remote sensing

MDPI

Article

OptiX-Based SAR Image Simulation with Residual-Learning Refinement

Ming Yang [1,2,3], Yueting Zhang [1,2,*], Guangyao Zhou [1,2], Jiayi Guo [1,2], Xia Ning [1,2], Yansong Wang [1,2,3] and Peifeng Li [1,2]

[1] Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; yangming24@mails.ucas.ac.cn (M.Y.); zhougy@aircas.ac.cn (G.Z.); guojy@aircas.ac.cn (J.G.); ningxia@aircas.ac.cn (X.N.); wangyansong23@mails.ucas.ac.cn (Y.W.); lipf@aircas.ac.cn (P.L.)
[2] Key Laboratory of Target Cognition and Application Technology, Chinese Academy of Sciences, Beijing 100190, China
[3] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: zhangyueting06@mails.ucas.ac.cn

Highlights

What are the main findings?

- An OptiX-based physical simulator combines ray tracing, Monte Carlo material-scattering modeling, and coherent image formation to generate physically interpretable multi-view SAR images.
- The integrated OptiX simulation and residual-refinement pipeline preserves large-scale structural consistency while compensating for fine-scale discrepancies, increasing the average PSNR from 16.63 to 23.88 dB and SSIM from 0.76 to 0.83, while reducing LPIPS from 0.24 to 0.20 on SARMV3D.

What is the implication of the main finding?

- Combining physics-based simulation with residual learning improves image fidelity without sacrificing the physical interpretability of the simulation process.
- When measured data are limited, the framework has the potential to provide controllable augmented multi-view SAR data for target recognition and related downstream tasks.

Abstract

Synthetic aperture radar (SAR) imagery plays an important role in disaster monitoring, environmental monitoring, and related applications. However, the acquisition of measured SAR data is limited by high costs and the difficulty of covering diverse environmental conditions, and the volume of SAR imagery is far smaller than that of optical remote sensing imagery, which restricts learning-based SAR applications. This paper proposes a multi-view SAR image simulation method based on OptiX and residual learning. The proposed method uses the NVIDIA OptiX ray-tracing API to simulate electromagnetic wave propagation, uses Monte Carlo sampling to model material interactions and coherent summation to form SAR images, and generates multi-view SAR image sequences of buildings. The simulator uses a scalar multiple-bounce approximation, and simulation, refinement, and evaluation use only the HH channel. The OptiX-based simulator is coupled with a residual U-Net that predicts the residual between simulated and measured images and compensates for missing fine-scale details. The method supports efficient multi-view generation. Quantitative validation on the SARMV3D dataset shows that the refined simulated images achieve low pixel-level error and high structural and perceptual similarity, yielding average values of 23.88 dB for the peak signal-to-noise ratio (PSNR), 0.83 for the structural similarity index measure (SSIM), and 0.20 for learned perceptual image patch similarity (LPIPS). The

Check for updates

Academic Editor: Yixiang Luomei

Received: 13 August 2026
Revised: 10 October 2026
Accepted: 14 October 2026
Published: 19 October 2026

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

method provides an effective means for generating multi-view SAR images of buildings, has potential for use in downstream tasks such as target recognition, is physically interpretable, and shows promising generalization capability across the evaluated building scenes.

Keywords: SAR image simulation; ray tracing; residual learning; NVIDIA OptiX; multi-view generation; image refinement

1. Introduction

Synthetic aperture radar (SAR) is an active remote sensing technology widely used in geological exploration, disaster monitoring, and environmental monitoring. Unlike optical sensors, SAR transmits microwave signals and records the returned echoes. This active imaging mode supports day-and-night operation and remains effective under many adverse weather conditions. SAR images also contain scattering information that cannot be directly obtained from optical imagery [1,2]. However, because SAR measurements are affected by a wide range of system and acquisition parameters, it is difficult to obtain representative SAR data.

SAR systems may use different flight trajectories, frequency bands, and sensor configurations. Even for the same scene, these settings affect image resolution and the recorded material responses. Target signatures also vary with the observation angle, while factors such as vegetation and soil moisture introduce additional environmental differences. Consequently, images of the same target may differ substantially across acquisition conditions.

In recent years, intelligent methods such as deep learning have significantly improved the utility of remote sensing images, and data volume has become a crucial factor in remote sensing applications. Because the availability of SAR imagery is constrained by acquisition time, viewing geometry, satellite coverage, and related factors, SAR datasets are substantially smaller than natural-image or optical remote sensing datasets, which restricts the application of intelligent methods. The MSTAR dataset [3] is one of the most widely studied datasets in the SAR community. Although it contains relatively complete actual measurement angle data of several targets, it still cannot cover the full space of a large number of targets or complex parameters and environmental changes. The actual measured SAR data is still scarce. Therefore, remote-sensing and monitoring applications require multi-view SAR image simulation.

The existing SAR simulation methods can be divided into three categories: physics-based SAR image simulation, deep learning-based SAR simulation, and a mixed method combining the two.

Physics-based SAR image simulation generates SAR echoes or images by explicitly modeling electromagnetic wave propagation, scattering mechanisms, and radar parameters. Because there is a direct connection between the simulated response and these physical processes, the simulation results have good physical interpretability.

Electromagnetic models are usually used to describe the interaction between the target and the background. For example, the Bidirectional Analytic Ray Tracing (BART) method has been used to calculate the radar cross section (RCS) of electrically large three-dimensional targets above rough surfaces [4]. BART can efficiently calculate the complex electromagnetic scattering of large three-dimensional ship targets over rough sea surfaces, and supports numerical RCS calculations under both monostatic and bistatic configurations. The improved Generalized Gaussian Coherent Scatterer (GGCS) model achieves a more general and realistic representation of SAR images by constraining the number of scatterers and flexibly selecting the relevant parameters of the Gaussian scattering distribution [5].

In addition, the Facet-Based Asymptotical Model (FBAM), combined with the geometrical optics/physical optics (GO/PO) method, has also been used to calculate the electromagnetic scattering of complex ship targets over rough sea surfaces [6], and its calculation accuracy has been verified by measured data.

Image-domain simulation generates SAR images from physically modeled scattering responses. The Mapping and Projection Algorithm (MPA) for polarimetric SAR imaging has been used for SAR image simulation of complex terrain scenes [7]. This method derives the relevant expressions for polarimetric scattering SAR imaging, and successfully realizes SAR imaging simulation under various configuration conditions. In addition, the real-time SAR simulation system based on graphics processing units (GPUs) uses rasterization technology to efficiently simulate single- and double-bounce returns, thus significantly improving the computational speed of SAR image simulation [8].

Although physics-based simulation methods have made considerable progress, they still face challenges in dealing with complex multiple-bounce scattering, large-scale scene diversity, and computational efficiency. Therefore, developing more efficient physics-based modeling methods is still an important research direction for improving the realism and diversity of SAR image simulation. Recent studies have further extended these methods. For example, large-scale SAR image simulation using EMPRISE software [9] has been validated against measured airborne SAR data and demonstrates high-fidelity simulation over complex terrain, although at high computational cost. FEKO-based SAR image simulation of marine geographical features [10] achieves accurate simulation of dynamic sea-surface scattering and increases the diversity of simulated data.

For SAR image generation, generative adversarial networks (GANs) have been widely used to synthesize SAR images at novel azimuth angles. GANs can generate high-quality images from measured SAR imagery through adversarial learning, which is particularly useful when only a limited number of samples are available. Aspect-angle synthesis methods generate SAR images at specified azimuth angles through azimuth-angle interpolation [11]. Many studies have also applied GAN-based generation to SAR data augmentation [12–15]. CycleGAN has been used for angular analysis of SAR aircraft targets [16]. An azimuth-angle discriminator has also been introduced into an improved deep convolutional generative adversarial network (DCGAN) to generate SAR images at different azimuth angles [17]. PeaceGAN has been used to estimate the target aspect angle and category of SAR target images [18].

Despite their ability to improve target recognition accuracy through synthetic samples, GAN-based methods still produce images that differ from measured SAR images in fine details and physical consistency. The absence of explicit attribute parameters and scattering mechanisms also limits the physical interpretability of the generated SAR images. Improving the realism of GAN-generated images, particularly for complex targets and environments, therefore remains an important research direction. Other deep neural networks have also been used for SAR image simulation. A differential-vector-based deep feature transformation method generates realistic samples by considering labels, azimuth angles, and target features [19]. An adversarial autoencoder image generation network substantially improves recognition accuracy under limited-sample conditions [20]. An improved recurrent neural network has been used to model sequences of SAR target images ordered by azimuth angle and predict SAR images at missing azimuth angles [21]. However, these methods generate SAR images at different azimuth angles primarily from the perspective of neural image representation, without accounting for the actual scattering mechanism and imaging projection geometry of SAR systems. A sparse modeling aspect-angle synthesis method based on images available in the training data simulates nearby aspect angles using the anisotropic scattering behavior of view-dependent scattering centers of interest [22],

but it cannot generate SAR images at different elevation angles and requires substantial manual annotation. A differentiable-simulator-based method for analyzing strong scattering in SAR images [23] improves simulation realism for complex targets by visualizing the correspondence between targets and images, although its sensitivity to noise remains to be improved. By combining the scalability of deep learning with the interpretability of physical simulation, this approach provides a new path for few-shot azimuth generation.

Hybrid methods use a physical simulator to generate the main SAR response and a neural network to reduce the sim-to-real gap. Cha et al. [24] combined a point-scatterer simulator with a deep residual refiner. The refiner mapped simulated MSTAR vehicle images to the measured-data domain, and the refined images were evaluated in an ATR task. This is an early example of residual sim-to-real refinement for SAR.

In recent years, advances in differentiable rendering have opened a new direction for novel SAR image generation. Differentiable rendering places the rendering process within a gradient-based optimization loop. Errors measured in the rendered image can then be propagated back to 3D scene parameters, allowing inverse problems to be solved while retaining the underlying rendering model. Much of the early work in computer graphics used rasterization as its basis. Rasterization is computationally efficient, but operations such as visibility determination and pixel assignment are discrete and do not have straightforward derivatives. OpenDR handles this difficulty by approximating image derivatives from local color changes during backpropagation [25]. Kato et al. define approximate gradients for the rasterization step so that the renderer can be embedded in a neural network [26]. SoftRas adopts a continuous relaxation instead, assigning probabilistic contributions to triangles and blending their effects at each pixel through a weighted average [27]. In these methods, useful gradients are obtained by approximating or smoothing the discrete rasterization process. Physically based differentiable rendering takes a different approach by differentiating the light-transport process itself. This generally provides greater physical fidelity but requires more computation [28]. Neural rendering has recently become an important branch of differentiable rendering. For example, neural radiance fields (NeRF) use multilayer perceptrons and positional encoding to represent 3D scenes implicitly, and synthesize novel views through volume rendering [29]. Compared with conventional neural networks, the operator structure of differentiable rendering is more amenable to the incorporation of physical mechanisms. Researchers have therefore sought to embed physical constraints into differentiable rendering for novel SAR image generation. Inspired by SoftRas, the differentiable SAR renderer (DSR) was proposed as a weighted operator based on SAR imaging principles for network-based target rendering and reconstruction [30]. Convolutional neural networks have been used to infer 3D target information from rendered SAR images, with DSR employed for angular interpolation [31]. A method combining feature rasterization and neural rendering has achieved SAR image rendering [32], although its performance is limited by the accuracy of the reconstructed mesh and it has difficulty generating background noise. Based on NeRF, SAR-NeRF has been proposed and combined with MPA to synthesize novel-view SAR images [33]. Similarly, Radar Fields extend radiance fields to SAR [34]. However, these methods have limited view-inference capability under few-shot conditions, and their computational overhead is substantial. Their modeling of background noise and complex materials also requires further improvement. SAR-GS [35] introduces SAR differentiable Gaussian rasterization (SDGR) and combines it with an electromagnetic scattering model to enable differentiable SAR image rendering and target reconstruction. By retaining physical scattering mechanisms while using gradient optimization, SAR-GS substantially improves rendering accuracy and geometric consistency under few-shot conditions, providing a new efficient route for differentiable rendering in SAR applications.

