






GANInSAR: Deep Generative Modeling for Large-Scale InSAR Signal Simulation

Zhongrun Zhou , Graduate Student Member, IEEE, Xinyao Sun , Fei Yang , Zheng Wang , Ryan Goldsbury , and Irene Cheng , Senior Member, IEEE

Abstract—Interferometric synthetic aperture radar (InSAR) technology is widely used to create digital elevation models and measure dynamics on the Earth’s surface, including monitoring ground displacements. The lack of or limited-collected ground-truth data, however, often poses a bottleneck in validating the research outcome, particularly at high precision and resolution levels. To mitigate the gap, we introduce a new deep generative model (DGM) for the simulation of linear deformation rate maps. We demonstrate that our adversarial DGM architecture with carefully designed preprocessing and postprocessing modules performs well for InSAR deformation signal synthesis, even when limited data are available. We also introduce a dimensionality reduction method, based on the distance between the real-world and generated image feature vectors, to address the lack of quantitative evaluation for data simulation. Furthermore, we introduce a hybrid evaluation metric integrating quantitative and qualitative measures, which is more intuitive than the existing methods and makes it easier for domain experts to participate in the evaluation. We compare the results of our model with established methods. The comparison result illustrates the superior performance of our proposed method.

Index Terms—Deep generative modeling, generative adversarial network (GAN), interferometric synthetic aperture radar (InSAR), radar signal processing, remote sensing images, simulation.

I. INTRODUCTION

INTERFEROMETRIC synthetic aperture radar (InSAR) technology has been widely used to understand the Earth’s surface and subsurface movements [1], [2]. In certain applications, InSAR delivers much better results compared to traditional photogrammetry, level sensors, and light detection and ranging technologies [3], [4]. The major advantage of InSAR is that it uses an active microwave remote sensing detection mode, which lets it work in 24 h lighting conditions and in any weather (even with clouds or smoke). InSAR can extract line-of-sight shifts with accuracy from centimeters to millimeters by utilizing the phase difference between two time-and/or-space-separated SAR images of the same area [4], [5]. Due to the highly precise and

dense measurements provided by InSAR, gathering adequate ground-truth data for validating InSAR measurements becomes challenging. The scarcity or limited availability of collected ground-truth data poses a significant hurdle in validating InSAR pipelines and the related research findings. In addition, stakeholders are consistently intrigued by the minimum detectable size of deformation and displacement, but quantifying these factors in practical scenarios presents inherent difficulties. In order to generate surface motion information with high spatial resolution and large coverage, there is a pressing need for a reliable and efficient image synthesis method. This study aims to mitigate this research gap by utilizing deep generative models (DGMs) to simulate linear deformation rate maps, which are a common product format derived from InSAR measurements.

A. Existing Synthetic InSAR Data

To the best of our knowledge, there are two synthetic InSAR data methods: unconditional and conditional data generation. For unconditional data generation, no previous generative adversarial network (GAN) model has developed an InSAR signal simulator for generating realistic complex features. Related works of unconditional synthesis method only simulate basic geometric features, such as cone, peak, slope, and square [6], [7], [8], [9], [10]. They use various linear mathematical functions for drawing basic geometric features [3], [11]. Conditional synthetic method generate InSAR data [12], [13], [14], [15], [16], [17] based on digital elevation model (DEM)/InSAR interferograms. All of these methods are also based on mathematical models to randomly generate basic geometric features like the previous methods, but the difference is then used to add randomly crop DEM/InSAR interferograms data, which is combined and used as training data.

1) *Unconditional Synthesis of InSAR Data*: In terms of unconditional synthetic of InSAR data, Duan et al. [3] presented a simulated annealing algorithm to achieve deformation results with high accuracy and high efficiency. The main contribution is that they designed different functions of acceptance criterion and random solution generator. Furthermore, the random solution generator introduced an exponential function as the mapping function to transform the random number to speed up the convergence of the final generation. As a result, the work alleviated the drawbacks of traditional deformation estimation methods, such as local extremums and computational efficiency. In the same year, Anantrasirichai et al. [11], [18] proposed a

Manuscript received 14 July 2023; revised 25 October 2023 and 1 January 2024; accepted 16 January 2024. Date of publication 2 February 2024; date of current version 29 February 2024. This work was supported by the NSERC under Grant CRDPJ 543428 - 19. (Corresponding author: Zhongrun Zhou.)

Zhongrun Zhou, Xinyao Sun, Fei Yang, and Irene Cheng are with the Computing Science, University of Alberta, Edmonton, AB T6G 2R3, Canada (e-mail: zhongrun@ualberta.ca; xinyao1@ualberta.ca; fei5@ualberta.ca; locheng@ualberta.ca).

Zheng Wang and Ryan Goldsbury are with RnD, 3V Geomatics Inc., Vancouver, BC V5Y 0M6, Canada (e-mail: zwang@3vgeomatics.com; rgoldsbury@3vgeomatics.com).

Digital Object Identifier 10.1109/JSTARS.2024.3361444

synthetic training dataset for detecting volcanic deformation in radar interferograms and InSAR time series based on the U.K. velocity map characteristics. Since they lacked real deformation signals, they trained with synthetic signals from point sources and tunnels. To create the synthetic datasets, they used a simple linear function with two components: the deformation and the turbulent atmosphere. The authors utilized two deformation models commonly associated with coal mining and tunneling in the U.K. With the help of the synthetic training dataset, Anantrasirichai et al. [18] adapted a pretrained CNN model to classify sustained volcanic deformation in InSAR. The same team [11] also adapted a pretrained CNN model to detect ground deformation caused by mining, landslides, and dewatering. By generating large volumes of training data using basic geometric features, the authors could detect patterns related to geological hazards. Sun et al. [9] introduced DeepInSAR, a deep convolutional neural network-based model capable of intelligently resolving phase filtering and coherence estimation challenges. Through evaluations on simulated and real data, the teacher–student framework addresses the issue of missing clean InSAR ground truth, demonstrating superior performance in both quantitative and qualitative assessments.

Sun et al. [10] introduced a two-stage black-box optimization framework designed to jointly estimate average ground deformation rates and terrain DEM errors. The method leverages an iterative grid search to obtain coarse candidate solutions, followed by the application of a covariance matrix adaptive evolution strategy to refine the final local results. Performance assessments using both simulated and real datasets, including quantitative and qualitative comparisons with different optimizers, affirm the reliability and effectiveness of their approach. These two papers provide a simulator that generates InSAR signals, as illustrated in the left columns of Figs. 1 and 2. However, these related works only use basic geometric features to simulate basic isolated features. They do not simulate realistic characteristics of composite features found in InSAR images.

2) *Conditional Synthetic of InSAR Data*: In terms of DEM/InSAR interferograms-based conditional synthetic of InSAR data, Rouet-Leduc et al. [14] demonstrated the use of synthetic data to train an autoencoder model as part of a generative model for detecting ground deformation signals from noisy InSAR time series. The researchers added atmospheric and topographic noise to interferograms to create synthetic time series of deformation, which were then used to train two independent models. One model was trained to recover point source deformation, while the other was trained to recover deformation on faults. Both models were designed to map synthetic noisy time series to synthetic cumulative displacements [14]. The use of synthetic data in this way enables the autoencoder model to autonomously untangle ground deformation from noise in InSAR time series and detect deformation signals without prior knowledge of the location or slip behavior of a fault. By successfully training models to detect deformation signals in the presence of noise, this method has the potential to improve the accuracy of ground deformation monitoring using InSAR technology. The following year, Brengman et al. [15] also adapted a CNN to detect, locate, and classify surface

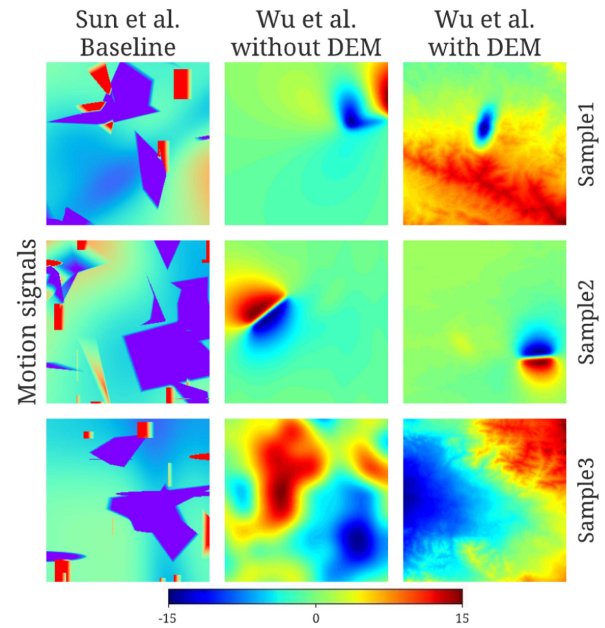


Fig. 1. From left to right, we show samples of the simulated motion signals (Sun et al. [8], [9] as baseline, and Wu et al. [16], [17] without DEM and with DEM as compared work). All samples have a value range of $[-15, +15]$ cm/year.

