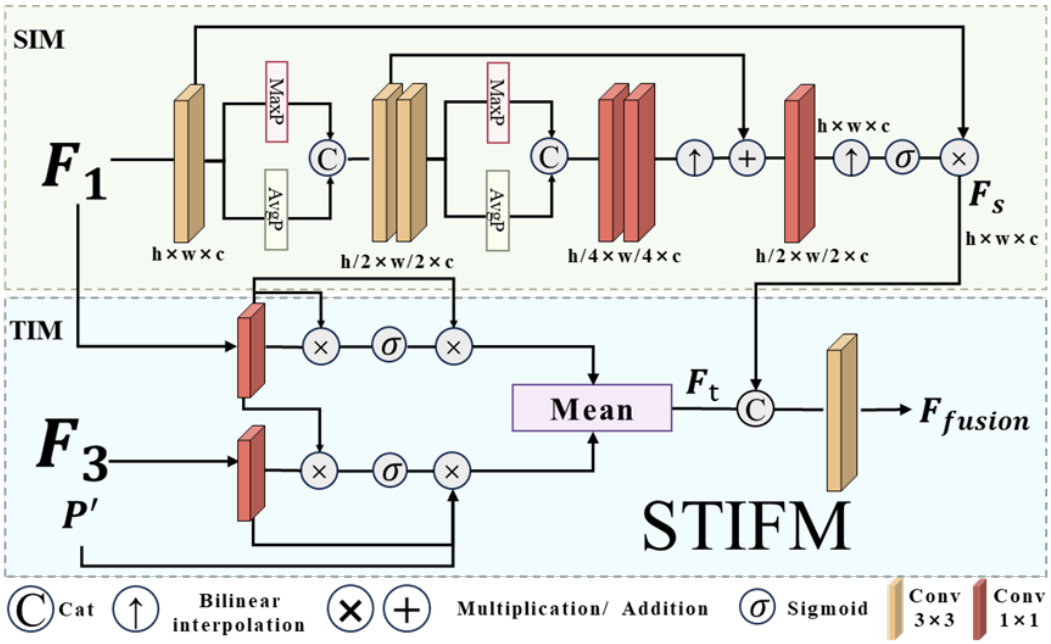

Figure 2. The structure of Spatiotemporal Information Fusion Module (STIFM).

parameters, maintaining the physical scattering interpretability of PolSAR data.

2.2.2 Auxiliary Change Detection Network (FC-EF)

To guide the denoising process, an auxiliary change detection network (FC-EF) (Daudt et al., 2018) is introduced to provide temporal stability priors. FC-EF adopts a lightweight ResNet-based encoder–decoder structure (Fig. 1(b)) and is pretrained using the outer temporal pair (T_1, T_3) to detect temporally stable regions. Feature maps extracted by ResNet are transformed into compact tokens via a Semantic Transformation (ST) module. Tokens from the two time points are concatenated along the channel dimension and processed by an encoder to enhance global representation. The fused tokens are then split, merged with their corresponding feature maps, and reconstructed by a decoder. The difference between the reconstructed feature maps yields a change detection prior, denoted as P_1, which highlights regions with minimal temporal variation.

During TGSD-Net training, the prior P_1 is converted into spatial attention weights through a 1×1 convolution followed by a Sigmoid activation. These weights recalibrate the fused spatiotemporal feature maps from STIFM, emphasizing stable regions while suppressing those affected by significant temporal changes. By employing FC-EF solely as an auxiliary guidance module rather than a direct output branch, TGSD-Net effectively balances speckle suppression with preservation of fine structural and polarimetric details. This cooperative architecture enhances denoising robustness under heterogeneous land-cover conditions and maintains consistent scattering behavior across multi-temporal acquisitions.

2.3 Spatiotemporal Information Fusion Module (STIFM)

To exploit spatiotemporal redundancy, TGSD-Net incorporates a Spatiotemporal Information Fusion Module (STIFM) at the denoising stage, which effectively integrates temporal correlations and multi-scale spatial features while leveraging the change-detection prior to reduce interference from temporal variations. STIFM consists of two complementary submodules: the Temporal Information Module (TIM) and the Spatial Information Module (SIM), as illustrated in Fig. 2.

2.3.1 Temporal Information Module (TIM)

The Temporal Information Module (TIM) aims to capture temporal correlations between consecutive feature maps. Its goal is to highlight temporally stable structures and suppress speckle noise and scene dynamics. TIM receives feature maps F_1 and F_3 as input, corresponding to time points T_1 and T_3. The output is the temporally fused feature F_t.