In summary, although deep-learning-based methods have made remarkable progress in SAR image generation and can synthesize new images using limited samples through adversarial learning, they still lack physical interpretability with respect to the underlying scattering mechanism and imaging geometry. Therefore, for complex targets and variable environments, the simulation realism of such methods is still insufficient.

Although differentiable rendering combines the rendering process with deep learning and provides an inverse-optimization route, it is still limited by mesh reconstruction accuracy, background noise simulation, and few-shot view inference. Its computational efficiency and generalization performance still need to be further improved. Therefore, this study adopts a physics-based SAR image simulation strategy that accurately models electromagnetic wave propagation and scattering to ensure the physical realism and interpretability of the generated images. However, the existing physics-based method usually has high computational complexity when modeling multiple-bounce scattering and complex material interactions, which leads to low simulation efficiency; at the same time, there is still room for further improvement in the simulation of fine-scale details such as texture, speckle, and local geometric deviations. Building on residual sim-to-real refinement for SAR, this study develops a multi-view SAR simulation and refinement framework for building imagery based on OptiX ray tracing. The framework introduces Monte Carlo ray tracing to simulate surface interactions such as diffuse scattering and specular reflection; uses coherent summation to form the final image; and uses GPU parallel acceleration to improve computational efficiency, so as to achieve more realistic multi-view SAR image simulation. Among them, the physics-based component is mainly used to reproduce large-scale structures, including coarse target contours and dominant scattering features. A residual-learning-based refinement module is used to correct the remaining fine-scale errors. For each simulated image, a deep neural network estimates the residual with respect to the corresponding measured SAR image. The predicted residual is then added to the simulated image to obtain the refined result. Because the large-scale response is generated by the physical model, it remains interpretable in terms of electromagnetic scattering mechanisms. The network predominantly learns fine-scale corrections while retaining the coarse structural prior supplied by the physical model, allowing the refinement module to be trained with a limited amount of measured data rather than a large annotated dataset.

2. Materials and Methods

This section introduces the proposed SAR image simulation and refinement method based on ray tracing. First of all, the SAR imaging principles and mathematical model are introduced. The physical quantities involved in the simulation process and their interrelationships are clarified. Then, the simulator is explained from two aspects: the overall architecture and the principal operations. The principal operations include ray tracing for SAR simulation, Monte Carlo modeling of material properties, coherent summation, and GPU-based parallel implementation. After completing these physical simulation stages, the residual-learning-based refinement module is further introduced. The specific procedure used to estimate the residual between the simulated and measured images is introduced together with the module.

The whole framework is implemented using the NVIDIA OptiX SDK (version 9.1.0; NVIDIA Corporation, Santa Clara, CA, USA). Its GLTF_render module is adapted to the SAR imaging process. Multi-view images can be generated efficiently. This method focuses on the balance between physically interpretable simulation and computational efficiency. This method aims to alleviate the problem of insufficient existing SAR data and provides diversified simulation images for potential use in downstream tasks such as target recog-

nition and environmental monitoring. The framework combines physically interpretable simulation with data-driven refinement. It significantly reduces the dependence on a large amount of measured training data.

2.1. Principles of SAR Imaging

Synthetic aperture radar (SAR) is an active remote sensing imaging technology. It transmits microwave signals through antennas mounted on moving platforms such as aircraft and satellites. Echoes reflected from the target area are then received, thus forming high-resolution two-dimensional images. SAR illuminates the scene with microwave signals and records the returned echoes. Unlike passive optical sensors, SAR does not depend on sunlight. It is also less affected by clouds and adverse weather.

As the radar platform moves, echoes are collected at successive positions. These echoes are coherently combined to form a synthetic aperture, thus providing high azimuth resolution. Range resolution is governed mainly by the signal bandwidth. SAR can therefore acquire images over long distances during both day and night. The image formation process, however, is more involved because both target scattering and coherent signal processing must be considered. Figure 1 illustrates the SAR imaging geometry.

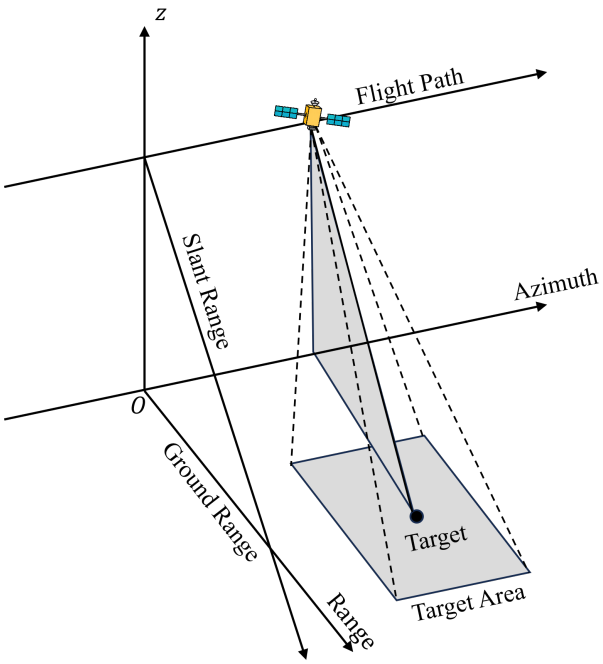

Figure 1. Schematic diagram of SAR imaging.

SAR systems usually use linear frequency modulation (LFM) signals as transmitted waveforms. By performing pulse compression on the received echo, the signal bandwidth can be exploited to obtain fine range resolution. For an LFM pulse, the transmitted signal can be expressed as:

$$s_t(t) = A \exp\left[j2\pi\left(f_c t + \frac{1}{2}K_r t^2\right)\right], \quad |t| \le \frac{T_p}{2}. \tag{1}$$

Among them, (A) represents the signal amplitude, (f_c) is the carrier frequency, and $(K_r = \frac{B}{T_p})$ is the chirp rate. (B) is the signal bandwidth, and (T_p) is the pulse duration. The time constraint $(|t| \le \frac{T_p}{2})$ indicates that the signal is only valid within a finite pulse duration. The bandwidth (B) determines the range resolution $(\delta_r = \frac{c}{2B})$. (c) is the

speed of light. The quadratic phase term makes the signal frequency vary linearly with time. This facilitates subsequent pulse compression processing.

When electromagnetic waves encounter the target, scattering occurs, including specular reflection, diffuse scattering, and multiple-bounce scattering. These scattering phenomena are related to factors such as the material properties, surface roughness, incidence angle, and wavelength. The reflected echo signal is superimposed from the contributions of all point scatterers in the scene. For the point-target model, the echo signal can be expressed as:

$$s_r(t,\eta) = \sum_{i=1}^{N} a_i \exp\left[j2\pi\left(f_c(t-\tau_i) + \frac{1}{2}K_r(t-\tau_i)^2\right)\right]. \quad (2)$$

where η denotes slow time, namely azimuth time corresponding to platform motion; a_i is the effective complex scattering coefficient of the i-th point scatterer, whose value depends on material properties, incidence angle, and wavelength; $\tau_i = \frac{2R_i(\eta)}{c}$ is the two-way delay; $R_i(\eta)$ is the instantaneous slant range from the platform to the point scatterer; N is the total number of point scatterers. The slant range $R_i(\eta)$ is usually approximated by a hyperbolic model:

$$R_i(\eta) = \sqrt{R_0^2 + (v\eta)^2}. \quad (3)$$

where R_0 is the nearest slant range and v is the platform velocity. The azimuth resolution is $\delta_a = \frac{\lambda}{(2\theta_s)}$, where $\lambda = \frac{c}{f_c}$ is the wavelength and θ_s is the synthetic aperture angle. This model accounts for Doppler frequency-shift effects: the azimuth velocity gradient induces frequency variation, which must be compensated through matched filtering.

2.2. Ray-Tracing-Based Simulation Framework

The simulation framework of this paper is based on the ray tracing method, which can model the propagation, scattering, and reflection processes of electromagnetic waves in 3D scenes. The framework adopts a high-frequency geometrical-optics (GO) ray-tracing formulation combined with Monte Carlo BRDF sampling. OptiX is used to determine the propagation paths, shadowing, and multiple reflections, while the resulting complex path contributions are coherently accumulated to form the SAR signal. With the help of OptiX hardware acceleration, this method can handle complex multiple-bounce scattering, which effectively reduces the computational cost compared with conventional analytical calculations.

The whole pipeline includes 3D-scene and SAR-parameter preparation, ray tracing, scattering computation, coherent summation, and SAR image formation, and finally outputs a multi-view SAR image sequence. Its viewing geometry and imaging parameters can be adjusted, with the azimuth angles ranging from 0° to 360° and the elevation angles ranging from 0° to 90°, so that data for augmentation can be generated.

This high-frequency ray approximation follows electromagnetic theory and provides a physical basis for the simulation. The resulting images can reproduce large-scale structures, especially coarse target contours and dominant scattering features, and provide a physically consistent structural prior for subsequent residual learning.

2.2.1. Ray-Tracing Simulation of the SAR Imaging Process

When ray tracing is applied to SAR simulation, the electromagnetic wave propagation process can be regarded as a bundle of rays emitted from a virtual radar position. After the rays interact with the scene, their amplitude and phase information is returned. The GLTF_render module of OptiX is modified to support SAR-specific ray emission and tracing. The framework can directly import standard 3D models in glTF format, including public building models and customized scenes. This avoids the need to manually define

complex geometries, makes complex targets easier to handle, and improves the usability and scalability of the framework. The triangulated glTF geometry is used to build an OptiX BVH acceleration structure. This acceleration structure is used for efficient intersection computation during ray tracing. The overall simulation framework is shown in Figure 2.