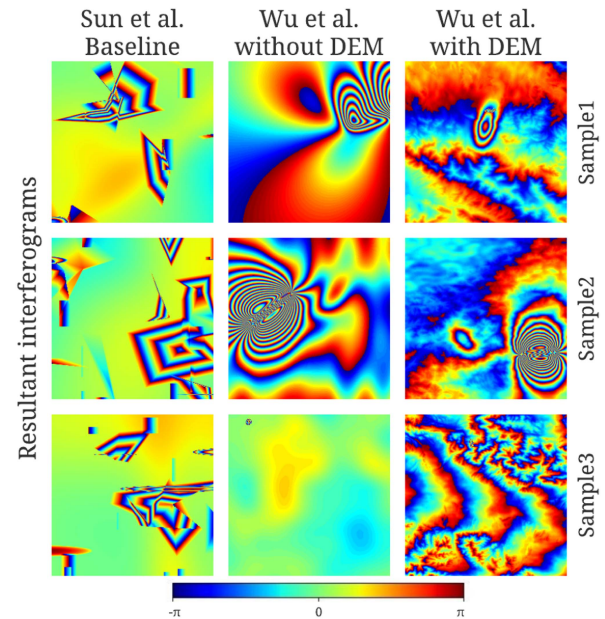


Fig. 2. From left to right, we show samples of the simulated resultant interferograms (Sun et al. [8], [9] as baseline, Wu et al. [16], [17] without DEM and with DEM as compared work). All samples are the interferometric of phase wrapping $[-\pi, +\pi)$ – (Blue: $-\pi$; Red: $+\pi$).

deformation by using synthetic data. Specifically, they generated synthetic interferograms containing wrapped and unwrapped surface deformations with synthetic atmospheric and topographic noise. To generate the noise, they used a noise correlation scale to create atmospheric noise. Then, they randomly added noise related to the landscape by scaling DEMs and mixing in the noise. These steps created a final noisy image with a millimeter to centimeter scale noise amplitude that was spatially and topographically linked. As a result,

generated atmospheric and topographic noise on interferograms as synthetic data can make the CNN network capable of detecting, locating, and classifying surface deformation with high accuracy and high efficiency.

Wu et al. [16], [17] used deep convolutional neural networks to detect and map rapid subsidence induced by mining activities using time-series Sentinel-1 SAR images. They developed a deformation detection network to automatically identify swiftly subsiding areas from wrapped interferograms and a phase unwrapping network designed to unwrap interferogram patches centred on detected subsidence locations [16], as illustrated in the middle columns of Figs. 1 and 2. They provided interferogram simulation strategies that generate diverse training samples incorporating distorted 2-D Gaussian surfaces [19] and fractal Perlin noise [20]. At the same year, Wu et al. [17] presented an innovative approach to enhance the reliability of phase unwrapping in interferograms. This article introduces the discontinuity estimation network tailored to predict phase discontinuities with remarkable precision. It utilizes a multichannel input approach, encompassing interferogram data, range/azimuthal phase gradients, and residue maps, providing robust guidance for discontinuity prediction, as illustrated in the right columns of Figs. 1 and 2. However, these methods are based on mathematical models that randomly generate various shapes and combine them with randomly cut DEM/InSAR interferograms data as training data.

B. Deep Generative Models

DGMs refer to neural networks with many hidden layers that are used to approximate complex high-dimensional probability distributions [21]. A deep neural network is used as a generator in DGMs to transform a simple distribution, like a univariate Gaussian, into a more complicated distribution [21]. However, the performance of DGMs heavily depends on four hyperparameters, including the customized network model, the training datasets, the regularization, and the training algorithms [21]. Three popular DGM approaches are normalizing flows (NFs) [22], [23], variational autoencoders (VAEs) [24], [25], [26], and GAN [27], [28], [29]. We choose GAN because it has demonstrated superior performance in generating realistic images compared to NF and VAE [30]. However, selecting the appropriate hyperparameters is crucial because the performance of DGMs can be very different depending on the application and data distribution.

In 2014, Goodfellow et al. [27] presented GAN as an alternative way of producing visual similar data based on the characteristics of the original data. This original GAN model has a generator and a discriminator with fully connected neural layers [27]. Its uncomplicated structure has limitations and can only be applied to some simple datasets, such as MNIST and CIFAR-10. The original GAN algorithm was designed to approximate data distributions, but a simple architecture may not be sufficient to generate high-quality images [31]. Two years later, Radford et al. [29] proposed a deep convolutional generative adversarial network (DCGAN), which is the first model that uses the deep convolutional structure for the generator. This model provides some guidelines for how convolutional layers can be used with GANs. The generator of DCGAN has the

spatial upsampling ability of the deconvolution, which allows it to generate higher resolution images. Furthermore, DCGAN can synthesize more natural images using complex random noise generation and a unique loss function [31]. Progressive GAN (PROGAN) was presented in 2018 by Karras et al. [32]. The core idea of PROGAN is to grow the generator and discriminator gradually, which is a progressive neural network and leads to faster and more stable training producing higher quality results. Therefore, this network is popular for learning complex task sequences. In the next year, one novel method for high-resolution image synthesis is StyleGAN [33], which Nvidia researchers had shown to work reliably on various datasets. The same researchers exposed and analyzed several of its characteristic artifacts. To address these issues, they proposed changes in both model architecture and training methods. The results of evaluating the StyleGAN model on the FFHQ dataset and the CELEBA-HQ dataset show that StyleGAN is better than traditional GAN generator architectures in every way.

Traditional CNN can only capture local spatial information. It is difficult for CNN-based GANs to learn multiclass image datasets, and some key parts in the generated images may be displaced. Therefore, in 2019, Zhang et al. [34] employed a self-attention mechanism in the generator and discriminator of traditional GAN to design self-attention GANs (SAGANs). This mechanism enables SAGAN to learn global and long-range dependencies to generate images and it is widely applied in large feature images to improve GANs. BigGAN [35] model was proposed by Brock et al. in 2019 using SAGAN as the baseline. This model proves that scaling up GAN training can improve the performance benefits of larger models and larger batches. BigGAN improves the state-of-the-art inception score (IS, higher is better) from 52.52 to 166.5 and Fréchet Inception Distance (FID, lower is better) from 18.65 to 7.4, respectively, on ImageNet at 128×128 resolution. The authors also evaluated the BigGAN at 256×256 and 512×512 resolutions on ImageNet: IS obtained 232.5 (256×256) and 241.5 (512×512), and FID achieved 8.1 (256×256) and 11.5 (512×512). Compared with DCGAN and SAGAN, BigGAN generates more realistic and large-scale images. However, the computational cost of BigGAN is relatively high, and therefore the model should be selected appropriately according to the application requirements. We are also aware of potential problems with IS evaluation metric. In the training of GAN networks, larger image scales pose greater difficulty and yield poorer results. The FID aligns with the general consensus, indicating that as the image scale increases, the generated results deviate further from the real-world dataset. The IS metric provide the opposite conclusion. This discrepancy arises because the IS metric does not directly measure the similarity between synthetic and real images.

Therefore, subsequent articles on GAN only use the FID criterion and rarely use the IS for evaluation, as researchers have doubts about the reliability of IS. In 2020, Schonfeld et al. [36] proposed a U-Net GAN model based on the BigGAN model. They substitute the original discriminator with an alternative U-Net-based discriminator, which can help the GAN model focus on both the global structure and local details of images following

the principle of segmentation. The new U-Net discriminator reuses the original discriminator classification architecture as the encoder and its decoder part uses the building blocks of the generator network [36]. Simulating global and local coherent images with object shapes and textures that cannot be distinguished from real-world images is one of the main challenges faced by GANs [36]. The specially designed discriminator in the proposed U-Net GAN model can effectively alleviate this problem. The authors evaluated U-Net GAN on FFHQ, CelebA, and the proposed COCO-Animals datasets. Compared with the BigGAN model, a better result was obtained with an average decrease of 2.7 FID points.

Since then, the research in this area has continued to expand and improve to address the shortcomings of earlier networks for processing natural images. The generative adversarial modeling's inherent research objective is that the generator and discriminator are trained together to maintain an equilibrium [37]. Recent advances in this area have helped to achieve this goal and have led to more effective models for generating images. Nevertheless, GAN has been successful only in image generation tasks for specific datasets and is mainly limited to photorealistic image generation. Further research is needed to expand its capabilities for addressing its limitations, and exploring a wider range of applications.

In this article, our main contributions are as follows.

- 1) We introduce deep generative modeling to generate large-scale data for InSAR signal analysis.
- 2) Our signal simulator can generate data to help evaluate different kinds of complex features, comparing to related methods, in satellite remote sensing for practical industrial applications.
- 3) We introduce a hybrid evaluation criteria (qualitative and quantitative) to evaluate the proposed GAN-InSAR approach on synthetic signals.

II. MATERIALS AND METHODS

A. Mathematical Modeling for InSAR Simulator

InSAR processing usually starts with flat-Earth phase removal (with orbit data) and topographic phase correction (with a DEM). The remained phase difference (after flat-Earth phase removal and topographic correction) is mainly affected by five factors: 1) surface deformation, 2) topographic errors (i.e., phase contributed by DEM height errors), 3) atmospheric delay, 4) orbital errors, and 5) decorrelation noise [10], [12], [38], with each representing a different phenomenon [39], [40]. For a given pixel location l in InSAR, the differential interferogram phase is described as follows:

$$\begin{aligned}\phi_l &= \phi_{\text{def},l} + \phi_{\text{topo},l} + \phi_{\text{atm},l} + \phi_{\text{orbit},l} + \phi_{\text{noise},l} \\ \hat{\phi}_l &= \arctan2(\sin(\phi_l), \cos(\phi_l))\end{aligned}\quad (1)$$

where ϕ_{def} represents ground deformation motion phase components, ϕ_{topo} means topographic errors (i.e., phase contributed by DEM height errors) [41], atmosphere phase ϕ_{atm} includes height-dependent dry atmosphere and long-wavelength wet atmosphere. The atmosphere phase is removed by an empirical

model to estimate the vertical structure of the dry phase delay and an spatial filter to remove the long-wavelength wet atmosphere signal. ϕ_{orbit} is the phase due to orbit inaccuracies, it behaves as global phase ramps and can be removed by a linear fitting. It is usually insignificant in modern satellite datasets because of precise orbit control. ϕ_{noise} denotes decorrelation noise that can be reduced by phase filtering, and the observed phase is $\hat{\phi}_l$.

Surface deformation can result from load changes at the surface or ruptures and pressure changes in the subsurface [3], [12]. Domain specialists usually consider ground deformation motion information to be valuable (a crucial signal in an interferogram) and difficult to obtain. Simultaneously, the ground deformation motion has no displacement in the majority of locations (the value is 0), which is detrimental to the generative network's training and convergence due to class imbalance.