In the Temporal Information Module (TIM), the input feature maps are processed separately using self-attention and cross-attention mechanisms. The self-attention focuses on enhancing the internal structural consistency of T_1, while the cross-attention captures temporal correlations between T_1 and T_3, highlighting temporally stable structures and suppressing inconsistencies caused by speckle noise or scene dynamics. To improve robustness against dynamically changing regions, the change-detection prior P generated by the auxiliary change-detection network (FC-EF) is introduced as a soft spatial mask in the cross-attention computation. This guides the network to down-weight unreliable areas while preserving stable regions during feature fusion. Finally, TIM fuses the features from different time points by computing their temporal mean, producing a temporally consistent and structurally stable feature representation.

The computation of attention coefficients and temporal fusion can be expressed as:

$$\begin{aligned} A_1 &= \sigma(\mathrm{Conv}_1(F_1) \otimes \mathrm{Conv}_1(F_1)), \\ A_3 &= \sigma(\mathrm{Conv}_1(F_1) \otimes \mathrm{Conv}_3(F_3)) \end{aligned} \tag{2}$$

where Conv_1 and Conv_3 are convolutional blocks for feature

extraction from T_1 and T_3, respectively. The operator $\otimes$ denotes element-wise multiplication. The Sigmoid function $\sigma(\cdot)$ normalizes the attention coefficients to the range $[0, 1]$. A_1 represents self-attention of T_1, emphasizing structural consistency within the same time frame. A_3 represents cross-attention between T_1 and T_3, capturing temporal correlations.

The temporally fused feature incorporating the change-detection prior is calculated as:

$$F_t = \text{mean}(A_1 \otimes \text{Conv}_1(F_1),\ A_3 \otimes \text{Conv}_3(F_3) \otimes P') \quad (3)$$

where P' is the spatially aligned and smoothed version of the prior P generated by FC-EF. As a soft mask, $P' \in [0, 1]^{H \times W}$ indicates temporal stability for each pixel: higher values correspond to unchanged regions, and lower values correspond to changed or unreliable regions. Applying P' only in the cross-attention branch suppresses features from dynamic areas in T_3 while retaining stable regions for reliable temporal fusion. Finally, the recalibrated features from T_1 and T_3 are averaged to obtain the temporally fused feature F_t, emphasizing consistent structures and mitigating the influence of speckle and scene dynamics.

2.3.2 Spatial Information Module (SIM)

The Spatial Information Module (SIM) focuses on extracting multi-scale spatial features from the feature map of a single time point T_1. SIM enhances the network's sensitivity to structural details while preserving contextual coherence. The processing of SIM can be divided into three stages: feature pyramid construction, multi-scale feature fusion, and spatial attention modulation.

First, the input feature map F_i is organized into a three-level pyramid. Each level applies convolution, average pooling, and max pooling to capture spatial information at different scales. Low-resolution layers provide global context, while high-resolution layers retain fine-grained details. The combination of convolution, pooling, and upsampling operations across layers allows the network to progressively integrate low-resolution contextual information with high-resolution details, producing rich spatial representations.

After multi-scale fusion, channel-wise attention is applied to the fused features to generate the spatially attentive feature map F_s. This map is then multiplied element-wise with the original features to enhance and integrate spatial information:

$$F_s = \sigma\Big(\text{Conv}_S\big(\text{Concat}(\text{AvgPool}(F_1), \text{MaxPool}(F_1))\big)\Big) \otimes F_1, \quad (4)$$

where AvgPool and MaxPool denote average pooling and max pooling, respectively. Concat$(\cdot)$ represents channel-wise concatenation. This process enables the network to adaptively emphasize important channels and spatial regions, improving its ability to represent fine-grained details.

Finally, the spatially re-weighted feature map F_s is concatenated with the temporally fused feature F_t along the channel dimension and linearly transformed to obtain the final spatiotemporal representation:

$$F_{\text{fusion}} = \text{Conv}_{\text{fusion}}(\text{Concat}(F_t, F_s)), \quad (5)$$

By combining temporal consistency from TIM with multi-scale spatial refinement from SIM, the STIFM produces a fused feature map F_{fusion} that is both temporally stable and spatially rich. This comprehensive representation provides a robust feature foundation for structure-aware and physically consistent PolSAR despeckling in complex heterogeneous environments.