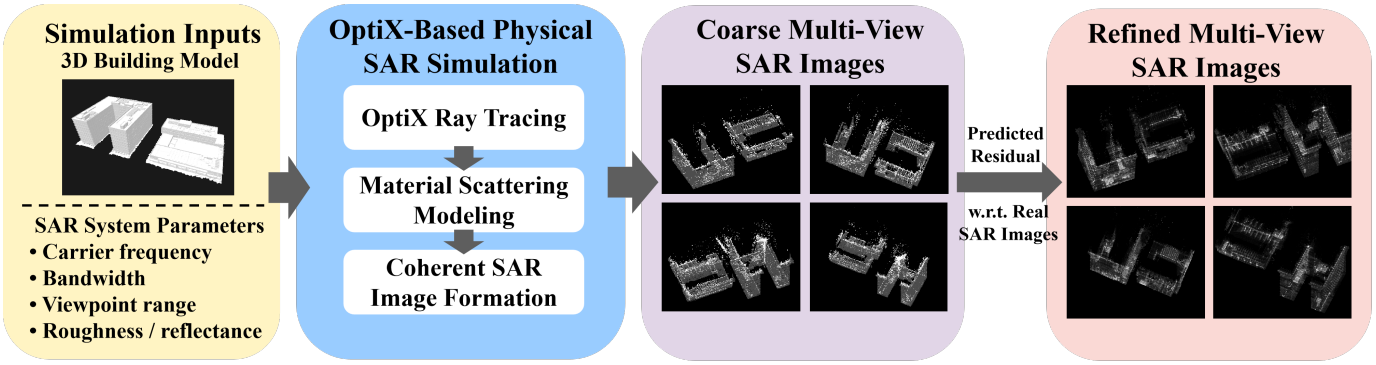

Figure 2. Overall simulation framework flowchart.

The specific process is as follows. Primary rays are emitted from the virtual radar position P_r, which is determined by SAR parameters such as platform altitude, velocity, and trajectory, toward each pixel on the imaging plane. The ray directions are computed based on projection geometry to ensure coverage of the target area. During ray propagation, the intersection point with the scene surface, P_i, the path length d, and the phase

$$\phi = -2\pi f_c \frac{d}{c} + \phi_{\text{scatter}}$$

are computed. Among them, ϕ_{scatter} represents the phase shift caused by scattering, such as random phase for rough surfaces. The amplitude is adjusted according to distance attenuation and material properties:

$$\sigma \propto \frac{1}{d^2} \times \rho_{material},$$

where $\rho_{material}$ represents the reflectance, and this term accounts for the range factor in the radar equation. Scene parameters are adjustable. For example, different environmental conditions can be represented by varying the surface roughness.

Multiple-bounce scattering is modeled in the following way: whenever a ray intersects a surface, a secondary ray is launched. The secondary ray then follows the selected reflection or scattering direction. Path continuation is controlled by a throughput-dependent stochastic termination rule beginning at path depth 2, and tracing is capped at a maximum path depth of 6. The parameter convergence analysis in Section Sampling and Path-Depth Analysis evaluates this configuration. The contributions of all retained paths are then accumulated in the complex echo of the corresponding pixel:

$$s(x,y) = \sum_p C_v g(R_p) \left[\prod_{m=1}^{M_p} \gamma_{HH,p,m} \right] \exp(-jkL_p). \quad (4)$$

where C_v denotes the system- and antenna-dependent scale factors that remain fixed for a given viewing direction; R_p is the one-way-equivalent slant range; $g(R_p)$ is the two-way propagation factor, which satisfies $g(R_p) \propto R_p^{-2}$ under the monostatic far-field approximation; L_p is the complete propagation-path length; $k = 2\pi/\lambda$; and $\gamma_{HH,p,m}$ is the effective complex HH co-polarized scattering coefficient associated with the m-th surface

interaction. Its magnitude and phase depend on the material properties, incidence angle, Fresnel response, and surface roughness.

The main feature of this process is the use of NVIDIA RT Cores for parallel ray traversal. This allows millions of ray paths to be processed efficiently. It also enables the framework to model multiple-bounce scattering, which is difficult to represent using rasterization. As a result, the framework can capture important physical effects in SAR imaging, including shadowing and multipath propagation. Phase continuity among the ray contributions is also maintained. These ray contributions are accumulated to reproduce the amplitude distributions observed in measured data. The resulting signals provide reliable large-scale structural information for the refinement module.

2.2.2. Material Property Simulation

Material simulation is a key module to improve the realism of SAR images. The framework uses Monte Carlo ray tracing to model material-dependent scattering, including diffuse scattering and specular reflection. The method is inspired by differentiable ray-tracing approaches for learning SAR surface-scattering parameters [36]. It samples the interaction between each ray and the surface. This allows rough surfaces and different material types to be represented more accurately.

Monte Carlo integration is directly related to the physical scattering model. It does not require inverse optimization and therefore avoids the additional computational cost. The method can represent different material responses in SAR images. For example, metals usually produce strong specular reflection, while vegetation mainly shows diffuse scattering. Since the method is stochastic, a sufficient number of samples is required to reduce sampling noise.

The material model is based on the bidirectional reflectance distribution function (BRDF). For diffuse scattering, a Lambertian model is used. The reflection directions are sampled over the incident hemisphere, and the Lambertian BRDF is

$$f_r = \frac{\alpha}{\pi}, \tag{5}$$

where α is the material-dependent diffuse-scattering coefficient. For specular reflection, the reflection direction is calculated according to the law of reflection:

$$\omega_o = \omega_i - 2(\omega_i \cdot n)n, \tag{6}$$

where n is the surface normal. The reflection coefficient is further adjusted by the Fresnel term.

The Fresnel reflection coefficient depends on the complex permittivity of the material and the incidence angle. It also varies with frequency in microwave bands. Therefore, the same material can produce different reflection responses at different radar bands.

Monte Carlo integration is used to estimate the average reflection. A total of N random directions are sampled from the incident hemisphere. For each sample k, the BRDF value $f_{r,k}$ is weighted by $\cos\theta_k / p(\omega_k)$. Here, $p(\omega_k)$ is the probability density of the sampled direction ω_k. For uniform hemisphere sampling, $p(\omega_k) = 1/(2\pi)$. For cosine-weighted importance sampling, $p(\omega_k) = \cos\theta_k / \pi$. The sampled values are then accumulated and averaged:

$$\sigma = \frac{1}{N} \sum_{k=1}^{N} f_{r,k} \frac{\cos\theta_k}{p(\omega_k)}. \tag{7}$$

The sampling process is implemented in parallel using the random number generator in OptiX. The implementation reads material parameters from extended attributes

in the glTF models or from external configuration files. For the 15.2-GHz configuration, the frequency-dependent complex permittivity and conductivity are parameterized according to Recommendation ITU-R P.2040-4. The nominal RMS surface height is set to 3 mm, and the same material-parameter configuration is used for all observation views. Monte Carlo sampling is integrated directly into the ray-tracing pipeline. The sampling and ray-tracing calculations therefore use the same workflow, which keeps the additional computational cost low.

The BRDF characterizes the bidirectional surface-scattering response, and Monte Carlo sampling evaluates its directional integral. The $1/R^2$ term describes two-way geometrical spreading, and the BRDF, Fresnel, and roughness terms characterize the local surface response in the forward model.

2.2.3. SAR Image Generation by Coherent Summation

Ray tracing yields complex scattering contributions for each resolution cell. Contributions assigned to the same range–azimuth coordinate are added coherently to form the corresponding complex scattering response. These responses provide the amplitude and phase terms of the SAR echo model. An imaging algorithm then converts the echo data into the final SAR image. Because each contribution remains complex-valued, its phase is retained during summation. Together, these operations follow the SAR signal-processing chain.

The coherent summation process sums all ray contributions for the range-azimuth coordinates (r, a):

$$E(r, a) = \sum_{k=1}^{M} b_k \exp(j\phi_k). \tag{8}$$

where M is the number of contributions to the resolution cell, b_k is the field-amplitude weight, and ϕ_k is the corresponding phase. This superposition preserves phase information and generates interference and speckle patterns through coherent accumulation.

In the final inverse-transform stage of the Range-Doppler algorithm, $S(u, v)$ denotes the processed two-dimensional spectrum, where (u, v) are frequency-domain coordinates. The final complex image is generated using a two-dimensional inverse Fourier transform:

$$I(x, y) = \mathcal{F}^{-1}\{S(u, v)\}. \tag{9}$$

where $\mathcal{F}^{-1}$ denotes the inverse Fourier transform and (x, y) are spatial-domain coordinates. The magnitude of $I(x, y)$ is used for image display and normalization. The process simulates the standard Range-Doppler algorithm, including range cell migration correction (RCMC) and azimuth compression, to compensate for motion-induced phase shifts. Gaussian noise can be added during processing to simulate receiver thermal noise and ensure realistic image signal-to-noise ratios. The simulated image generated by this module serves as the input to residual learning and provides a physically consistent coarse representation.

2.2.4. GPU Parallel Acceleration and System Implementation

The whole framework is developed in C++ with the NVIDIA OptiX SDK, and all core modules, including ray emission, intersection computation, Monte Carlo sampling, and coherent summation, are executed on the GPU in a highly parallel manner. The bounding volume hierarchy (BVH) acceleration structure provided by OptiX significantly optimizes the intersection computation between rays and complex 3D scenes, which are represented as triangular facets converted from glTF models. NVIDIA RT Cores accelerate BVH traversal and ray–triangle intersections during multiple-bounce tracing for a 512×512 ray-launch grid. With 10 Monte Carlo samples per pixel, each view contains 2,621,440 primary rays.

With GPU parallel acceleration, this implementation supports large-scene simulation and multi-view sequence generation. The input parameters include the glTF model, SAR configuration such as carrier frequency and bandwidth, viewing-angle range, and scene parameters such as surface roughness and reflectance. The output is a SAR image sequence with adjustable resolution, such as 512×512 pixels.

The key feature of this implementation is the adoption of a fully GPU-based end-to-end pipeline. It improves computational efficiency in large-scale scenes, supports efficient multi-view simulation, and can be readily extended to distributed computing. In the performance benchmark conducted on an RTX 3080 Ti, the complete pipeline took an average of 5.830 s/view over the eight views. All code modules are unit tested to ensure numerical stability. In addition, the framework can also be seamlessly integrated with deep learning modules, such as PyTorch-based residual prediction, enabling end-to-end simulation and refinement.

2.3. Image Refinement Based on Residual Learning

In order to further improve the realism of images generated by physics-based simulation, the framework introduces a residual-learning-based refinement module. This module mainly focuses on fine-scale details, such as speckle texture, local geometric deviation, and subtle variations in scattering intensity. Its core idea is based on the natural complementarity between physics-based simulation and data-driven learning methods. The OptiX-based ray-tracing simulation framework has already been able to capture large-scale structural information in SAR imaging, including the overall target outline, dominant specular-reflection and diffuse-scattering paths, locations of multiple-bounce scattering paths, and angle-dependent scattering characteristics. From the perspective of electromagnetic theory, these components have strong physical interpretability and physical consistency.

However, relying only on physics-based simulation, there are still clear discrepancies in fine-scale image content, such as the typical speckle distribution in measured SAR images, local intensity fluctuations caused by microscopic material texture, and small geometric projection errors. These discrepancies mainly arise from factors such as ray sampling density, Monte Carlo noise, and simplified assumptions in surface roughness modeling.

Residual learning provides an effective way to address this issue. Rather than asking the network to generate a complete SAR image from scratch, this module focuses the network on predicting the residual between the simulated image I_{sim} and the registered measured reference image $I_{\text{meas}}^{\text{reg}}$, where $I_{\text{meas}}^{\text{reg}}$ denotes the measured reference image mapped into the simulation coordinate system. The residual map R is defined as

$$R = I_{\text{meas}}^{\text{reg}} - I_{\text{sim}}. \tag{10}$$

The final refined image is then obtained by simple addition:

$$I_{\text{refined}} = I_{\text{sim}} + \hat{R}. \tag{11}$$

This design makes full use of the large-scale structural prior in the simulated image and guides the network toward predominantly local texture and scattering-intensity corrections, thereby reducing data requirements and training complexity.