For each pixel, the ground deformation motion (m_r) phase component can be modeled as a linear function for each surface deformation phase as follows:

$$\begin{aligned}\phi_{\text{def}}^k &= -\frac{4\pi}{\lambda}(d^{\text{first}} - d^{\text{second}})^k \cdot m_r \\ &= -\frac{4\pi}{\lambda}\Delta\text{days}^k \cdot m_r \\ &= \text{conv}_{\text{def}} \cdot \Delta\text{days}^k \cdot m_r\end{aligned}\quad (2)$$

where Δdays is the temporal difference between two acquisitions on distinct days (d^{first} , d^{second}) used to form the interferogram, and conv_{def} is unit conversion factor [10].

This article will focus on simulating ground deformation rate imagery. Each pixel value is the linear deformation rate over a certain observation period. It is estimated by a network of interferograms with all other signals (such as atmosphere phase, height-error) removed. The matrix form of fitting the linear rate for a single point is as follows:

$$\mathbf{V} \times \mathbf{D} = \Phi \quad (3)$$

where \mathbf{V} is the rate, \mathbf{D} is a matrix of temporal/time difference (days), and Φ is phase of a network of interferograms for this point.

B. Proposed Method

1) *Preprocessing of InSAR Signals:* InSAR signals are denoted by complex floating point arithmetic, including both the real and imaginary components [42]. According to the latest article [17], the imaginary component can be introduced with decorrelation noise before calculating wrap for simulation InSAR interferograms. The end-to-end simulation of InSAR signals is valuable, however, the generation of deformation phases is much more widely used, and various influencing factors can be added according to actual applications. For example, without adding terrain-related/topographic-errors phase [16] and adding terrain-related/topographic-errors phase [17], there are separate application scenarios. Therefore, this article only focuses on the simulation of InSAR real floating point numbers. Unlike the natural images using natural number arithmetic often in the range of 0–255, the conventional neural network training algorithms for image processing do not take into account the real

floating number processing [42]. Consequently, these algorithms can cause severe performance deterioration, training divergence, and undesirable results. We implemented several signal preprocessing methods and redesigned the network structure to adapt the real floating number data format of InSAR signals to solve this problem.

In addition, we designed conditional filter algorithms so that the neural network training algorithms can handle InSAR data. When the random patches do not meet the conditions in the data loader phase, they will directly be filtered out and will not be included in training. For screening the training patches, we defined a threshold value θ , which denotes the percentage of pixels within a patch with a motion value in the range $[-a, +a]$ cm/year, as illustrated in Fig. 4. We observed that when $\theta = 30\%$ and $a = 3$ cm/year, the proposed model produces the best results. For real-world datasets, the proposed model consistently produces the best visual outputs that are difficult to differentiate from target data samples. It is argued that this kind of algorithm, similar to strengthening features, will cause information loss. First, in a ground deformation application, nonregions of interest are discarded by the algorithm. Our model is expected to learn regions with motions that have practical significance. At the same time, we mainly use this algorithm at the initial stage of model training to make the model converge faster. When the generated results are close to convergence, we can turn OFF this algorithm or change the constraints in some training cycles. When the algorithm is used in practical applications or related fields, the parameters can be changed accordingly to fit the actual situation. For example, the conditional parameters can be loosened when the contour features are close to the real-world image and the texture details start to get close to the training image before the network model converges after a certain training period, which was chosen based on the network model and the training dataset. We can determine this specific training cycle by manually observing the model's real-time generated results and the evaluation criteria score.

Since our study focuses on motion rate, which is used to detect wide area ground deformation, we would expect a large amount of zero values, i.e., no deformation. Data distribution analysis of the real-world motion rate data confirms that most of the data area is zero (no motion). However, this is unsuitable for network training. With most areas zero, similar to some large areas in natural images, which are white (or black), and only a few areas have other values. It is challenging to learn the characteristics because the network constantly generates images with zero values are very close to convergence. In order for the network to pay attention to the change range of these values, we implemented a random value offset/shift preprocessing technique. To perform data preprocessing for training, the data loader module shifts the entire patch by a nonzero constant value randomly chosen from the range $[-5, 5]$. With the help of the random value offset/shift technique, our method is able to converge faster and get simulation close to realistic data. In order to avoid the loss of information or the possibility of changing the numerical distribution, we only use this algorithm at the beginning of network training (for example, the first ten epochs) and then terminate the algorithm.

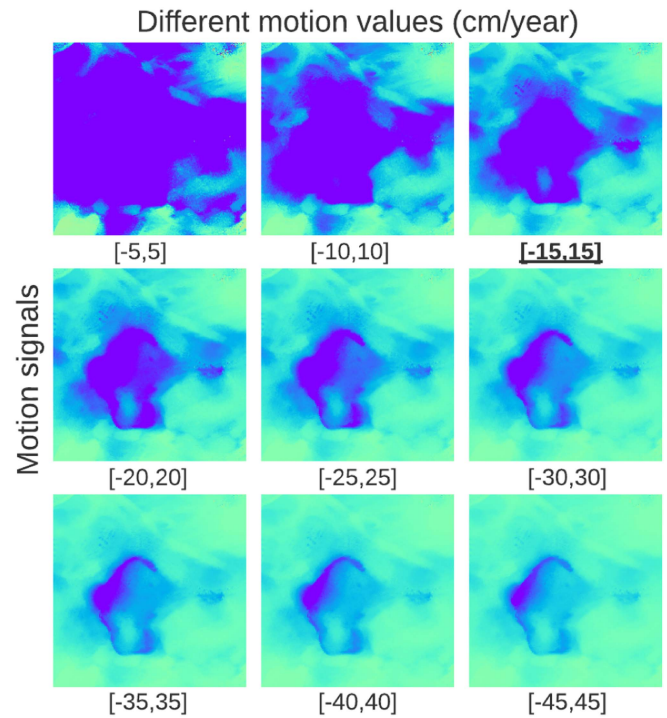


Fig. 3. Different motion values (cm/year).

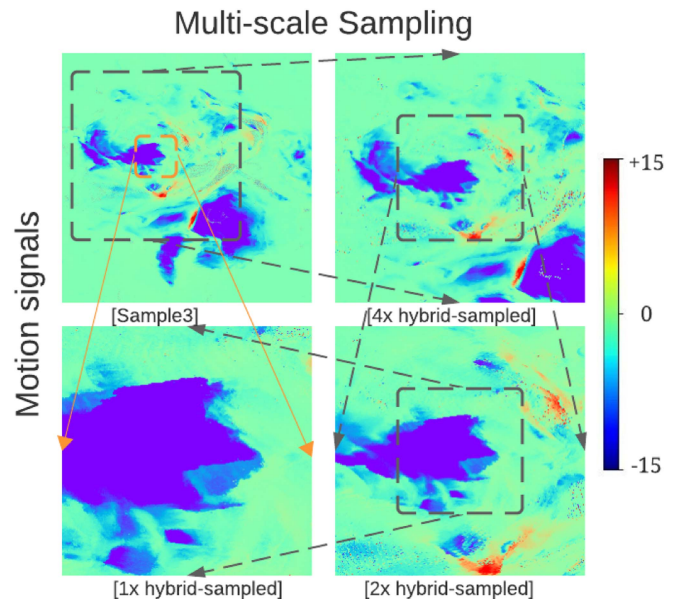


Fig. 4. Multiscale sampling.

2) *Data Augmentation and Multiscale Sampling*: In all learning methods, model quality depends on diverse and balanced training data. However, in our case, we lack sufficiently large-scale and diverse datasets. One reason is the huge amount of human effort and time required to collect ground-truth information manually, which is difficult to accomplish. The second reason is that there is no easy access to certain geolocations, such as hilly areas and dense forests, to increase data diversity. We introduce a data augmentation technique to increase the diversity of training data without collecting new samples. We

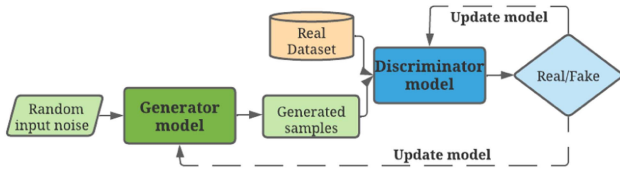


Fig. 5. Typical GAN architecture.

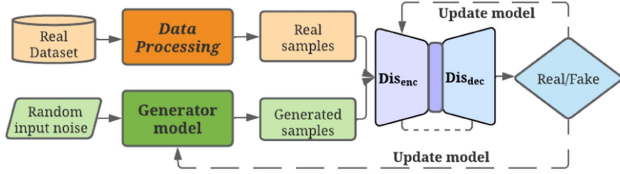


Fig. 6. Proposed network architecture for InSAR signals.

implemented image-level transformations, such as cropping, flipping, scaling, rotation, and so on [43].

To solve the variable value and spatial scale problem, we designed a multiscale / mixture patch technique for the network to learn multilevel and multiscale features. In the data loader phase, original patches with different hybrid-sampled rates represent real-world data at different scales in the discriminator. For example, 2x hybrid-sampled means the original input patch size of the data loader is doubled from 256×256 to 512×512 pixels, and then downsampled to 256×256 . This means the hybrid-sampled step does not affect the data quality (without losing information and introducing irrelevant information), as illustrated in Fig. 4. Meanwhile, this technology adapts the motion rate data to the receptive field of the network structure. The results show that the model generated results have more dense and small scale features. Consequently, the model trades off between different value scales and spatial scales of the signals.

3) *Proposed Network Architecture*: The typical GAN has two distinct models: generator and discriminator models, as illustrated in Fig. 5. The generator model spawns synthetic images that resemble the training images. The discriminator model assesses an image and decides whether it is a real-world training image or a synthetic (fake) image generated by the generator. During the training phase, the generator continuously improves the generated images to fool the discriminator, while the discriminator improves its ability to distinguish between real and fake images. At the same time, the adversarial game reaches equilibrium when the data generated by the generator are indistinguishable from the training data, and the discriminator can only guess whether the generator output is real or fake with a 50% confidence level. However, the existing GANs cannot be directly applied to signal generation applications. As shown in Fig. 6, we have added a data preprocessing module. We have optimized the backbone network, network modules, hyperparameter optimization.