2.4 Loss Function

To guide the despeckling network while incorporating temporal consistency and leveraging change detection priors, the overall loss function is formulated as a weighted combination of the primary despeckling loss and the auxiliary change detection loss:

$$\text{loss} = w_1 \times \mathcal{L}_{\text{despeckle}} + w_2 \times \mathcal{L}_{\text{CD}}, \quad (6)$$

where $\mathcal{L}_{\text{despeckle}}$ represents the main self-supervised PolSAR despeckling loss, and $\mathcal{L}_{\text{CD}}$ denotes the loss associated with the change detection network. The weights w_1 and w_2 control the relative contributions of these terms. In practice, w_1 is assigned a larger value than w_2 to ensure that despeckling remains the dominant objective, while the change detection term plays a supporting role to enforce temporal and structural consistency.

The despeckling loss is computed by comparing the network output $f_\theta(T_1)$ with the pseudo-target T_3, but only over regions with minimal temporal changes, as indicated by the mask derived from the change detection prior:

$$\mathcal{L}_{\text{despeckle}} = \sum_{i=1}^{M} \left\| M_{P_1} \odot f_\theta(T_1) - M_{P_1} \odot T_3 \right\|_2, \quad (7)$$

where f_θ denotes the despeckling network with parameters θ, $\| \cdot \|_2$ represents the L2 norm, $M_{P_1} = 1 - P_1$ is the complementary mask from the change detection output, and $\odot$ denotes element-wise multiplication. By focusing the loss computation on temporally stable regions, the network is encouraged to preserve scattering structures and fine details, while ignoring areas that may have changed between time points.

The change detection network is trained using the standard cross-entropy loss:

$$\mathcal{L}_{\text{CD}} = \frac{1}{H \times W} \sum_{h=1}^{H} \sum_{w=1}^{W} l(P_{hw}, Y_{hw}), \quad (8)$$

where $l(P_{hw}, Y_{hw})$ is the cross-entropy between the predicted probability P_{hw} and the corresponding label Y_{hw} for pixel (h, w), and $H \times W$ is the spatial size of the input image. The output of this network is not directly used for despeckling; instead, it provides the prior mask M_{P_1} that guides the despeckling network to focus on regions with minimal temporal variation.

Overall, this composite loss function ensures that TGSD-Net achieves a balanced trade-off: w_1 enforces accurate noise suppression while maintaining polarimetric structures, and w_2 guarantees temporal consistency by leveraging the change detection prior. This design allows the network to effectively handle multi-temporal SAR imagery, suppress speckle noise, and preserve both structural and scattering fidelity across varying land-cover types.

Table 1. Quantitative comparison of despeckling performance on different land covers.

Land cover	Urban		Agricultural		Mountainous	
	ENL	EPD	ENL	EPD	ENL	EPD
Noise	5.4458	–	15.0032	–	6.3023	–
SAR2SAR	11.8342	0.9827	48.1123	0.9851	14.2034	0.9789
SAR-DRN	10.2568	0.9758	41.5090	0.9774	12.0045	0.9861
deSpeckNet	14.6721	0.9883	69.3345	0.9920	17.1178	0.9817
Ours	**19.2046**	**0.9869**	**80.1120**	**0.9973**	**20.8452**	**0.9894**

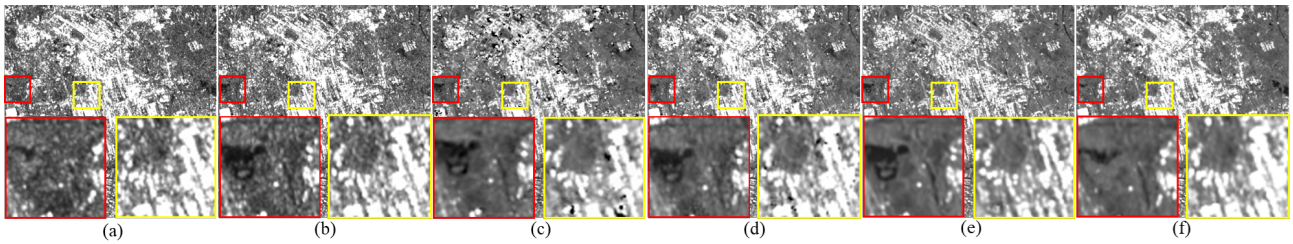

Figure 3. The visualization results. (a) Noisy image (T_1). (b) Noisy image (T_3). (c) SAR2SAR. (d) SAR-DRN. (e) despeckNet. (f) TGSD-Net.