During data preparation, each simulated–measured image pair is independently normalized and placed on a common pixel grid. Strong-scattering candidates are extracted from the smoothed intensity images using intensity thresholding and grid-based non-maximum suppression. Point-to-point ICP [37] is then used to estimate a two-dimensional rigid transformation consisting of rotation and translation. The measured reference image

is mapped into the simulation coordinate system before the residual target is calculated. Mild affine and noise perturbations are used during training to improve robustness to residual registration offsets.

The network architecture adopts a U-Net with residual blocks [38]. The encoder-decoder framework of U-Net extracts multi-scale spatial features and restores them during decoding. Residual blocks from ResNet [39] facilitate fine-scale residual learning and alleviate gradient vanishing. These components together improve pixel-level image refinement. In U-Net, symmetric skip connections integrate the multi-scale features extracted by the encoder with the spatial details recovered by the decoder, preserving local textures and edge information while maintaining the overall target structure. The residual blocks are further conducive to the learning of fine-scale residuals. They can mitigate gradient vanishing in deep networks and guide the network to pay more attention to the detailed differences between simulated and measured images. As a fully convolutional network, the architecture can process input images of arbitrary size. Because it does not rely on adversarial training, the training process is also more stable, making it particularly suitable for regression tasks under limited-data conditions. To further enhance the network's attention to local fine-scale details, convolutional block attention modules (CBAMs) are incorporated at key locations [40] to improve refinement performance.

The loss function also considers pixel-level accuracy, structural similarity, and deep-feature similarity. Mean squared error (MSE) is used here as the pixel-level loss to accurately match residual values. The structural loss is defined as $(1 - \mathrm{SSIM})$, where SSIM measures local structural similarity between the predicted and reference images. The perceptual loss computes mean squared error on the feature layers of a pretrained deep network to measure high-level visual similarity in texture and speckle.

The total loss is a weighted combination of the above three losses. Each weight is determined through hyperparameter search to balance pixel accuracy, texture smoothness, and perceptual quality. Pixel accuracy, structural consistency, and perceptual fidelity are taken into account by this combined loss, which can further improve speckle texture reproduction and overall visual quality.

Training follows a supervised learning paradigm. Registered paired measured and simulated images are used as training samples. The dataset contains eight groups of multi-view image pairs. Of these, six groups are used for training, one for validation, and one for independent testing. This division ensures effective learning under limited-data conditions and provides a certain degree of generalization ability. Data augmentation strategies are also introduced during training to improve robustness to noise, slight registration errors, and angular changes. Specifically, they include random mild affine transformations, such as translation, rotation, and scaling, as well as Gaussian noise injection.

During training, the network uses the Adam optimizer, and the initial learning rate is set to (1×10^{-4}). Through a cosine annealing schedule, the learning rate is gradually reduced to its minimum value. The batch size is set to 4, and the maximum number of training epochs is 200. To prevent overfitting, an early stopping strategy is introduced: training is terminated when the validation loss does not decrease for 15 consecutive epochs. Data augmentation includes random rotations ($\pm 5^{\circ}$), translations (± 10 pixels), scaling (0.95–1.05), and Gaussian noise injection with a standard deviation of 0.01. After image normalization, all augmentations are randomly applied with a probability of 0.5. The loss weights are selected through hyperparameter search based on the combined validation performance in PSNR, SSIM, and LPIPS. The selected settings are kept fixed across folds. Each fold uses validation loss for early stopping and checkpoint selection.

During inference, only the newly generated simulated image is fed into the network to obtain a predicted residual map. The predicted residual map is then directly added to

the simulated image to obtain the refined result. For quantitative evaluation, the measured image corresponding to an independent test view is used only to estimate the registration transformation so that the image pair is placed in a common coordinate system, and to calculate the evaluation metrics. The measured image is not supplied to the refinement network or used to update the network parameters.

While maintaining the large-scale structural consistency of physics-based simulation, the refinement module compensates for missing fine-scale details. It has the potential to provide higher-quality input data for downstream SAR image processing tasks such as target detection, semantic segmentation, and 3D reconstruction. Therefore, this method has strong theoretical significance and engineering value. The combined training loss used by the refinement network is defined as

$$\begin{aligned} L_{\text{total}} = \lambda_{\text{mse}} L_{\text{mse}}(R, \hat{R}) + \lambda_{\text{ssim}} \left(1 - \text{SSIM}(I_{\text{meas}}^{\text{reg}}, I_{\text{sim}} + \hat{R})\right) \\ + \lambda_{\text{perc}} L_{\text{perceptual}}(I_{\text{meas}}^{\text{reg}}, I_{\text{sim}} + \hat{R}). \end{aligned} \tag{12}$$

The overall residual-learning refinement procedure is summarized in Algorithm 1.

Algorithm 1. Residual-learning-based SAR image refinement

Input: Paired simulated and measured SAR images $\{(I_{\text{sim}}^i, I_{\text{meas}}^i)\}_{i=1}^N$, validation image pairs, and a new simulated image $I_{\text{sim,new}}$.
Output: Refined SAR image I_{refined}.
Parameters: U-Net with residual blocks and CBAM modules; Adam optimizer; early-stopping strategy.

Normalize each paired I_{sim}^i and I_{meas}^i.
Map I_{meas}^i into the simulation coordinate system using rigid ICP to obtain $I_{\text{meas}}^{i,\text{reg}}$.
Compute the residual target $R^i = I_{\text{meas}}^{i,\text{reg}} - I_{\text{sim}}^i$.
Initialize the refinement network.
for each training epoch **do**
  for each mini-batch **do**
    Apply random affine augmentation and Gaussian noise injection.
    Predict $\hat{R}$ using the refinement network.
    Compute L_{total} using Equation (12).
    Update the network parameters with Adam by backpropagation.
  end for
  Evaluate the validation loss.
  if the validation loss does not decrease for 15 consecutive epochs **then**
    Stop training.
  end if
end for
Normalize the new simulated image $I_{\text{sim,new}}$.
Predict $\hat{R}_{\text{new}}$ using the trained refinement network.
Generate $I_{\text{refined}} = I_{\text{sim,new}} + \hat{R}_{\text{new}}$.
return I_{refined}.

2.4. Experimental Setup

The publicly available SARMV3D series of SAR microwave vision 3D imaging datasets [41–43] is used as the primary benchmark. Released by the Aerospace Information Research Institute, Chinese Academy of Sciences, the dataset focuses on SAR microwave-vision 3D imaging and supports the fusion of multi-view array interferometric SAR (InSAR) data, laser point clouds, and optical data. It is suitable for 3D reconstruction and simulation validation of typical targets such as buildings. The distinctive advantage of the SARMV3D series is that it combines measured data acquired by a microwave-vision 3D SAR (MV3DSAR) system and provides high-resolution 3D point-cloud ground truth together with multi-directional SAR images. This makes it particularly suitable for validating the accuracy of physics-based simulation frameworks. In this experiment, we mainly use the first subset of SARMV3D-3.0, namely the small unmanned aerial vehicle (UAV)-borne full-polarimetric array InSAR data subset. This subset was collected in Suzhou in October

2022 using the MV3DSAR system. It contains data acquired from eight different directions around two adjacent typical buildings, along with contemporaneously collected laser point cloud data and oblique optical photogrammetry data. The data cover multiple incidence angles and polarization modes, including HH, HV, VH, and VV, with a resolution better than 0.2 m × 0.2 m. This subset is especially suitable for multi-view SAR simulation because it provides measured multidirectional observations, accurate 3D point-cloud ground truth, and corresponding SAR images. It can therefore be used for direct comparison and validation of the simulated images generated by the proposed framework. SAR images and 3D models covering eight directions are selected from this subset as the benchmark for qualitative and quantitative evaluation. The dataset is chosen because of its public availability and completeness. Compared with other SAR datasets, SARMV3D places greater emphasis on multi-view coverage of building scenes, which closely matches the focus of this study. In addition, the laser point cloud data in the dataset can be directly imported into the proposed OptiX-based simulation framework after conversion to glTF format, simplifying experimental preparation and providing high-quality paired measured and simulated images for subsequent residual learning.

To evaluate cross-scene generalization, we conducted an additional experiment on SARMV3D-2.0 using a LiDAR-derived building model and eight corresponding measured SAR views. The refinement network trained on SARMV3D-3.0 was directly applied to the simulated images of SARMV3D-2.0. The network weights were kept fixed, without additional training or fine-tuning on the target scene.

Although the dataset provides all four polarization channels, all simulations, network training, and quantitative evaluations reported in this study use the HH-polarized R1 channel. The simulated image and its measured reference are both represented as single-channel HH amplitude images, so the physical simulator and the refinement network operate on the same co-polarized channel.

Published physics-based SAR simulation results are generally obtained using different scenes, operating frequencies, polarization channels, spatial resolutions, and evaluation procedures. To provide comparisons under consistent experimental conditions, controlled internal baselines are evaluated on the same SARMV3D-3.0 data. The physical-simulator baselines examine single-bounce propagation, constant-reflectivity material interaction, and phase-free accumulation. The refinement-network baselines examine conventional U-Net prediction, residual prediction, residual blocks, skip connections, CBAM, and different combinations of the MSE, SSIM, and perceptual losses.

The experiments are conducted on a server equipped with an NVIDIA GeForce RTX 3080 Ti GPU with 12 GB of memory, an Intel Core i5-13600KF CPU (Intel Corporation, Santa Clara, CA, USA), and 32 GB of RAM. This configuration makes full use of the hardware acceleration capabilities of the RTX series and supports parallel ray tracing with the NVIDIA OptiX SDK. It provides sufficient computational resources for the subsequent PyTorch (version 2.5.1) integration of the residual learning module. The SAR parameter configuration follows the actual acquisition parameters of the dataset and is adjusted accordingly to meet the simulation requirements. The carrier frequency used in the present experiments is 15.2 GHz, corresponding to a wavelength of approximately 0.01972 m. The bandwidth is selected to achieve a range resolution better than 0.2 m.

The platform is simulated as a UAV at an altitude of approximately 300–500 m. Its trajectory is adjusted according to the eight viewing directions in the dataset to simulate circular flight. The simulation resolution is set to 512 × 512 pixels. The ray-launch grid contains 10 independent Monte Carlo samples per pixel, and the maximum path depth is set to 6. These settings are selected according to the sampling-convergence and path-depth analysis reported in Table 1. The same physical-simulation configuration is used to

generate all inputs for the residual-learning experiments, providing a consistent sampling distribution across training, validation, and testing.

Table 1. Sampling-number and maximum-path-depth analyses. Sampling statistics are reported as mean ± standard deviation over eight observation views; runtime was measured for a representative view on an NVIDIA GeForce RTX 3080 Ti.

(a) Monte Carlo sampling analysis at maximum path depth 6

Samples/Pixel	Runtime (s)	ENL	Rayleigh KS
1	0.733	0.3771 ± 0.0312	0.4037 ± 0.0350
3	0.840	0.5707 ± 0.0379	0.2586 ± 0.0249
7	1.197	0.7062 ± 0.0643	0.2402 ± 0.0208
10	1.550	0.7070 ± 0.0745	0.2275 ± 0.0788

(b) Maximum-path-depth analysis at 10 samples per pixel

Maximum Depth	Runtime (s)	Returned-Energy Fraction	NCC vs. Depth 6
2	1.519	0.8733	0.9505
4	1.573	0.9943	0.9980
6 (reference)	1.532	1.0000	1.0000

The simulation covers the eight viewing directions in the dataset and generates the corresponding multi-view SAR image sequences. The refinement experiments use an eight-fold cross-view evaluation protocol. Six views are used for training in each fold, one view for validation, and one view serves as the independent test view. The test role is rotated across all eight views. Each fold is repeated using three independent random seeds, resulting in 24 held-out test evaluations for each network configuration.