Our proposed model is based on UNetGAN model, which is also based on the BigGAN model. The generators in the network structure of BigGAN and UNetGAN models are based

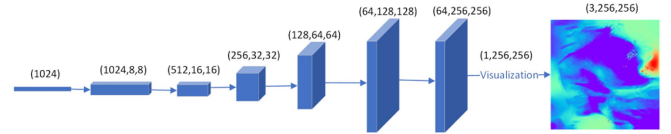


Fig. 7. Generator for InSAR signals.

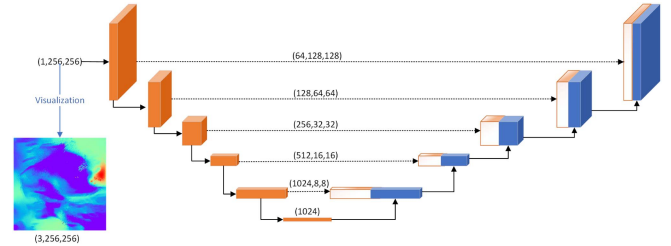


Fig. 8. U-Net discriminator for InSAR signals.

on residual block. Resnet blocks (residual blocks / residual connections) have been successful in deep learning, particularly in image recognition [44]. Resnet blocks can train deeper networks and help obtain better results. The generator introduces some new modules to make the distribution of the generated signals closer to the real-world data, as illustrated in Fig. 7. The main contribution of the UNetGAN model, utilizing an alternative U-Net-based discriminator, can encourage the GAN model to focus on both the global structure and local detail information of images using the principle of segmentation [36], as illustrated in Fig. 8. We made changes in the generator model to adapt the InSAR signals. The last activation layer was removed to meet the floating point arithmetic output requirements. The new U-Net discriminator reuses the original discriminator classification architecture as the encoder and its decoder part is built by the building blocks of the generator network [36], [45], as illustrated in Fig. 6. It provides detailed per-pixel and global image responses to the generator while maintaining the local and global coherence of the generated images. Meanwhile, the per-pixel feedback of the discriminator, a per-pixel consistency regularization technique (CutMix data enhancement), was designed to help the U-Net discriminator pay more attention to the semantic and structural changes [36]. Our proposed model has the ability to generate synthetic images with simultaneously balanced object shape and texture, global and local coherence features, which cannot be distinguished even by domain experts.

4) *Hybrid Evaluation Metric*: Since there is few GAN related research in the InSAR literature, we considered the evaluation metrics for generative models [46]. We explored the feasibility of combining visual assessment with quantitative evaluation. Our industrial domain experts do not find it practical to use the common quantitative evaluation methods in the computer vision domain, i.e., comparing the numerical size of sample scores generated by multiple models, and they recommend the visual inspection approach for real-world InSAR applications. We introduced hybrid evaluation metric, integrating qualitative and quantitative measures, to evaluate the results. The qualitative evaluation step t-distributed stochastic neighbor embedding

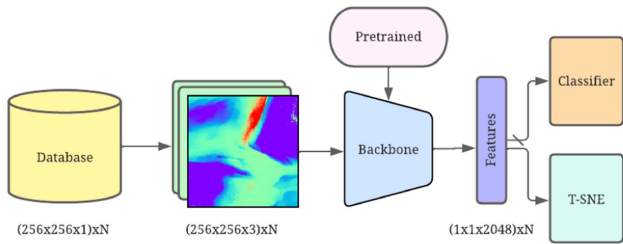


Fig. 9. Our qualitative evaluation pipeline: the pretrained Inception v3 network backbone extracts feature maps from images, and then T-SNE visualization analyzes the images. The pipeline generates the inference using T-SNE.

(T-SNE) [47], [48] allows domain experts to participate more intuitively in evaluating and comparing various models. They then make the validate their visual assessment with the quantitative evaluation metric (FID) [49].

Visual/manual inspection [50], [51] can be time-consuming, subjective (incorporate domain experts’ own biases and opinions) and cannot handle a huge diverse sample. We introduced T-SNE as a qualitative evaluation method to address this issue [47]. T-SNE is a nonlinear dimensionality reduction technique that transforms the visualization of high-dimensional data in a low-dimensional space (2-D or 3-D). Initially, T-SNE constructs a probability distribution that captures the mutual distance relationships between points in the initial high-dimensional space. Subsequently, the algorithm creates a low-dimensional space with similar relationships between the points. We employ T-SNE to visualize the embeddings and feature vectors generated by a pretrained neural network model [48]. The T-SNE analysis plots (visualization in low-dimensional space) can illustrate the similarity of each model to the real-world data distribution. When there is more overlap and a smaller distance between the real-world data point set and the compared model point set, this model distribution is more similar to the real-world data distribution. On the contrary, the other model’s generated images are less similar to the real-world data’s distribution when there is less overlap and the distance between class points is greater. It should be noted that the distance only works within the same T-SNE analysis plot, and comparing the distances of different plots does not provide useful information. The T-SNE analysis method can further investigate and analyze outliers or anomalies. Since each point is an image patch, we can look at the image features that the outliers represent point by point and teach the model to focus on features that are closer to those of domain experts and real-world application requirements. More details can be found in Section IV. The limitation is that its similarity distance is difficult to quantify and is not a numerical evaluation criterion. Therefore, we have the quantitative measure in our hybrid evaluation approach.

The Inception v3 model is a convolutional neural network trained on more than a million images from the ImageNet dataset to learn rich feature representations for a wide range of images. A classification network usually has a backbone that pulls out features from an image and a classifier that uses this information to make classification predictions. To assess the network’s feature map data, we executed the network inference up to the layer

right before the classifier layer. To extract the last feature map before the classifier head, we reimplemented the Inception v3 model by running the inference without the final classifier. This model lets us load the weights that have already been trained on ImageNet so that we can get 2048-D feature maps. Then, we used T-SNE to analyze and visualize the feature maps extracted by the Inception v3 model. Our qualitative evaluation pipeline is shown in Fig. 9.

For generative models, the two most popular quantitative evaluation methods are IS (IS) [52] and FID [49]. Both are single-value metrics and rely on a pretraining classification model trained on ImageNet. IS [52] computes the Kullback–Leibler divergence [53] between the conditional class distribution (quality) and the marginal class distribution (diversity) for each synthetic image. A major drawback is that IS does not capture how synthetic images compared to real-world images. FID [49] calculates the Fréchet distance (squared Wasserstein distance) between multivariate Gaussians fitted to the embedding space of generated and real-world images. As the quantitative evaluation criterion of the generator, FID can balance the realism and diversity of generated images. It is consistent with human inspection and sensitivity to modest changes (such as small artifacts and slight blurring) in the real distribution. Furthermore, it is able to detect intraclass mode collapse in the generative models’ training phase. It is important to note that FID needs a large sample size (normally above 50 K) to achieve a reliable FID score. A smaller number of samples can result in overestimation [54].

The calculation of FID involves computing the Fréchet distance between two multivariate Gaussian distributions fitted to the feature representations of the Inception v3 network [49]. The following formula calculates the FID between two multivariate Gaussians $X_1 \sim N(\mu_1, C_1)$ and $X_2 \sim N(\mu_2, C_2)$:

$$d^2(X_1, X_2) = \|\mu_1 - \mu_2\|^2 + \text{Tr}(C_1 + C_2 - 2 * (C_1 * C_2)^{1/2}). \quad (4)$$

In the activations of the coding layer of the Inception v3 model, X_1 and X_2 denote the generated and real-world samples, respectively. μ_n and C_n represent the mean and covariance of the coding layer activations for all generated and real-world samples. To ensure full rank covariance and avoid complex numbers or the “nans” problem when calculating the square root, the number of samples used to compute the Gaussian statistics should exceed the dimension of the coding layer, which was extracted as 2048-D feature maps by the Inception v3 model without the final classifier. We deployed a visual assessment and introduced a hybrid evaluation approach, integrating qualitative and quantitative measure, to evaluate the outcome for the InSAR research field. The details of the evaluation criteria comparison results are presented in Section IV-B.

III. EXPERIMENTS

To establish a ground truth for comparison, we used the linear deformation rate data from our industrial collaborators, which are processed based on real-world datasets captured by TerraSAR-X in StripMap mode [55]. The color scale of each

pixel in the image represents the ground linear motion rate (cm/year) of the corresponding Earth location. The real-world dataset includes nine stacks of motion rates close to clean reference data as training samples to train a GAN simulator. Each original stack has a dimension/resolution of 1500×1500 pixels, which exceeds the computing power and memory limits of GPUs and fails previous state-of-the-art GAN models. Even for good performance complex generative models, which can run large images, the time performance is not acceptable. To address this issue, we randomly selected patches with the size of 256×256 pixels from nine real-world motion rate signals. We trained the model on motion rate signals in order to assess the model's learning capacity, and the results are shown in Section IV Results.

We implemented the network using PyTorch Lightning and Hydra with PyTorch backend. The implementation was executed on an NVIDIA 1080 GPU with 8 GB GPU RAM. We trained the DGMs by extracting patches with the size of 256×256 pixels from nine stacks (1500×1500 pixels) real-world motion rate interferograms, as illustrated in Figs. 16–18. In our proposed model, we employed the original training parameters of BigGAN and UnetGAN, utilizing the Adam optimizer [56] with learning rates of $1e-4$ for the Generator and $5e-4$ for the discriminator. The training process extended for 81 epochs. Our stopping criteria are based on the hybrid evaluation criteria mentioned in our article, which involves both qualitative (T-SNE) and quantitative (FID) evaluations. During training, we observed a convergence trend starting at 76 epochs, indicating that the model was stabilizing. However, we noticed that the generated features were somewhat dense. To address this, we turned OFF the preprocessing selection patch condition, and by the end of the 81st epoch, the model was able to generate samples with both sparse and dense features. The final model, used for evaluating the proposed model samples in the article, was derived at this point.