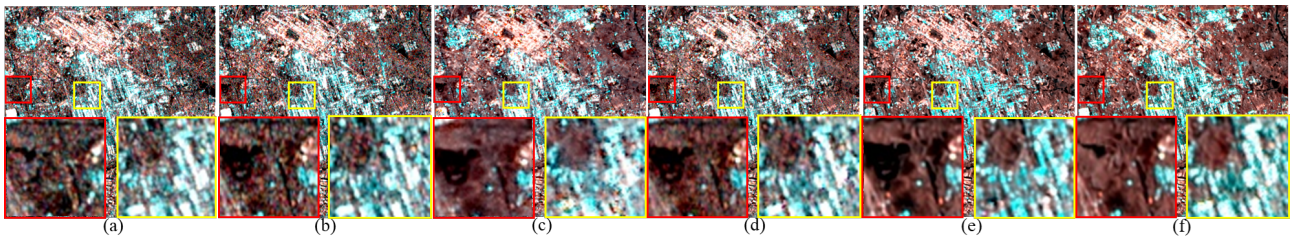

Figure 4. False-color Images. (a) Noisy image (T_1). (b) Noisy image (T_3). (c) SAR2SAR. (d) SAR-DRN. (e) despeckNet. (f) TGSD-Net.

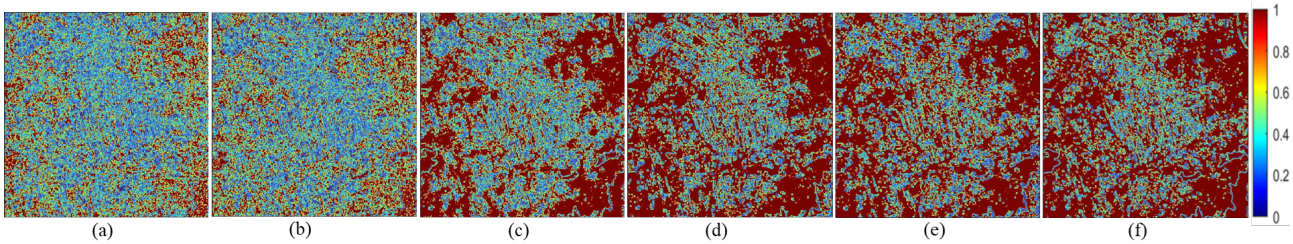

Figure 5. ENL color Images. (a) Noisy image (T_1). (b) Noisy image (T_3). (c) SAR2SAR. (d) SAR-DRN. (e) despeckNet. (f) TGSD-Net.

3. Experiment and analysis

This study uses C-band dual-polarization (VV and VH) Sentinel-1 SAR data covering ten consecutive time points over a mixed urban, vegetation, and water area. The data were acquired in IW mode with approximately 10 m ground resolution. Standard pre-processing, including orbit correction, radiometric calibration, and multi-looking, was applied to reduce speckle and ensure geometric and radiometric consistency. Covariance matrix elements and Raney polarization decomposition parameters (Raney et al., 2012) were generated using PolSARpro (v6.0), along with change detection labels to guide self-supervised training by identifying temporally stable regions. For network training, the first eight temporal images were used, forming 35,000 training and 5,000 validation patches, while the last two time points were reserved for testing. The network was trained using the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, an initial learning rate of 0.001 halved every 25 epochs, for a total of 100 epochs.

For quantitative evaluation, the proposed TGSD-Net is compared with three representative self-supervised or deep-learning-based despeckling networks: SAR2SAR (Dalsasso et al., 2021), SAR-DRN (Zhang et al., 2018), and DeSpeckNet (Mullissa et al., 2020). Two objective metrics are considered: the Equivalent Number of Looks (ENL), which measures the degree of speckle suppression in homogeneous regions, and the Edge Preservation Degree (EPD), which quantifies the ability to retain structural details. Higher ENL values indicate more effective noise reduction, whereas higher EPD values reflect stronger edge preservation capability.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-477-2026 | © Author(s) 2026. CC BY 4.0 License.