Under limited-data conditions, this design can ensure the effectiveness and generalization capability of the hybrid framework. The network adopts a U-Net with residual blocks as the backbone network. The encoder contains five downsampling layers. Each downsampling layer is followed by two residual blocks. The decoder adopts a symmetric upsampling structure and embeds CBAM attention modules in the skip connections to enhance local texture perception. The network takes a single-channel HH amplitude image as input and outputs the predicted residual map.

The evaluation metrics include peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), learned perceptual image patch similarity (LPIPS), and mean squared error (MSE). These metrics are used to jointly quantify the similarity improvement from physical simulation to refined images. PSNR and SSIM are computed on log-compressed SAR intensity images normalized to [0,1].

PSNR and MSE measure pixel-level agreement, SSIM measures local structural similarity, and LPIPS measures deep-feature similarity. Samples are taken from five matched, non-overlapping 12×12 low-structure regions within the target footprint and normalized by their regional RMS. The Rayleigh KS distance measures their deviation from a unit-RMS Rayleigh distribution. Normalized cross-correlation (NCC) compares Gaussian-smoothed image maps ($\sigma = 3$ pixels) within the same footprint to assess the spatial distribution of scattering responses.

For the architecture ablations, the data partitions, data augmentation, training settings, and complete loss function are kept unchanged while residual prediction, residual blocks, skip connections, and CBAM are varied. For the loss-function ablations, the complete residual U-Net with CBAM is retained while the loss is varied among MSE only, MSE + SSIM, and MSE + SSIM + perceptual loss. The mean and standard deviation of PSNR, SSIM, LPIPS, and MSE are reported.

Sampling and Path-Depth Analysis

A controlled analysis was conducted using representative sampling numbers of 1, 3, 7, and 10 samples per pixel at a maximum path depth of 6. Sampling convergence was evaluated over the eight observation views using the equivalent number of looks (ENL) and the Kolmogorov–Smirnov distance to the Rayleigh amplitude distribution, while direct rendering of a representative view was used to measure runtime. The maximum path depth was then varied over 2, 4, and 6 at 10 samples per pixel and evaluated using the returned-energy fraction and normalized cross-correlation (NCC) with the depth-6 result.

The mean Rayleigh KS statistic decreases as the number of samples per pixel increases, with the largest reduction occurring between 1 and 3 samples per pixel. Between 7 and 10 samples per pixel, the mean ENL changes from 0.7062 to 0.7070, while the mean Rayleigh KS statistic decreases from 0.2402 to 0.2275. At 10 samples per pixel, the representative runtime is 1.550 s. Increasing the maximum path depth from 2 to 4 raises the returned-energy fraction from 0.8733 to 0.9943 and the NCC from 0.9505 to 0.9980, while depth 6 retains the complete evaluated path contribution with comparable runtime. Accordingly, 10 samples per pixel and a maximum path depth of 6 are used throughout the experiments.

3. Results

All experiments use the setup described in Section 2.4. First, the OptiX-based simulation results are compared with measured SAR images from eight viewing directions. We then evaluate the residual-learning refinement using visual comparisons and objective image-quality metrics. Section 4 discusses the implications of these results and the limitations of the framework.

3.1. Results of OptiX-Based Physical Simulation

Before evaluating the complex building scene, the ray-propagation and coherent-integration stages were validated at 15.2 GHz using a 1 m $\times$ 1 m PEC square plate and a 1 m $\times$ 1 m PEC right dihedral. For the square plate, the analytical and simulated broadside RCS values were 45.0926 and 45.0931 dBsm, respectively, giving an absolute difference of 0.0006 dB and a main-lobe RMSE of 0.0013 dB. For the right dihedral, the boresight RCS values obtained by analytical calculation and simulation are 48.1029 and 48.0996 dBsm, respectively. The absolute difference between the two is 0.0032 dB, and the angular-response RMSE is 0.0037 dB. These results show that the high-frequency response obtained from the simulation is in close agreement with the analytical results, and verify the ray-intersection geometry, two-way propagation phase, coherent integration, and double-reflection path construction.

This experiment is mainly used to validate the practical performance of the physics-based simulation framework in generating multi-view SAR images. By comparison with measured data, the accuracy and limitations of the framework in simulating complex building scenes are evaluated. The experiment first converts the 3D laser point cloud from the SARMV3D-3.0 subset into a glTF model as input, and combines it with matching SAR system parameters to generate simulated image sequences corresponding to the eight viewing directions.

For each viewing direction, simulation is conducted independently to model electromagnetic interactions under the corresponding acquisition conditions, including multiple-bounce scattering and coherence effects, and to further examine the response of the framework to angular variation.

Qualitative analysis: Figure 3 compares simulated SAR images from several azimuth directions with measured SAR images. The simulated images successfully reproduce the coarse outlines of the buildings, dominant specular reflections such as bright wall

lines, multiple-bounce scattering such as corner-reflector responses, and layover effects. Strong agreement with the measured images can be observed. The framework can better handle projection deformation and scattering-characteristic differences caused by azimuthal variation, while maintaining cross-view consistency in target structure.

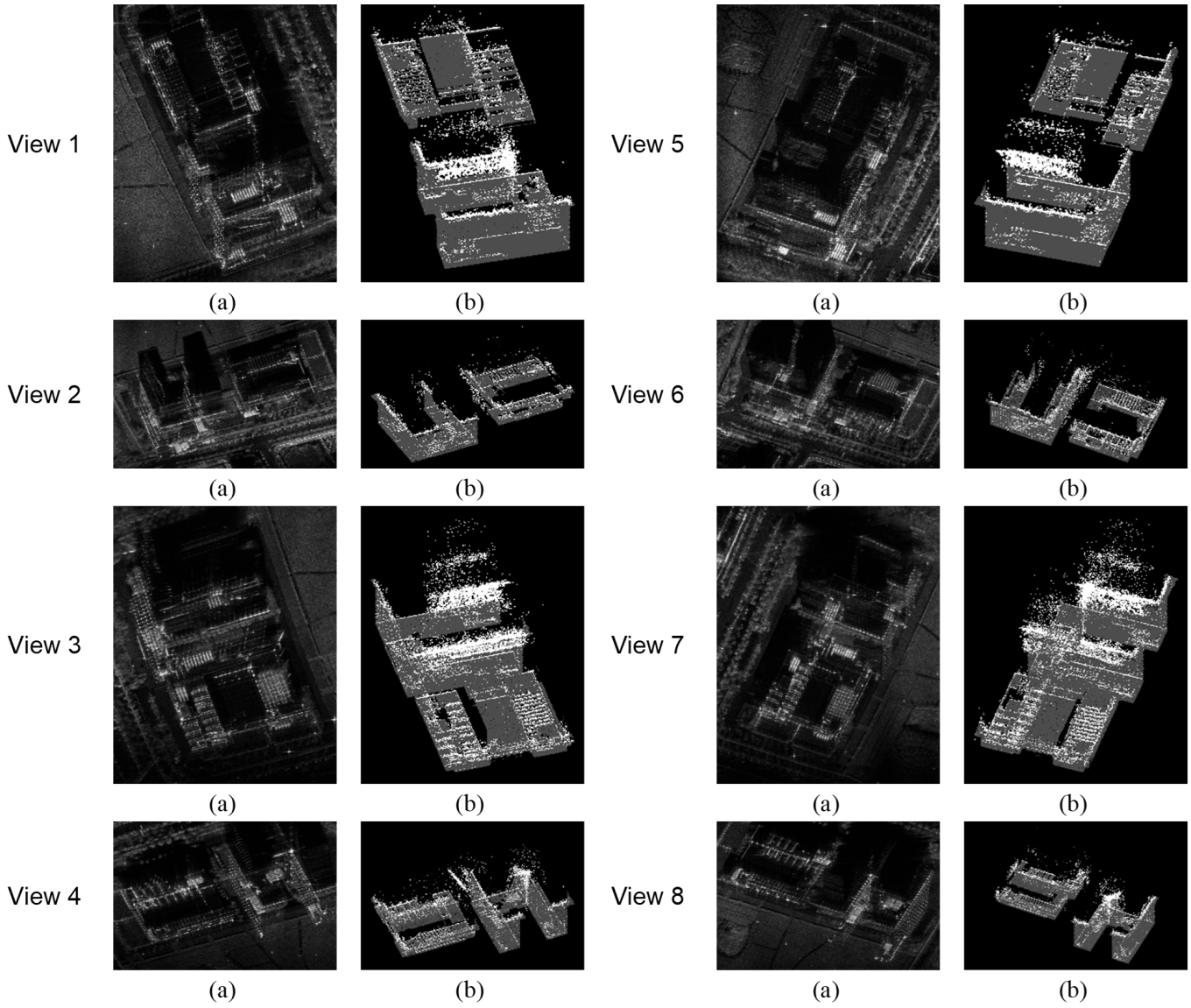

Figure 3. Multi-view SAR simulation and comparison with measured SAR data: (**a**) measured SAR image; (**b**) simulated SAR image.

The ray-tracing framework shows strong physical consistency in large-scale structures, including coarse outlines, dominant scattering paths, and angle-dependent scattering characteristics. However, some discrepancies remain in the simulated images in terms of fine-scale details due to Monte Carlo sampling noise, simplified assumptions in surface roughness modeling, and incomplete simulation of system noise. These discrepancies include the speckle distribution in measured SAR images, subtle local fluctuations in scattering intensity, and edge sharpness.

The simulation method can better reproduce large-scale structures such as dominant scattering paths, geometric shadows, and deformation caused by angular variation, and achieves high visual similarity to measured data. The remaining fine-scale discrepancies provide a basis for subsequent refinement. Owing to the end-to-end integration of ray

tracing and the avoidance of parameter mismatch commonly observed in traditional methods, the simulated images can also maintain stable physical consistency across different viewing directions.

Quantitative analysis: To quantify the simulation performance, this experiment computes the evaluation metrics for each viewing direction and summarizes the results in Table 2. Across the eight viewing directions, the average PSNR is 16.6346 dB, the SSIM is 0.7613, the LPIPS is 0.2373, and the MSE is 0.0256. The simulated images and measured data are highly consistent in structural similarity, especially in terms of edge sharpness and overall layout.

These differences mainly arise from the following factors. The conversion of the 3D model is constrained by the accuracy of the original details, so there are inherent discrepancies between the model and the actual target in terms of geometry and surface detail. The measured data also contain factors that are difficult to control, such as system noise, speckle, and polarization characteristics. In addition, the intensity distribution and local scattering behavior of the simulated images largely depend on controllable parameters such as material properties and surface roughness. Although these parameters can be flexibly adjusted, it is still difficult to fully reproduce the electromagnetic response of the actual scene.