IV. RESULTS

A. Visual Evaluation

This section empirically demonstrates the effectiveness of the proposed model using real-world ground deformation scenarios for generating large-scale data in InSAR signal analysis. Fig. 10 shows unwrapped samples, and Fig. 11 shows the samples with phase wrapping $[-\pi, +\pi)$. In the middle columns of Figs. 10 and 11, we use the latest InSAR simulator described in [8], [9], [10], which is regarded as the state-of-the-art, to generate synthetic ground deformation maps as the baseline for comparison. The baseline includes various signal geometries: 1) triangle, 2) Gaussian bubble, 3) rectangle, and 4) ellipse. These geometric shapes are stochastically created and added together to produce arbitrary combinations of various irregular spatial features, as illustrated in the middle columns of Figs. 10 and 11. In the right columns of Figs. 10 and 11, the features of the motion signals generated by our proposed network are more similar to real-world motion signals and resultant interferograms, without introducing artificial features or features that do not exist in the real-world motion signals (as highlighted by the dotted

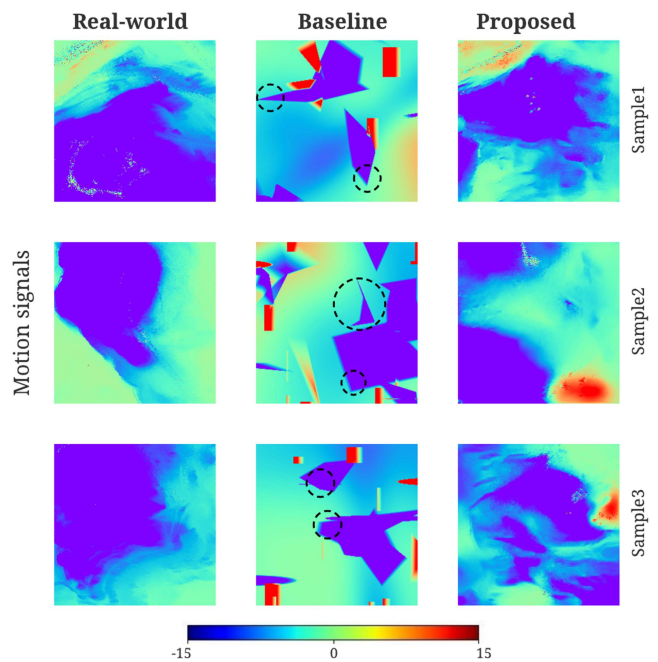


Fig. 10. From left to right, we show samples of the real-world motion signals, simulated motion signals (baseline [8], [9]), and motion signals generated by the proposed network. All samples have a value range of $[-15, +15]$ cm/year.

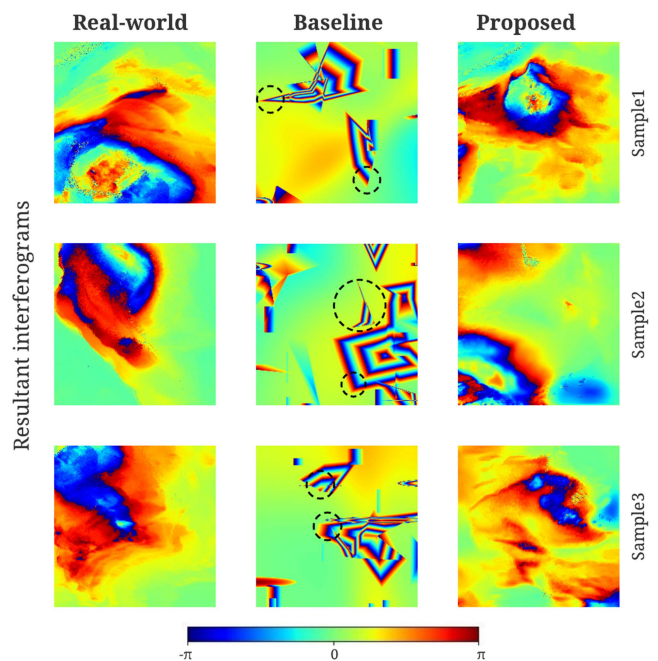


Fig. 11. From left to right, we show samples of the real-world resultant interferograms, simulated resultant interferograms (baseline work) [8], [9], and resultant interferograms generated by the proposed network. All samples are the interferometric of phase wrapping $[-\pi, +\pi)$ (Blue: $-\pi$; Red: $+\pi$).

circles in Figs. 10 and 11). The following visual inspection and quantitative evaluation demonstrate that our model is capable of mimicking the ground motions successfully and can capture the target signal distribution with adversarial training.

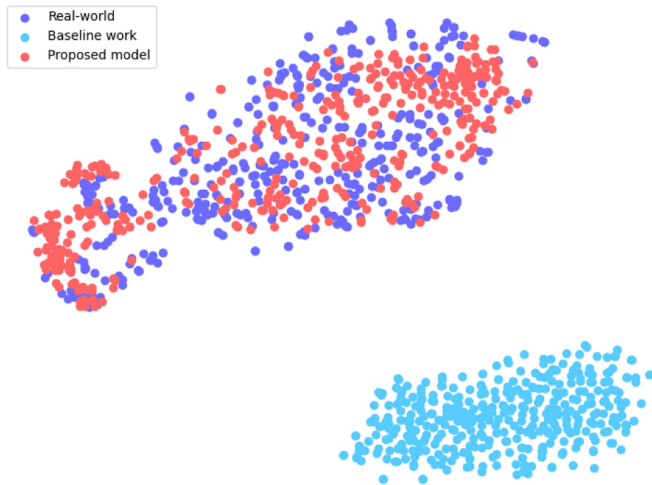


Fig. 12. T-SNE visualization for comparison of real-world, baseline work [8], [9], and proposed model (see Section II-B3). The purple and red points overlap closely, while the blue points are far from the real-world data. Visual analysis shows that the real-world and proposed model simulations are closer in features and textures than the current state-of-the-art [8], [9].

B. Hybrid Evaluation Metric

Although several approaches and measures have been introduced to evaluate generative modeling performance, there is no universally agreed definition or benchmark for generative models [57]. Visual inspection is time-consuming and subjective and cannot capture distributional characteristics [58]. Distribution characteristics refer to the properties and patterns of the generated samples in relation to the target distribution. These characteristics capture the similarity, diversity, and overall fidelity of the generated samples indicating how well the GAN model captures and reproduces the underlying data distribution compared to the real data.

We introduced a hybrid evaluation criterion, integrating qualitative and quantitative measures, to assess the performance of our InSAR simulator. The qualitative evaluation is T-SNE, and the quantitative evaluation is derived from FID. T-SNE is a powerful visualization technique to transform data statistics into lower-dimensional spaces. T-SNE is a nonlinear dimensionality reduction technique. Unlike principal component analysis, which involves an iterative optimization that takes time to converge and few parameters can be tweaked [59]. T-SNE constructs a probability distribution that captures the mutual distance relationships between points in the initial high-dimensional space and then transforms this characteristic into a low-dimensional space while preserving these relationships. By applying T-SNE to these features and analyzing the results, we found that the projection kept the original distance relationships between points in high- and low-dimensional space. For example, if two points were close in the high-dimensional space at the start, they stayed close in the low-dimensional space after the projection. If they were far apart in high-dimensional space, they remained far apart in low-dimensional space. We are able to show that the projections from our model and the real-world representation are highly overlapped (see Fig. 12), indicating that the proposed

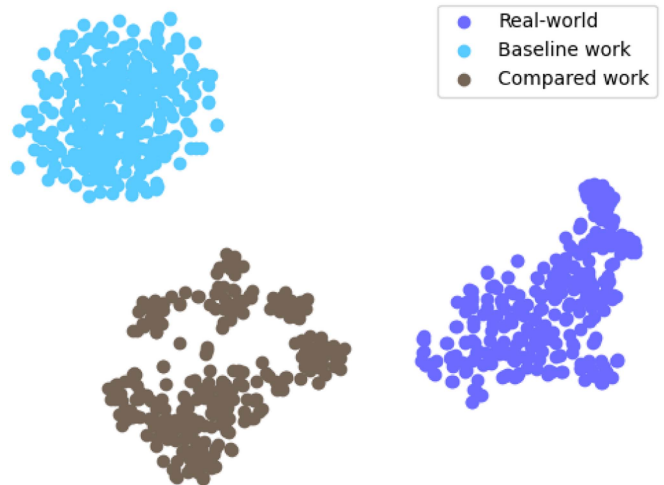


Fig. 13. T-SNE visualization for comparison of real-world, baseline work [8], [9], and compared work [16], [17]. The purple and gray points are relatively close, while the blue points are far from the real-world data. Visual analysis shows that there are some differences in the features between the real world and the compared work, and the differences between the real world and baseline work features are even greater.

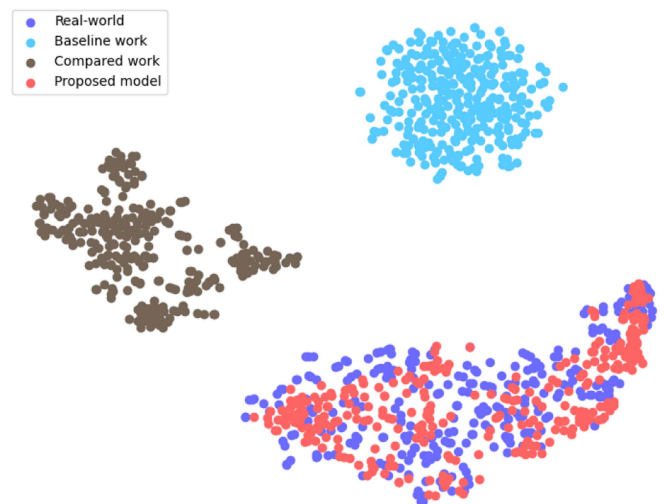


Fig. 14. T-SNE visualization for comparison of real-world, baseline work [8], [9], compared work [16], [17], and proposed model (see Section II-B3). The purple and gray points overlap closely, while the blue points and gray points are far from the real-world data. Visual analysis shows that the real-world and proposed model simulations are closer in features and textures than the current state-of-the-art papers of each of the two approaches [8], [9], [16], [17].

model performs better than previous work [8], [9]. The Fig. 13 shows that there are some differences in the features between the real world and the compared work [16], [17], and the differences between the real world and baseline work features are even greater. Fig. 14 shows that the real-world and proposed model simulations are closer in features and textures than the current state-of-the-art papers of each of the two approaches [8], [9], [16], [17].