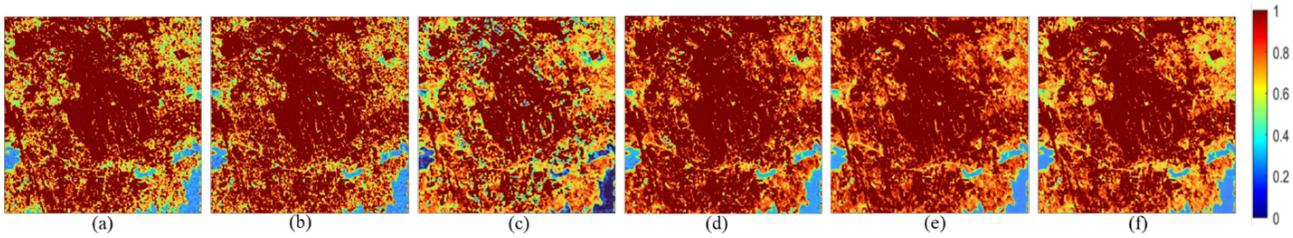

Figure 6. Polarization Decomposition Images (ODD). (a) Noisy image (T_1). (b) Noisy image (T_3). (c) SAR2SAR. (d) SAR-DRN. (e) despeckNet. (f) TGSD-Net.

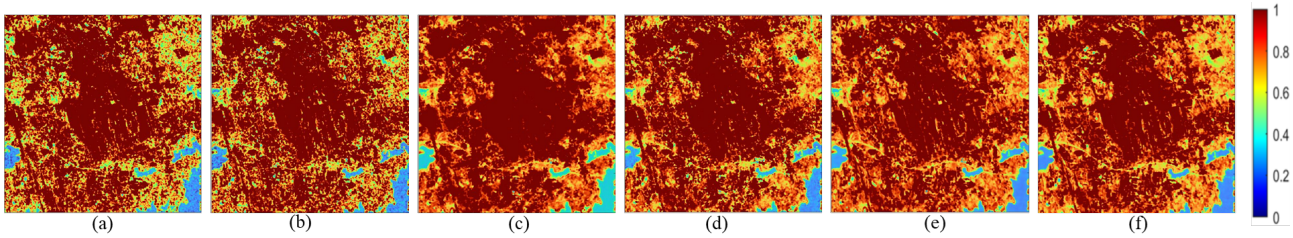

Figure 7. Polarization Decomposition Images (DBL) (a) Noisy image (T_1). (b) Noisy image (T_3). (c) SAR2SAR. (d) SAR-DRN. (e) despeckNet. (f) TGSD-Net.

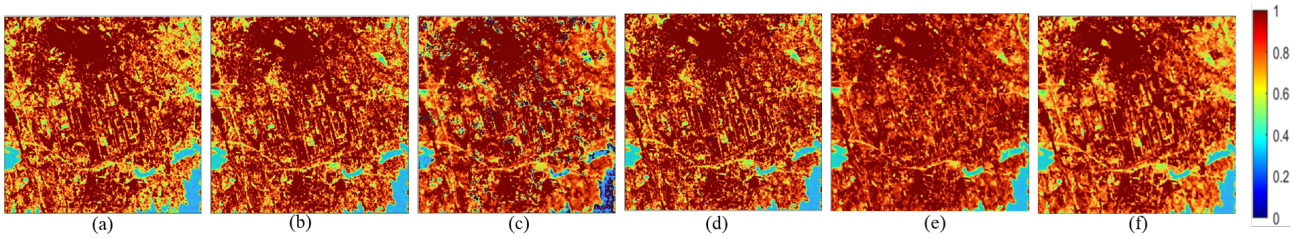

Figure 8. Polarization Decomposition Images (RND) (a) Noisy image (T_1). (b) Noisy image (T_3). (c) SAR2SAR. (d) SAR-DRN. (e) despeckNet. (f) TGSD-Net.

3.1 Real experiment

From Table 1, it can be observed that TGSD-Net significantly surpasses the competing methods in both ENL and EPD. Specifically, TGSD-Net achieves an ENL improvement of over 38.9% compared with DeSpeckNet and more than 70% compared with SAR2SAR, indicating a substantial enhancement in homogeneous region smoothing. The EPD metric, which remains close to 1.0, further confirms that edge structures are faithfully preserved. These consistent improvements across both indicators demonstrate that TGSD-Net achieves a superior balance between speckle suppression and edge detail retention.

Figure 3 presents the qualitative results of the compared methods. The proposed TGSD-Net produces smoother intensity distributions with fewer residual speckles, particularly in homogeneous and low-texture regions. Meanwhile, fine structural boundaries are preserved, avoiding the over-smoothing effect visible in SAR-DRN and DeSpeckNet. Figure 4 provides pseudo-color visualizations, where different regions are more clearly distinguishable in the TGSD-Net results. Figure 5 further visualizes ENL maps, where TGSD-Net exhibits both higher ENL values and a more uniform spatial distribution, confirming that the network maintains statistical consistency across spatially varying terrains.