These quantitative results confirm the reliability of the simulation method in capturing the core mechanisms of SAR imaging and can support efficient single-image generation, laying the foundation for large-scale multi-view data augmentation. Discrepancies still remain in the fine-scale content of the simulated images, such as speckle mismatch. These results therefore motivate the further introduction of the hybrid strategy.

Table 2. Quantitative metrics for multi-view physics-based simulation.

View	PSNR (dB)	SSIM	LPIPS	MSE
View 1	13.6357	0.6432	0.2989	0.0433
View 2	21.2006	0.8808	0.1639	0.0076
View 3	14.3080	0.7477	0.2799	0.0371
View 4	17.7746	0.7778	0.2299	0.0167
View 5	13.8226	0.6584	0.3058	0.0415
View 6	20.0417	0.8524	0.1479	0.0099
View 7	15.8449	0.8029	0.2231	0.0260
View 8	16.4490	0.7273	0.2491	0.0227
Average	**16.6346**	**0.7613**	**0.2373**	**0.0256**

3.2. Results of Residual-Learning-Based Image Refinement

This experiment evaluates the refinement effect of the residual learning module on the basis of the physics-based simulation results. Through before-and-after comparison, it examines the module's contribution to fine-scale detail compensation. The experiment includes data preprocessing, model training, and generalization testing, using paired measured and simulated images from eight directions as the benchmark. The focus is to verify the practical improvement of the hybrid framework under limited-data conditions.

Data preprocessing and registration: Each simulated–measured image pair is independently aligned using rigid ICP before residual construction, as illustrated in Figure 4. To evaluate the influence of residual misalignment, controlled translations of 0, 1, 2, 3, and 5 pixels were applied only to the simulated test inputs, while the measured references and trained networks were held fixed. At a 2-pixel displacement, the residual target changed by $74.4 \pm 9.5\%$ of the unperturbed residual norm, whereas the refined PSNR and SSIM changed by only -0.108 ± 0.087 dB and -0.0141 ± 0.0158, respectively. Relative to the correspondingly perturbed physics-based input, refinement retained gains of

4.715 ± 1.855 dB in PSNR and 0.0517 ± 0.0137 in SSIM. Positive PSNR and SSIM gains were maintained for all eight views throughout the tested 0–5-pixel range.

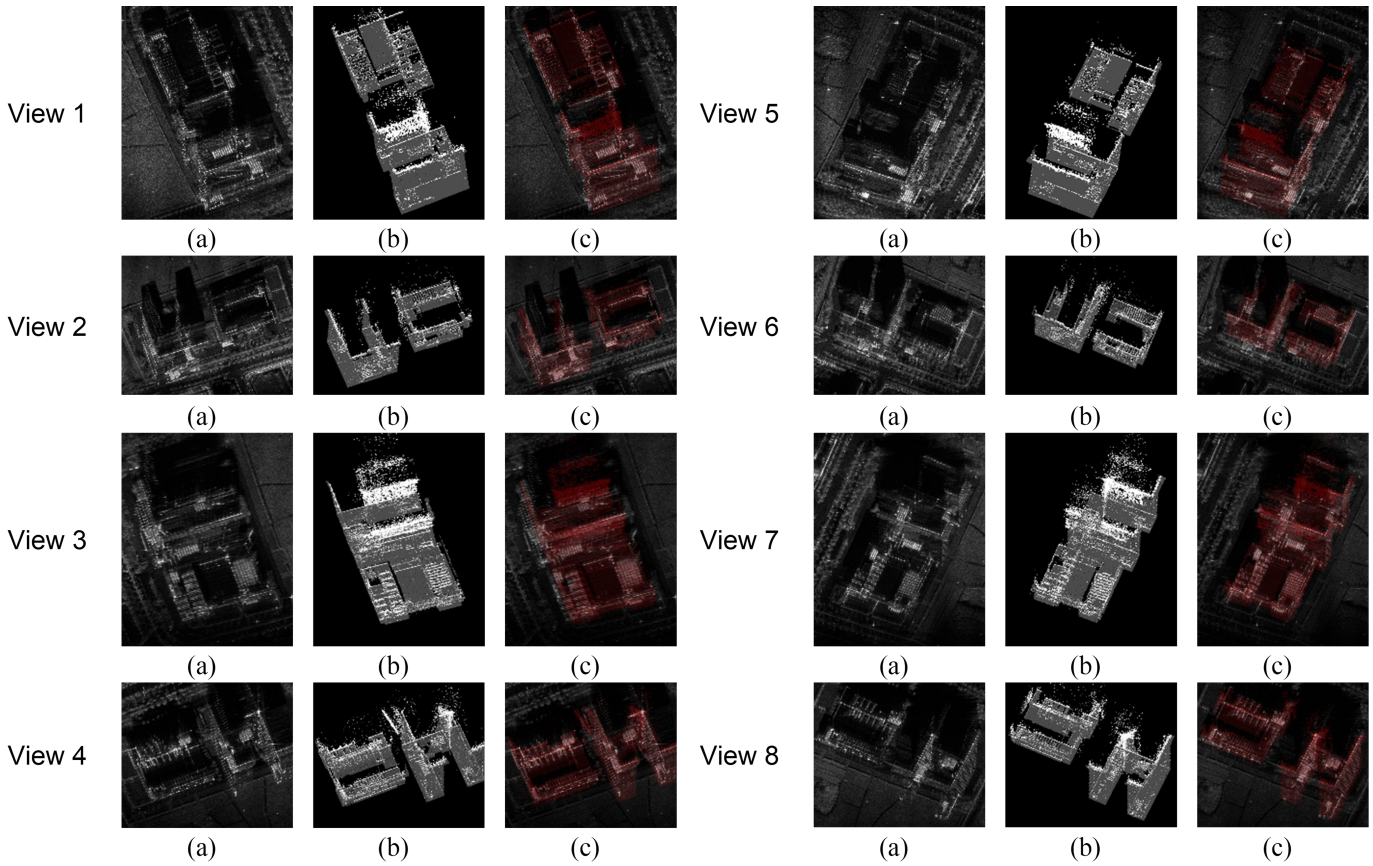

Figure 4. Image registration results: (**a**) measured SAR image; (**b**) corresponding simulated SAR image after pairwise alignment; (**c**) registration schematic.

Figure 5 compares the simulated images before and after refinement. Before refinement, the simulated images are relatively smooth overall, especially in terms of speckle texture and local details. Some microscopic textures that can be observed in the measured data are missing. After refinement, these details become clearer and more visible, while the large-scale target outline remains consistent with the original simulation.

The cross-view generalization results show a similar trend. The refined images contain fewer fine-scale artifacts while maintaining the consistency of the physics-based simulation. The shadow boundaries are more natural, and the scattering-center intensities are closer to those in the measured images. This indicates that the residual module predominantly learns fine-scale features from the measured data while largely preserving the coarse geometric organization provided by the physics-based simulation.

In complex building scenes, the refinement process improves local details and cross-view consistency. The hybrid strategy can therefore support its application to tasks such as disaster monitoring.

Table 3 summarizes the evaluation metrics after refinement across the eight viewing directions. After refinement, the average PSNR increases from 16.6346 dB to 23.8844 dB, a gain of 7.2497 dB. The pixel-level error is substantially reduced, especially in speckle-dense regions and regions with clear local intensity fluctuation. SSIM increases from 0.7613 to 0.8273, a gain of 0.0660. Structural similarity and texture-detail matching are markedly improved, and the network can effectively compensate for missing fine-scale image components. LPIPS decreases from 0.2373 to 0.2047, a reduction of 0.0326. This

indicates a substantial improvement in deep-feature similarity between the refined images and measured SAR data. MSE decreases from 0.0256 to 0.0044, a reduction of 0.0211. From a quantitative perspective, pixel-level agreement is significantly improved. These conventional metrics are complemented by the Rayleigh KS and NCC results in Table 4.

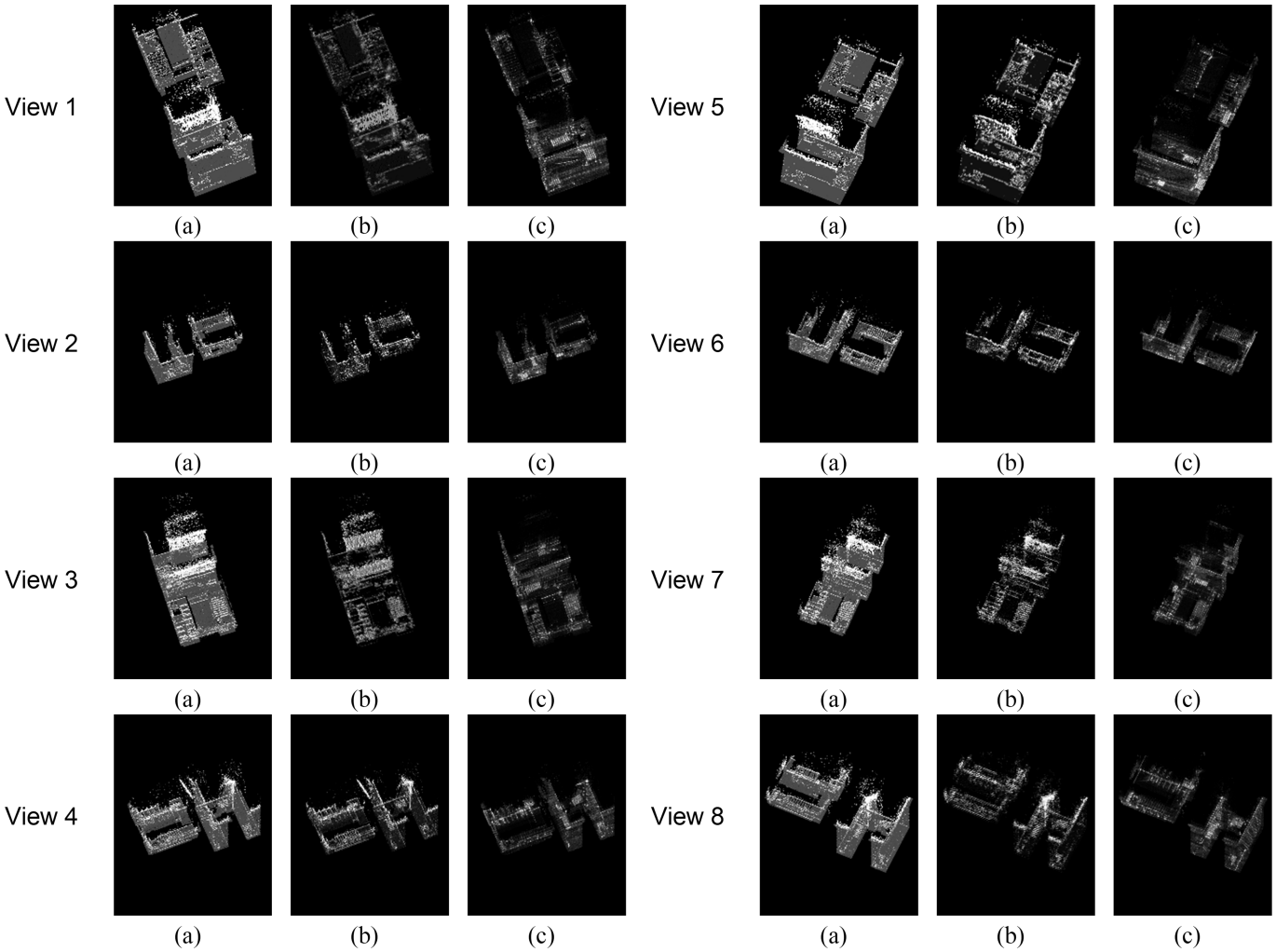

Figure 5. Comparison before and after residual-learning refinement: (**a**) image produced by physics-based simulation; (**b**) refined image; (**c**) measured SAR image.