FID is a commonly used metric for assessing the similarity between two sets of images, such as those GANs generated samples. The FID measure has been shown to correlate well with

TABLE I
FRECHET INCEPTION DISTANCE (FID, LOWER IS BETTER)

Baseline work [9]	Compared work [17]	Proposed model
283.27	193.64	35.40

human perception of visual quality. The lower the FID score, the more similar the generated samples are to real-world data. Table I shows that our proposed model (35.4) is much better than the previous work (283.27) [8], [9] and compared work (193.64) [16], [17], and is close to the distribution of real-world and training data. Our coarse-grained and fine-grained (hybrid) evaluation metrics are based on the same pretraining network model. Therefore, in essence, the feature maps of the comparison between the two evaluation criteria are the same. However, quantitative scores may or may not represent significant differences, and therefore we introduce a hybrid approach to add visual analysis (t-SNE) to elaborate numeric score (FID).

To summarize, we propose a hybrid assessment metric, integrating qualitative and quantitative measures, to evaluate the outcome for InSAR analysis. It combines the advantages of visual and quantitative evaluation. We adopt this hybrid approach as recommended by industrial domain experts.

V. DISCUSSION

In InSAR, high-level fringe frequency indicates regions on the ground that are moving quickly, which can cause the interferogram to have many phase jumps from $-\pi$ to $+\pi$ when it is wrapped. The ability to preserve structural detail is essential for effective InSAR simulators [8], [9], and our proposed GANInSAR demonstrates this capability. Table I shows that the proposed GANInSAR can generate interferogram maps similar to the training data structural details, whether unwrapped or wrapped. Furthermore, it demonstrates that our method can precisely simulate high-contrast features and clear boundaries in real-world site interferogram maps. We can clearly find that the results generated by the proposed model are better than baseline work [8], [9] and compared work [16], [17], and it is almost impossible to distinguish the proposed model generated data from the real-world data. Previous studies have mainly focused on generating simulated signals with randomized motion patterns with basic geometric features. However, these simulators still cannot capture the full complexity and diversity of real-world InSAR data. The results of both quantitative and qualitative evaluations confirm that our trained GANInSAR model can generalize the multiscale signal features (details can be found in Section II-B2). The proposed GANInSAR can effectively alleviate the inability to use data-hungry deep learning techniques due to the lack of data.

In general, our InSAR simulator based on the GAN demonstrates a significant advancement in generating more realistic synthetic data, in the absence of clean ground truth data of real-world images. Even the latest simulators can only generate random composite signals under the handcrafted traditional simulators based on mathematical models [8], [9], [16], [17], as illustrated in the Figs. 1 and 2. This objective condition adversely

affects the evaluation of the learning ability and generalization ability of the deep learning network in this field applied to the simulated data scenarios. The convolutional neural networks, being a data-driven technique, need diverse and large-size real-world training data when trying to solve complex tasks, in particular, when applying to real-world InSAR data. Existing simulators are designed to randomly generate basic geometric features as training data. In order to enhance the complexity of primitive features, some articles add randomly cropped DEM/InSAR interferograms data on the basic geometric features [17]. However, they still cannot fully simulate real-world scenarios' complex features. The GANInSAR framework has the potential to learn the distribution of simulated data and real-world data, even when the real-world data is from a small dataset with only nine stacks. Conversely, the details can be found in Section IV Results, and Figs. 1 and 2. By learning the data distribution from a limited dataset, a massive amount of simulation data close to the original data distribution is generated, thus giving researchers enough data to apply various deep neural network models to solve practical application problems.

The data generated by our generative model will help phase unwrapping and filtering applications, which is also one of the motivations for our research. Our lab is doing phase unwrapping and filtering research, and one of the related papers [60] was published in October 2023. Traditional approaches, which separate unwrapping and filtering algorithms, often introduce errors and changes in signal statistics. To overcome this, the paper [60] introduces a novel two-stage phase unwrapping deep neural network framework based on U-Net, enabling joint unwrapping and denoising of InSAR phase images. Integrating our data generation model with phase unwrapping and filtering work is one of our future research areas. We also noticed the relationship between the topographic phase and the deformation phase. We are following the common practice in the InSAR field described in the articles [16], [17]. The author mentioned in the paper [16] that their deformation part is a randomly generated distorted 2-D Gaussian surface. In order to enhance the complexity of primitive features, another paper from their lab adds randomly cropped DEM [17] as the topographic phase. We understand the reviewer's concern regarding the clarity of our motivation for using GANs to generate a deformation phase dataset. We have provided a new evaluation criterion and compared the generated samples from the two articles [16], [17]. The experiment showed that our results exceeded the samples generated by the previous articles [16], [17]. Nevertheless, we do not want to limit our technique to this application. We believe that further discussions on future work can benefit the development of this field and attract more scholars to participate in the research.

Ablation experiments are essential to validate the benefits of the modifications introduced to the deep learning model. These experiments will be designed to systematically analyze the impact of each modification on the model's performance, providing a more robust evaluation. Especially in the more mature tasks of recognition, classification, detection, segmentation, and tracking. Due to the availability of a large number of comparison models and a large number of open datasets, any model alteration ablation experiments have become a mandatory

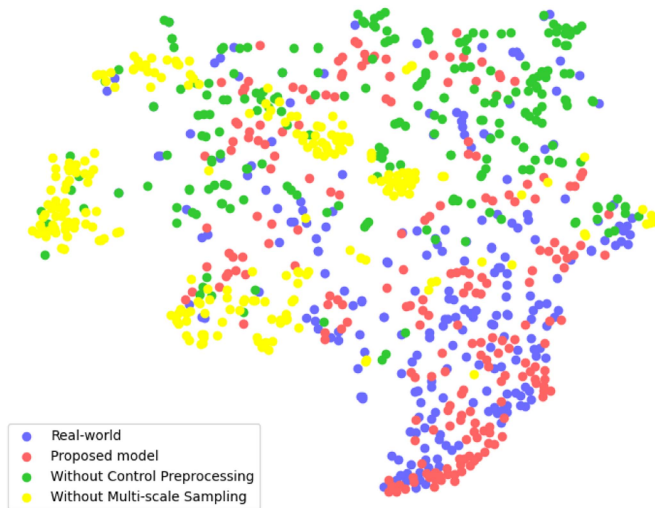


Fig. 15. T-SNE visualization for comparison of real-world, proposed model (see Section II-B3), without control preprocessing and without multiscale sampling. The purple and red points overlap closely, while the green points and yellow points are far from the real-world data. Visual analysis shows that the real-world and proposed model simulations are closer in features and textures than the without control preprocessing and the without multiscale sampling approaches.

TABLE II
FRECHET INCEPTION DISTANCE (FID, LOWER IS BETTER)

Proposed model	Without Control Preprocessing	Without Multi-scale Sampling
35.40	79.95	71.50

item for experiments. The modifications mentioned in our article are mostly unavoidable modifications. If the modifications mentioned previously are not made, the model cannot converge and learn any features. We encountered many challenges in order to conduct ablation experiments. Meanwhile, due to the lack of reference models and baseline datasets, and almost all generated models being truly massive public datasets, the data types and numerical distributions are completely different from our training data (as mentioned in the preprocessing section). We overcame the above-mentioned challenges and conducted these experiments to provide a more thorough evaluation of our approach. By applying T-SNE to these features and analyzing the results, Fig. 15 shows that the real-world and proposed model simulations are closer in features and textures than the without Control Preprocessing and the without Multi-scale Sampling approaches.

The lower the FID score, the more similar the generated samples are to real-world data. Table II shows that our proposed model (35.4) performs better than ablation experiments without control preprocessing module variables and turns OFF multiscale sampling modules. In the ablation experiments, we conducted the following two subexperiments: First, in the original experiment, we closed the preprocessing module at epoch 76 in order to produce samples that could contain sparse features. In this subexperiment, we did not disable the preprocessing and patch selection conditions to continue training. After loading the parameters of 76 epochs and continuing to train to 81 epochs,

the FID score obtained was 79.94, which is far worse than the proposed model's 35.40. In the second subexperiment, we trained a multiscale sampling module and retrained 81 epochs, resulting in an FID score of 71.50, which is also not as good as the proposed model's 35.40. The above-mentioned ablation experiment strengthens the validity of our originally proposed model solution, and the experimental results demonstrate the effectiveness of the two modules (preprocessing and multiscale sampling modules) in the proposed model solution.

Deep learning is a data-hungry technology, providing deep learning models with more diverse and higher quality data can improve the stability and robustness of deep learning models and facilitate the extension of deep learning to various potential and new application scenarios. ImageNet [61] dataset, which collected massive data from the Internet, makes the performance and accuracy of deep learning algorithms trained on ImageNet [61] exceed the performance of machine learning algorithms. However, many deep learning-based algorithms in the field of remote sensing, especially InSAR-related research, have failed to perform well due to the lack of this specific type of data. Two previously published papers [5, 6] from our lab mentioned that a small amount of real-world captured data is more effective than simulated data. These two papers [9], [10] were applied to phase filtering and coherence estimation [9], ground deformation and DEM error estimation in InSAR time series data [10], respectively. This is also one of the motivations that drives us to apply generative models to InSAR. These two articles use the same simulator for data simulation, and their simulator is used as the baseline model in this article for comparison. In future work, we will improve our generative model to test more different application scenarios.

VI. CONCLUSION

We propose GANInSAR, a GAN-based DGM for large-scale InSAR motion signal synthesis. This work provides three main contributions as follows.