During training, T_1 was used as the noisy input and T_2 as the reference target. Because methods without a temporal-change prior tend to adapt excessively to the later image, their outputs exhibit a bias toward the reference temporal frame T_2. In contrast, TGSD-Net, guided by the change detection prior, aligns more closely with the true structures of T_1, effectively mitigating the temporal mismatch effect. This observation demonstrates that the designed temporal-guided self-supervision can significantly enhance the robustness of the network when dealing with long-term multi-temporal SAR data.

3.2 Polarization decomposition experiment

The preservation of polarimetric information is critical for subsequent PolSAR applications such as land-cover classification and scattering mechanism analysis. Therefore, beyond despeckling intensity values, it is essential that the despeckled data retain physical consistency across polarization channels. Figures 6, 7 and 8 illustrate the three Raney decomposition components, which represent Odd (surface scattering), Dbl (double-bounce scattering), and Rnd (volume scattering), derived from the despeckled results.

The comparison reveals that SAR2SAR tends to underestimate Odd scattering, leading to an overly smooth appearance in local high-scattering regions. Conversely, SAR-DRN and DeSpeckNet occasionally overestimate the Odd and Rnd components, resulting in unrealistic regions with low or medium scattering intensity in non-vegetated regions. TGSD-Net, by contrast, maintains strong consistency with the original noisy input in both spatial distribution and intensity of all three components. The stable relative ratios among Odd, Dbl, and Rnd

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-477-2026 | © Author(s) 2026. CC BY 4.0 License.

further indicate that TGSD-Net preserves the underlying polarimetric scattering mechanisms without introducing systematic bias. These results confirm that the proposed network not only excels in noise reduction but also maintains polarimetric integrity—a property crucial for reliable downstream analysis.

4. Conclusion

In this work, we proposed a PolSAR despeckling framework that explicitly incorporates temporal variation and polarization decomposition information. The proposed TGSD-Net introduces a spatiotemporal feature fusion module that aggregates complementary information from multiple temporal acquisitions, enhancing the representational capacity of the network. A change detection–guided loss function constrains the learning process, allowing the model to adaptively distinguish between genuine temporal changes and speckle-induced fluctuations. Extensive experiments on multi-temporal Sentinel-1 datasets demonstrate that TGSD-Net achieves superior quantitative (ENL and EPD) and qualitative performance compared with existing deep learning–based methods.

The results verify that leveraging temporal priors and polarization features jointly can substantially improve PolSAR despeckling. Future research will extend this concept toward unified frameworks for multi-task remote sensing image enhancement, such as simultaneous despeckling, super-resolution, and semantic segmentation, with the ultimate goal of providing general-purpose models for PolSAR data quality improvement and application readiness.

References

Dalsasso, E., Denis, L., Tupin, F., 2021. SAR2SAR: A Semi-Supervised Despeckling Algorithm for SAR Images. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 14, 4321–4329.

Daudt, R. C., Le Saux, B., Boulch, A., 2018. Fully convolutional siamese networks for change detection. *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 4063–4067.

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T., 2018. Noise2Noise: Learning Image Restoration Without Clean Data. *arXiv preprint arXiv:1803.04189*.

Li, J., Lin, L., He, M. et al., 2024. Sentinel-1 Dual-Polarization SAR Images Despeckling Network Based on Unsupervised Learning. *IEEE Trans. Geosci. Remote Sens.*, 62, 1–15.

Molini, A. B., Valsesia, D., Fracastoro, G., Magli, E., 2021. Speckle2Void: Deep Self-Supervised SAR Despeckling with Blind-Spot Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–17.

Mullissa, A. G., Marcos, D., Tuia, D., Herold, M., Reiche, J., 2020. DeSpeckNet: Generalizing Deep Learning-Based SAR Image Despeckling. *IEEE Trans. Geosci. Remote Sens.*, 60, 1–15.

Raney, R. K., Cahill, J. T. S., Patterson, G. W., Bussey, D. B. J., 2012. The M-Chi Decomposition of Hybrid Dual-Polarimetric Radar Data with Application to Lunar Craters. *J. Geophys. Res. Planets*, 117(E12).

Zhang, Q., Yuan, Q., Li, J., Yang, Z., Ma, X., 2018. Learning a Dilated Residual Network for SAR Image Despeckling. *Remote Sens.*, 10(2), 196.