The performance improvements are evident in the held-out test views, indicating that training under limited-data conditions still has good generalization capability. Physics-based simulation provides a consistent large-scale structural scaffold, including dominant outlines and scattering paths. The controlled comparisons quantify the contributions of the physical-simulation components and the residual-refinement network. The residual network mainly learns the fine-scale residual mapping. This avoids the risk of physical distortion and artifacts that may occur when generating images from scratch.

The eight-fold cross-view evaluation demonstrates improved image similarity on the held-out test views. These results support the effectiveness and cross-view generalization of residual learning under limited-data conditions. These results confirm the effectiveness of residual learning and show its potential in SAR scenarios with limited training data. With only a limited number of samples, this method can achieve stable convergence and good generalization, using far fewer samples than conventional deep-learning methods.

Table 3. Quantitative metrics for multi-view refinement. Each row reports the independent test result obtained when the corresponding view serves as the test view in the eight-fold cross-view evaluation; the other views comprise six training views and one validation view. The values are averaged over three training runs with different random seeds.

View	PSNR (dB)	SSIM	LPIPS	MSE
View 1	21.1189	0.7271	0.2663	0.0077
View 2	25.8113	0.9051	0.1424	0.0026
View 3	24.0479	0.8125	0.2517	0.0039
View 4	23.8760	0.8390	0.2053	0.0041
View 5	21.2989	0.7583	0.2251	0.0075
View 6	25.7541	0.8908	0.1601	0.0027
View 7	24.6990	0.8607	0.1684	0.0034
View 8	24.4687	0.8246	0.2184	0.0036
Average	**23.8844**	**0.8273**	**0.2047**	**0.0044**

Table 4. Rayleigh KS distance and normalized cross-correlation over eight observation views.

Metric	Physics-Based Simulation	Refined Simulation
Rayleigh KS	0.2275 ± 0.0788	$\mathbf{0.1668 \pm 0.1060}$
Normalized cross-correlation (NCC)	0.4910 ± 0.2630	$\mathbf{0.7318 \pm 0.0855}$

Values are the mean and standard deviation across eight observation views. Lower Rayleigh KS indicates greater consistency with the Rayleigh speckle model in the selected low-structure regions. Higher NCC indicates closer agreement with the corresponding measured SAR images in the spatial distribution of scattering responses. These results use the predictions from the earlier eight-view residual-refinement evaluation; Table 3 reports the updated CBAM configuration.

As shown in Table 4, the decrease in Rayleigh KS after refinement indicates that the statistical distributions in the selected low-structure regions are more consistent with the Rayleigh speckle model. The higher NCC indicates closer agreement with the measured SAR images in the spatial distribution of scattering responses.

Table 2 provides the physics-based simulation baseline, and Table 5 reports the architecture and loss-function ablations. Residual prediction reduces LPIPS from 0.2638 to 0.1860 and MSE from 0.0068 to 0.0065 relative to direct prediction. Residual blocks further increase PSNR to 23.7943 dB and SSIM to 0.8247 while reducing MSE to 0.0046. CBAM provides additional improvements in PSNR, SSIM, and MSE, producing values of 23.8844 dB, 0.8273, and 0.0044, respectively, while the configuration without CBAM retains a slightly lower LPIPS. Removing the skip connections causes a marked degradation across all four metrics. Among the loss configurations, the complete MSE + SSIM + perceptual loss achieves the highest PSNR and SSIM and the lowest MSE, together with a lower LPIPS than the MSE-only and MSE + SSIM configurations.

To identify the contributions of the physical-simulation components under identical conditions, controlled ablations were conducted for multiple-bounce propagation, material interaction, and coherent phase accumulation. The results are summarized in Table 6.

The single-bounce configuration retains most of the returned energy, while the remaining reflection paths produce a relative L_2 difference of 0.2555. The energy-matched material baseline produces a relative L_2 difference of 0.2180, showing that material-dependent interaction weighting changes the spatial distribution of scattering intensity. The phase-free baseline substantially changes the peak amplitude and ENL, confirming the contribution of coherent phase accumulation to interference and amplitude statistics.

Table 5. Architecture and loss-function ablation study of the refinement network. All architecture comparisons use the complete loss, and all loss-function comparisons use the complete residual U-Net with CBAM. Results are reported as mean ± standard deviation over eight cross-view folds and three random seeds per fold ($n = 24$ per configuration). Higher PSNR and SSIM values and lower LPIPS and MSE values indicate better performance.

Configuration	PSNR (dB)	SSIM	LPIPS	MSE
Conventional U-Net, direct prediction	23.1583 ± 2.9180	0.8096 ± 0.0707	0.2638 ± 0.1110	0.0068 ± 0.0094
U-Net with residual prediction	23.1337 ± 2.7903	0.8085 ± 0.0733	**0.1860 ± 0.0505**	0.0065 ± 0.0076
Residual U-Net without CBAM	23.7943 ± 1.7893	0.8247 ± 0.0612	0.1993 ± 0.0462	0.0046 ± 0.0021
Residual U-Net + CBAM, complete loss	**23.8844 ± 1.7536**	**0.8273 ± 0.0587**	0.2047 ± 0.0434	**0.0044 ± 0.0020**
Complete model without skip connections	18.6149 ± 2.6259	0.7709 ± 0.0806	0.2203 ± 0.0556	0.0162 ± 0.0093
Residual U-Net + CBAM, MSE + SSIM	21.3004 ± 3.1234	0.7903 ± 0.0791	0.2097 ± 0.0515	0.0101 ± 0.0107
Residual U-Net + CBAM, MSE only	22.3063 ± 3.2121	0.7513 ± 0.0888	0.2359 ± 0.0660	0.0084 ± 0.0102

Table 6. Component ablation of the physical simulator. Values are averaged over eight views.

Component	Reduced Configuration	Metric	Complete Simulator	Reduced Configuration
Multiple-bounce propagation	Single-bounce only	Returned-energy fraction	1.0000	0.8823
Multiple-bounce propagation	Single-bounce only	Relative L_2 difference	0.0000	0.2555
Material interaction	Constant reflectivity	NCC	1.0000	0.9704
Material interaction	Constant reflectivity	Relative L_2 difference	0.0000	0.2180
Coherent phase accumulation	Phase-free accumulation	Peak amplitude	0.5082	1.9114
Coherent phase accumulation	Phase-free accumulation	ENL	0.4764	0.4032

3.3. Cross-Scene Generalization on SARMV3D-2.0

The network trained on SARMV3D-3.0 was evaluated on all eight observation views of SARMV3D-2.0 using fixed pretrained weights. Figure 6 compares the physics-based simulations, refined simulations, and corresponding measured SAR images.

Table 7 summarizes the quantitative results of cross-scene transfer across the eight target-scene views.

Direct transfer increases the average PSNR from 16.3347 to 19.4651 dB and SSIM from 0.6213 to 0.7180, while reducing LPIPS and MSE. The lower Rayleigh KS indicates that the statistical distributions in the selected low-structure regions are more consistent with the Rayleigh speckle model. NCC increases from 0.7836 to 0.9238, indicating closer agreement with the measured SAR images in the spatial distribution of scattering responses. These

improvements demonstrate cross-scene generalization of the learned residual correction from SARMV3D-3.0 to SARMV3D-2.0.

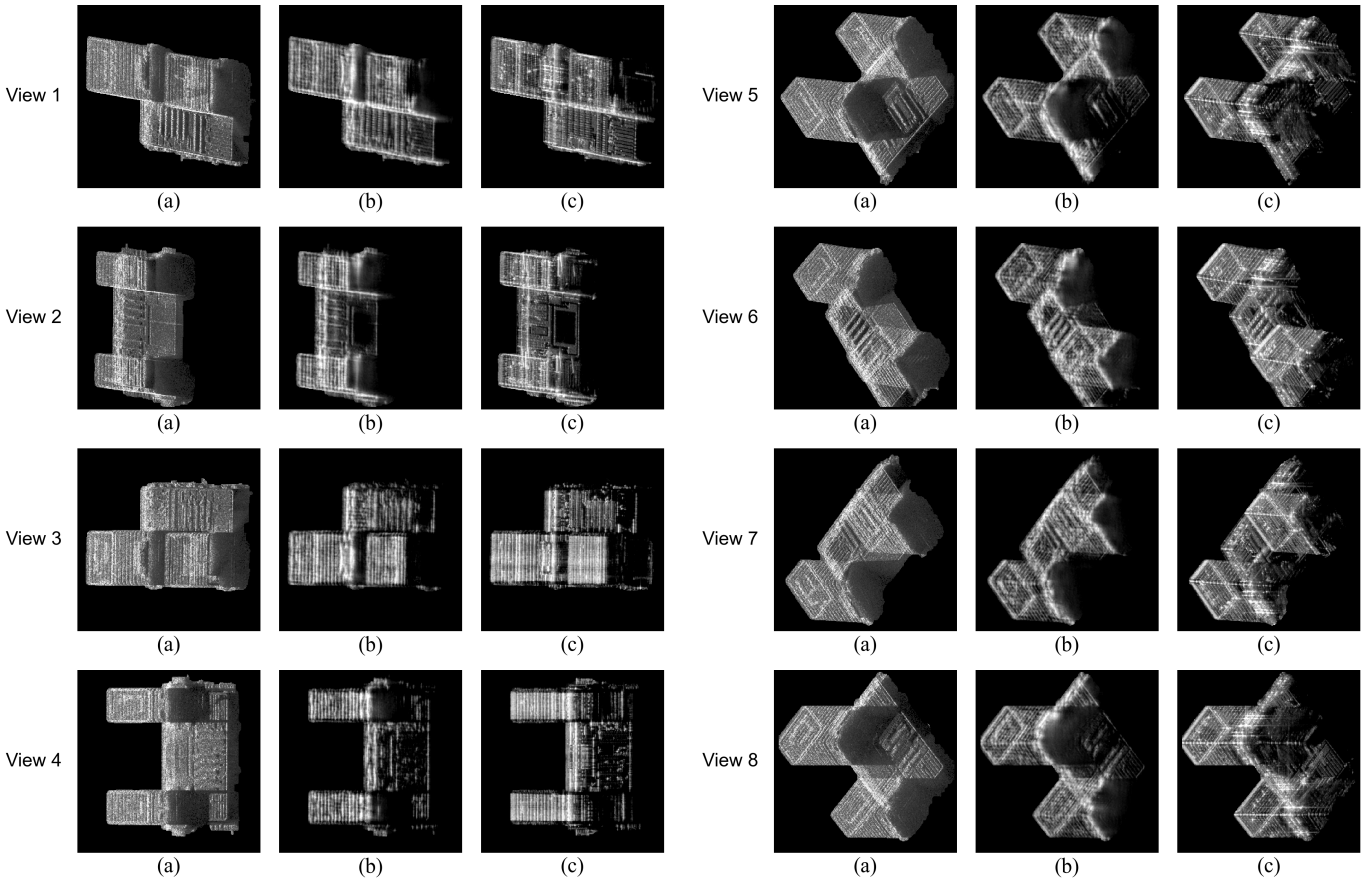

Figure 6. Comparison before and after residual-learning refinement on SARMV3D-2.0 using fixed weights trained on SARMV3D-3.0: (**a**) image produced by physics-based simulation; (**b**) refined image; (**c**) measured SAR image.