- 1) Designing a DGM architecture suitable for InSAR motion signal synthesis using only limited training data.
- 2) Implementing preprocessing and postprocessing modules for DGM architecture suitable for InSAR motion signal synthesis.
- 3) Introducing hybrid evaluation criteria for effective evaluation of synthetic signals.

This work focuses on synthesizing large-size and diverse samples and keeping the InSAR spatial information to compensate the limited data in this field. Our method was trained and evaluated on real-world motion signals. In both cases, the model successfully generated synthetic InSAR motion signals that are difficult to distinguish from target training signals. Visual comparisons in real-world scenarios reveal that our method outperforms the handcrafted traditional simulators based on mathematical models [3], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17]. In addition, our innovation introduced a hybrid evaluation criterion, integrating qualitative and quantitative measures, to assess our InSAR simulator. The qualitative step is based on T-SNE, and the quantitative step is based on FID. The

evaluation shows that the proposed model is able to generate realistic new features that have not been seen before. In conclusion, this study demonstrates that our proposed framework successfully mitigates the research gap of lacking or limited-collected ground-truth data in the InSAR signal applications. Moreover, our model can generate large-size and diverse samples with multispatial features, making it difficult for domain experts to distinguish our results from the original data. These illustrate the capability of our DGM for InSAR motion signal synthesis. At the same time, our model can be seamlessly integrated with traditional methods, allowing for the addition of atmospheric phases, topographic phases, orbital phases, decorrelation phases, and more. The adversarial DGM architectures are effective for generating a large amount of diversity simulation data for InSAR signal analysis. We demonstrate that it can generate deformation data and should also be able to generate other InSAR phase components. In future work, we will extend our model to generate the diversity of InSAR phase components, such as atmospheric, topographic, orbital, and decorrelation. Meanwhile, recent research on transformer neural networks [62] instead of convolutional neural networks show that they may become the next generation models of computer vision tasks, such as classification, detection, and segmentation. We also plan to investigate an InSAR simulator that combines GAN and transformer networks [63] with the goal to generate more diverse and sophisticated synthetic for other applications.

VI. ACKNOWLEDGMENT

The authors would like to acknowledge the anonymous reviewers for their valuable suggestions that have helped to improve this manuscript.

APPENDIX NINE STACKS OF REAL-WORLD MOTION DATA

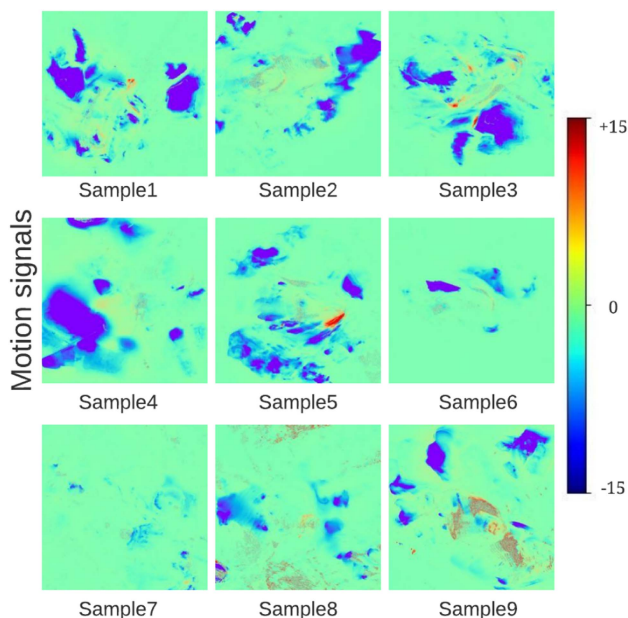


Fig. 16. All samples have a value range of $[-15, +15]$ cm/year.

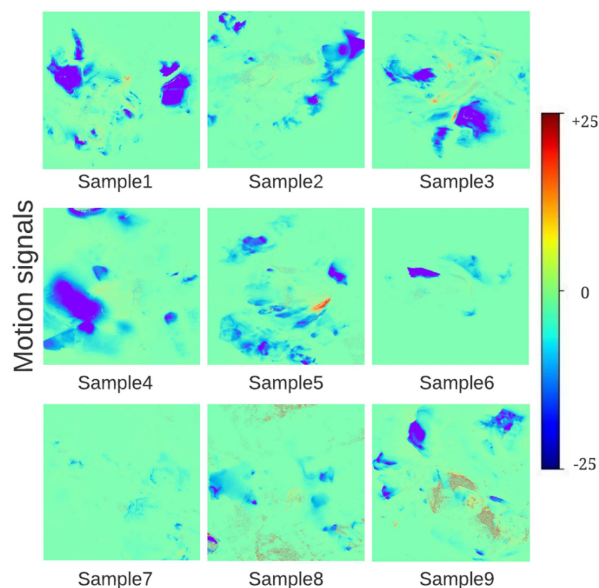


Fig. 17. All samples have a value range of $[-25, +25]$ cm/year.

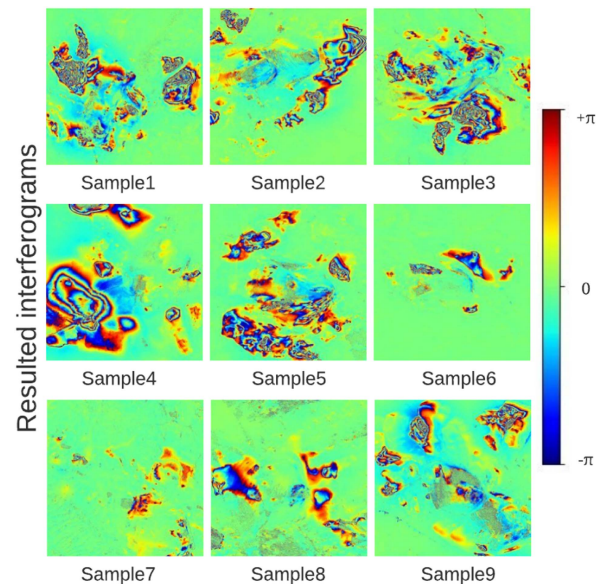


Fig. 18. All samples are the interferometric of phase wrapping $[-\pi, +\pi]$ - (Blue: $-\pi$; Red: $+\pi$).

REFERENCES

- [1] F. Mohammadimanesh, B. Salehi, M. Mahdianpari, B. Brisco, and M. Motagh, "Wetland water level monitoring using interferometric synthetic aperture radar (InSAR): A review," *Can. J. Remote Sens.*, vol. 44, no. 4, pp. 247–262, 2018.
- [2] J. J. Sousa, A. J. Hooper, R. F. Hanssen, L. C. Bastos, and A. M. Ruiz, "Persistent scatterer InSAR: A comparison of methodologies based on a model of temporal deformation vs. spatial correlation selection criteria," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2652–2663, 2011.
- [3] W. Duan, H. Zhang, and C. Wang, "Deformation estimation for time series InSAR using simulated annealing algorithm," *Sensors*, vol. 19, no. 1, p. 115, 2019. [Online]. Available: <https://www.mdpi.com/1424-8220/19/1/115>
- [4] S. Li, S. Zhang, T. Li, Y. Gao, Q. Chen, and X. Zhang, "Modeling the optimal baseline for a spaceborne bistatic SAR system to generate DEMs," *ISPRS Int. J. Geo- Inf.*, vol. 9, no. 2, p. 108, 2020. [Online]. Available: <https://www.mdpi.com/2220-9964/9/2/108>

- [5] S. Samsonov, A. Dille, O. Dewitte, F. Kervyn, and N. d'Oreye, "Satellite interferometry for mapping surface deformation time series in one, two and three dimensions: A new method illustrated on a slow-moving landslide," *Eng. Geol.*, vol. 266, 2020, Art. no. 105471.
- [6] F. Sica, D. Cozzolino, X. X. Zhu, L. Verdoliva, and G. Poggi, "InSAR-BM3D: A nonlocal filter for SAR interferometric phase restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3456–3467, Jun. 2018.
- [7] S. Mukherjee, A. Zimmer, X. Sun, P. Ghuman, and I. Cheng, "An unsupervised generative neural approach for InSAR phase filtering and coherence estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 11, pp. 1971–1975, Nov. 2021.
- [8] X. Sun, A. Zimmer, N. K. Kottayil, S. Mukherjee, and I. Cheng, "A benchmark InSAR simulator for phase filtering and coherence estimation," in *Proc. Conf. Arabian J. Geosciences*, 2020, pp. 11–13.
- [9] X. Sun, A. Zimmer, S. Mukherjee, N. K. Kottayil, P. Ghuman, and I. Cheng, "DeepInSAR—A deep learning framework for SAR interferometric phase restoration and coherence estimation," *Remote Sens.*, vol. 12, no. 14, p. 2340, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/14/2340>
- [10] X. Sun, A. Zimmer, S. Mukherjee, P. Ghuman, and I. Cheng, "IGSCMAES: A two-stage optimization for ground deformation and DEM error estimation in time series InSAR data," *Remote Sens.*, vol. 13, no. 13, p. 2615, 2021. [Online]. Available: <https://www.mdpi.com/2072-4292/13/13/2615>
- [11] N. Anantrasirichai et al., "Detecting ground deformation in the built environment using sparse satellite InSAR data with a convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 2940–2950, Apr. 2021.
- [12] G. Rongier, C. Rude, T. Herring, and V. Pankratius, "Generative modeling of InSAR interferograms," *Earth Space Sci.*, vol. 6, no. 12, pp. 2671–2683, 2019.
- [13] G. Rongier, C. Rude, T. Herring, and V. Pankratius, "An attempt at improving atmospheric corrections in InSAR using cycle-consistent adversarial networks," *EarthArXiv*, vol. 11, 2020, doi: [10.31223/X5M594](https://doi.org/10.31223/X5M594).
- [14] B. Rouet-Leduc, R. Jolivet, M. Dalaison, P. A. Johnson, and C. Hulbert, "Autonomous extraction of millimeter-scale deformation in InSAR time series using deep learning," *Nature commun.*, vol. 12, no. 1, pp. 6480, 2021. [Online]. Available: <https://www.nature.com/articles/s41467-021-26254-3>
- [15] C. M. Brengman and W. D. Barnhart, "Identification of surface deformation in InSAR using machine learning," *Geochemistry Geophys. Geosystems*, vol. 22, no. 3, 2021, Art. no. e2020GC009204.
- [16] Z. Wu, T. Wang, Y. Wang, R. Wang, and D. Ge, "Deep learning for the detection and phase unwrapping of millimeter-scale deformation in large-scale interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5216318.
- [17] Z. Wu, T. Wang, Y. Wang, R. Wang, and D. Ge, "Deep-learning-based phase discontinuity prediction for 2-D phase unwrapping of SAR interferograms," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5216516.
- [18] N. Anantrasirichai, J. Biggs, F. Albino, and D. Bull, "A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets," *Remote Sens. Environ.*, vol. 230, 2019, Art. no. 111179.
- [19] I. Baran, M. Stewart, and S. Claessens, "A new functional model for determining minimum and maximum detectable deformation gradient resolved by satellite radar interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 675–682, Apr. 2005.
- [20] K. Perlin, "An image synthesizer," *ACM SIGGRAPH Comput. Graph.*, vol. 19, no. 3, pp. 287–296, 1985.
- [21] L. Ruthotto and E. Haber, "An introduction to deep generative modeling," *GAMM-Mitteilungen*, vol. 44, 2021, Art. no. e202100008.
- [22] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2015, pp. 1530–1538.
- [23] I. Kobyzev, S. Prince, and M. Brubaker, "Normalizing flows: An introduction and review of current methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 3964–3979, Nov. 2021.
- [24] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. 2nd Int. Conf. Learn. Representations*, Banff, AB, Canada, 14–16, 2014.
- [25] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic backpropagation and approximate inference in deep generative models," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2014, pp. 1278–1286.
- [26] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *Found. Trends Mach. Learn.*, vol. 12, no. 4, pp. 307–392, 2019, doi: [10.1561/22000000056](https://doi.org/10.1561/22000000056).
- [27] I. Goodfellow et al., "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [28] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014. [Online]. Available: <https://arxiv.org/abs/1411.1784>
- [29] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. 4th Int. Conf. Learn. Representations*, San Juan, Puerto Rico, May 2–4, 2016.
- [30] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [31] Y. Ma, K. Liu, Z. Guan, X. Xu, X. Qian, and H. Bao, "Background augmentation generative adversarial networks (BAGANs): Effective data generation based on GAN-augmented 3D synthesizing," *Symmetry*, vol. 10, no. 12, p. 734, 2018. [Online]. Available: <https://www.mdpi.com/2073-8994/10/12/734>
- [32] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *Proc. Int. Conf. Learn. Representations*, 2018.
- [33] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4396–4405.
- [34] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2019, pp. 7354–7363.
- [35] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *Proc. Int. Conf. Learn. Representations*, 2019.
- [36] E. Schonfeld, B. Schiele, and A. Khoreva, "A u-net based discriminator for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8204–8213.
- [37] D. Berthelot, T. Schumm, and L. Metz, "BEGAN: Boundary equilibrium generative adversarial networks," 2017, *arXiv:1703.10717*.
- [38] A. Hooper, H. Zebker, P. Segall, and B. Kampes, "A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers," *Geophysical Res. Lett.*, vol. 31, no. 23, pp. 1–5, 2004.
- [39] R. F. Hanssen, *Radar Interferometry: Data Interpretation and Error Analysis*, vol. 2, Berlin, Germany: Springer Sci. and Bus. Media, 2001.
- [40] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.
- [41] C. Prati, A. Ferretti, and D. Perissin, "Recent advances on surface ground deformation measurement by means of repeated space-borne SAR observations," *J. Geodynamics*, vol. 49, no. 3/4, pp. 161–170, 2010.
- [42] N. K. Kottayil, A. Zimmer, S. Mukherjee, X. Sun, P. Ghuman, and I. Cheng, "Accurate pixel-based noise estimation for InSAR interferograms," in *Proc. IEEE SENSORS*, 2018, pp. 1–4.
- [43] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, 2019.
- [44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [45] U. Demir and G. Unal, "Patch-based image inpainting with generative adversarial networks," 2018, *arXiv:1803.07422*.
- [46] A. Borji, "Pros and cons of GAN evaluation measures: New developments," *Comput. Vis. Image Understanding*, vol. 215, 2022, Art. no. 103329.
- [47] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
- [48] L. Van Der Maaten, "Learning a parametric embedding by preserving local structure," in *Proc. 12th Int. Conf. Artif. Intell. Statist.*, PMLR, 2009, pp. 384–391.
- [49] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6629–6640.
- [50] E. L. Denton et al., "Deep generative image models using a Laplacian pyramid of adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1486–1494.
- [51] S. Zhou, M. Gordon, R. Krishna, A. Narcomey, L. F. Fei-Fei, and M. Bernstein, "HYPER: A benchmark for human eye perceptual evaluation of generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, Art. no. 310.
- [52] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2234–2242.
- [53] J. M. Joyce, "Kullback-Leibler divergence," in *International Encyclopedia of Statistical Science*, Berlin, Germany: Springer, 2011, pp. 720–722.
- [54] M. J. Chong and D. Forsyth, "Effectively unbiased FID and inception score and where to find them," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 6069–6078.

- [55] W. Pitz and D. Miller, "The TerraSAR-X satellite," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 2, pp. 615–622, Feb. 2010.
- [56] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, San Diego, CA, USA, May 7–9, 2015.
- [57] Z. Wang, Q. She, and T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2021.
- [58] D. Saxena and J. Cao, "Generative adversarial networks (GANs) challenges, solutions, and future directions," *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–42, 2021.
- [59] W. Li, J. E. Cerise, Y. Yang, and H. Han, "Application of t-SNE to human genetic data," *J. Bioinf. Comput. Biol.*, vol. 15, no. 4, 2017, Art. no. 1750017.
- [60] S. Vijay Kumar, X. Sun, Z. Wang, R. Goldsbury, and I. Cheng, "A u-net approach for InSAR phase unwrapping and denoising," *Remote Sens.*, vol. 15, no. 21, p. 5081, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/21/5081>
- [61] O. Russakovsky et al., "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211–252, 2015.
- [62] S. Khan, M. Naseer, M. Hayat, S. W. Zamir, F. S. Khan, and M. Shah, "Transformers in vision: A survey," *ACM Comput. Surv.*, vol. 54, 2021, Art. no. 200.
- [63] Y. Jiang, S. Chang, and Z. Wang, "TransGAN: Two pure transformers can make one strong GAN, and that can scale up," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 14745–14758.



Zhongrun Zhou (Graduate Student Member, IEEE) is currently working toward the Ph.D. degree in computing science with the Department of Computing Science, University of Alberta, Edmonton, AB, Canada.

He gained industrial experience at the Hong Kong Applied Science and Technology Research Institute before pursuing Ph.D. degree. His research interests include satellite remote sensing signal processing and simulation, generative models for computer vision, and applications of generative models for small data scenarios.



Xinyao Sun received the the Ph.D. degree in computing science from the University of Alberta, Edmonton, AB, Canada, in 2022.

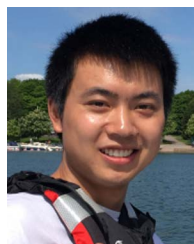
He is the CSO of Matrix Labs, Inc., and a Postdoctoral Fellow with the Multimedia Research Center, University of Alberta, Edmonton, AB, Canada. His Ph.D. research focused on accelerating the satellite signal processing pipeline for wide area monitoring in aerospace, utilizing deep learning and signal processing techniques. His DeepInSAR framework is now being used in the industry to serve clients all over

the world. In addition, with the University of Alberta, he mentors projects in the multimedia master's program, focusing on computational consensus blockchains, smart contract automation, decentralized financial systems, and cross-chain security.



Fei Yang received the B.S. and M.S. degrees in computing science from the University of Alberta, Edmonton, AB, Canada, in 2020 and 2022, respectively, where she is currently working toward the Ph.D. degree in computing science.

Her research interests include remote sensing image processing, computer vision, and their application.



Zheng Wang received the B.Eng. and M.Eng. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2012 and 2015, respectively, and the Ph.D. degree in Earth observation from Newcastle University, Newcastle upon Tyne, U.K., in 2019.

He is currently a Senior Data Scientist. His research interests include modeling and deep learning techniques in conjunction with remote sensing data to measure millimetre displacements and detect ground changes in the Earth's surface.



Ryan Goldsbury received the M.S. and Ph.D. degrees in astronomy and astrophysics from the University of British Columbia, Vancouver, BC, Canada, in 2011 and 2015, respectively.

He is the Chief Technology Officer with 3vGeomatics, Vancouver, BC, Canada. He was with 3vGeomatics in various roles, including Chief Technology Officer and Software Developer. His research interests include data analysis, image processing, and data modeling.



Irene Cheng (Senior Member, IEEE) received the Ph.D. degree in computing science from the University of Alberta (UofA), Edmonton, AB, Canada, in 2005.

She is currently a Professor and the Scientific Director with Multimedia Research Group, Department of Computing Science, UofA. She was a Postdoctoral Researcher with the University of Pennsylvania, Philadelphia, PA, USA. Her research interests include remote sensing and healthcare.

Prof. Cheng was a recipient of the UofA Alumni Horizon Award in 2008, which recognized her outstanding professional achievements. She was also a recipient of the Visiting Professorship of the highest rank by Institut National des Sciences Appliquées (INSA) de Lyon, France, in 2011. For more than 20 years, she delivered high quality scientific results, which include setting up the CROME (Computer Reinforced Online Multimedia Education) project and an interdisciplinary course for modeling and visualization, inventing the trademark "Zoomage" for a high resolution scanned device and imagery, bringing the Consortium for Aerospace Research and Innovation in Canada (CARIC) funding to the province. By extending her remote sensing research, her team contributes to various applications, e.g., Sustainable Agriculture and Environment.