Table 7. Quantitative results of cross-scene transfer from SARMV3D-3.0 to SARMV3D-2.0.

Metric	Physics-Based Simulation	Refined Simulation
PSNR (dB)	16.3347 ± 1.1718	$\mathbf{19.4651 \pm 1.0514}$
SSIM	0.6213 ± 0.0434	$\mathbf{0.7180 \pm 0.0584}$
LPIPS	0.3112 ± 0.0344	$\mathbf{0.2281 \pm 0.0403}$
MSE	0.0240 ± 0.0069	$\mathbf{0.0116 \pm 0.0027}$
Rayleigh KS	0.2343 ± 0.0319	$\mathbf{0.2039 \pm 0.0527}$
Normalized cross-correlation (NCC)	0.7836 ± 0.0319	$\mathbf{0.9238 \pm 0.0399}$

Values are reported as mean $\pm$ sample standard deviation across eight SARMV3D-2.0 observation views. All refined images were generated using fixed weights trained on SARMV3D-3.0, without target-scene training or fine-tuning.

4. Discussion

The above experiments analyze the performance of the proposed framework, as well as its theoretical significance and practical application value in SAR image processing from multiple aspects. The physics-based simulation results show that the OptiX-based ray-tracing method performs well in multi-view building scenes, especially in capturing large-scale structures such as the overall target outline, dominant specular reflection paths, multiple-bounce scattering paths, and angle-dependent projection deformation. These simulated structures are highly consistent with the measured SAR data. The average PSNR is 16.63 dB, and the SSIM is 0.76. This performance is mainly related to the physics-based

modeling of electromagnetic wave propagation. Material interaction is modeled using Monte Carlo integration, while speckle is simulated by coherent summation. Ray tracing also has clear advantages in handling complex geometry and multiple-bounce scattering, as shown by the experimental results.

However, there are still certain limitations when relying only on physics-based simulation, as shown by the qualitative comparison results. Fine-scale details in the simulated images have not been fully reproduced, including random speckle patterns, subtle local fluctuations in scattering intensity, and edge sharpness. Compared with the measured images, the simulated images are therefore usually smoother and contain relatively less noise. For some viewing directions, wall-related multiple-bounce scattering also differs from that in the measured images in terms of intensity and spatial distribution. The simulated multiple-bounce paths are sometimes weaker, and the reflection contribution in some local regions is insufficient. In addition, some local differences near sharp building boundaries may also arise because the current path model does not explicitly include GTD/UTD edge- and wedge-diffraction terms.

These differences mainly come from the following aspects. Monte Carlo sampling determines the statistical stability of fine-scale scattering. The parameter analysis shows that the mean ENL and Rayleigh KS values change relatively little between 7 and 10 samples per pixel. Accordingly, all physics-based inputs used for residual learning are generated using the fixed 10-sample configuration. Some details are also lost during 3D model conversion. For example, conversion from laser point clouds to glTF format may fail to fully preserve microscopic surface roughness or heterogeneous material attributes. The measured SAR data contain some uncontrollable environmental factors, such as atmospheric disturbance, soil moisture variation, and system thermal noise, which can usually only be approximated through simplified noise simulation. These fine-scale discrepancies and multiple-bounce-scattering discrepancies are common in SAR applications because measured data are often affected by dynamic environmental conditions. This limits the realism of purely physics-based methods in complex scenes.

The introduction of the residual learning module substantially mitigates the above shortcomings and forms an efficient hybrid physics-based/data-driven strategy. In the refinement module, this paper adopts a U-Net with residual blocks as the core network. Its structural design matches the requirements of residual learning. The skip connections of U-Net support the propagation of large-scale structural information, while the residual blocks facilitate local residual correction. Together, these components improve fine-scale details while maintaining the coarse structural consistency supplied by the physics-based simulation. The component-wise ablation results further demonstrate that the skip connections, residual-learning strategy, residual blocks, CBAM, and perceptual loss contribute to complementary aspects of the refinement process. Their combination improves pixel-level accuracy, structural preservation, and perceptual similarity. This provides quantitative support for the design of the complete refinement network.

In the experiments, after refinement, PSNR increases from 16.63 dB to 23.88 dB, an increase of 7.25 dB; SSIM increases from 0.76 to 0.83, a gain of approximately 8.7%; and LPIPS decreases by 0.0326. These quantitative improvements reflect the effectiveness of the network in compensating for fine-scale image components. Further analysis shows that the statistical distributions in the selected low-structure regions are more consistent with the Rayleigh speckle model. The refined images also show closer agreement with measured data in the spatial distribution of scattering responses. They still retain the large-scale structural scaffold provided by the physics-based simulation and avoid the physical distortion commonly observed in purely data-driven methods.

Local differences between some regions, especially dense wall-related multiple-bounce-scattering areas, and the measured SAR images still remain in the refined results. They are mainly manifested as slight discrepancies in reflection intensity or texture detail. This is mainly because the residual learning module is better at compensating fine-scale texture and speckle, while its ability to correct large-scale structural errors is relatively limited. When the physics-based simulation itself already contains large deviations in these regions, it is especially difficult for the network to fully reverse large-scale structural errors. This phenomenon is more noticeable in areas such as building walls. In the future, it can be considered to increase the Monte Carlo sampling density or introduce adaptive importance sampling to improve higher-order scattering accuracy. Material interaction simulation can also be further optimized by incorporating microwave-specific rough-surface scattering models, so as to provide a more accurate large-scale structural basis for residual learning.

This advantage is particularly prominent under limited-data conditions. The experiment uses only six groups of training image pairs, covering part of the eight viewing directions, and the network can still converge stably. The residual learning design simplifies the learning task into texture mapping rather than complete image reconstruction, thereby reducing the risk of overfitting and improving generalization. In addition, the composite loss balances pixel accuracy, structural similarity, and perceptual quality. The controlled translation analysis further shows that refinement maintains positive PSNR and SSIM gains over the physics-based input for all eight views throughout the tested 0–5-pixel range.

The experiments also reveal some potential challenges. Registration error can be alleviated to some extent by data augmentation methods such as mild affine perturbation, but in scenarios with large deformation, such as high-elevation views, registration error may still amplify residual artifacts. This method is also sensitive to parameter settings. Increasing the number of Monte Carlo samples can further reduce error, but it also increases computation time, so a trade-off between accuracy and efficiency is required. The dataset itself also has certain limitations. SARMV3D mainly focuses on building scenes, and its applicability to other types of targets, such as ships and aircraft, still needs to be further validated in the future. The material model may also need to be adjusted to capture band-specific reflection characteristics. The current model uses a scalar multiple-bounce approximation, and simulation and evaluation use only the HH channel. Performance in target recognition and other downstream tasks requires separate validation. Direct comparisons with other rough-surface scattering models and external computational baselines are needed.

The experimental results show that physics-based simulation and data-driven refinement have good complementarity. Physics-based simulation can preserve physical interpretability and large-scale structural consistency, while data-driven refinement can improve fine-scale details. The combination of the two can improve the overall quality of SAR image simulation. The framework also has the potential to alleviate data scarcity in downstream SAR tasks by providing controllable synthetic data. The above limitations show that there is still room for further improvement in this method.

5. Conclusions

This paper presents a multi-view SAR image simulation and refinement method that combines an OptiX-based ray-tracing framework with residual learning. This method aims to reduce the dependence on measured SAR data, because the acquisition cost of measured SAR data is high and it is difficult to collect under diverse environmental conditions. The main contribution is an integrated framework for multi-view building SAR simulation. It combines OptiX-based multiple-bounce ray tracing, Monte Carlo material sampling, coherent image formation, and residual refinement. The ray-tracing framework generates multi-view SAR image sequences by modeling electromagnetic wave

propagation, Monte Carlo material interaction, and coherent summation. While preserving physical interpretability, the framework can maintain the overall target structure and dominant scattering characteristics.

The residual-learning module further compensates for fine-scale details that are difficult to reproduce through physics-based simulation alone, including speckle texture and local intensity fluctuations. The refined images therefore show higher consistency with measured SAR data in pixel-level accuracy, structural and deep-feature similarity, and dominant scattering responses. The statistical distributions in the selected low-structure regions are also more consistent with the Rayleigh speckle model. The experiments are conducted on the SARMV3D dataset. Physics-based simulation performs well in multi-view consistency and large-scale structure reproduction, as shown by the results. After refinement, pixel-level error, structural similarity, and perceptual similarity are substantially improved. This indicates that the hybrid strategy has high effectiveness and good generalization potential under limited-data conditions in the evaluated building scenes.

The framework makes full use of GPU parallel acceleration and effectively reduces simulation runtime. It also provides a practical foundation for large-scale multi-view data generation. In practical applications, it has the potential to provide high-quality and diverse augmented data for downstream tasks such as SAR target recognition, semantic segmentation, 3D reconstruction, and disaster monitoring, thereby reducing dependence on expensive measured data. The method also provides a potential physical-learning hybrid paradigm for SAR imaging.

Future work can be further expanded in multiple directions. The diversity of scenarios can be enriched by applying the framework to more complex environments, such as urban agglomerations, sea ships, or vegetation-covered areas. The generalization ability can be improved by integrating more glTF models and environmental factors. Other advanced AI methods can also be integrated, such as combining generative adversarial networks (GANs) to assist residual learning, further optimizing high-frequency noise simulation, or introducing attention mechanisms to enhance multi-view consistency. Large-scale application verification can be carried out by integrating the framework into real SAR systems, conducting online data augmentation experiments, and evaluating its practical effectiveness in disaster response. The further application potential of the framework in multi-temporal observation scenarios can also be explored.

These extended directions will further improve the robustness and practical value of the method and promote the further development of SAR image processing toward intelligence and physics-driven approaches.

Author Contributions: Conceptualization, M.Y. and Y.W.; methodology, M.Y.; software, M.Y.; validation, Y.Z.; formal analysis, Y.Z.; investigation, Y.Z. and G.Z.; resources, G.Z., J.G., X.N. and Y.W.; data curation, M.Y., Y.Z., J.G. and X.N.; writing—original draft preparation, M.Y., Y.Z. and Y.W.; writing—review and editing, M.Y. and Y.Z.; visualization, M.Y.; supervision, M.Y., Y.W. and P.L.; project administration, X.N., Y.W. and P.L.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the National Natural Science Foundation of China under Grant 61991420 and Grant 61991421, and in part by the Key Laboratory of Target Cognition and Application Technology, Chinese Academy of Sciences, under Grant 2023-CXPT-LC-005.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The SARMV3D-3.0 and SARMV3D-2.0 datasets used in this study are available through the Journal of Radars data platform at https://www.radars.ac.cn/web/data/getData?newsColumnId=2f2748db-10ef-4ad0-bcc4-f087ce59b6f8 (accessed on 16 May 2026). The data generated during this study are available from the corresponding author upon reasonable request.

Acknowledgments: This study was experimentally validated using the SARMV3D dataset released by the Microwave Vision project team of the Aerospace Information Research Institute, Chinese Academy of Sciences. The public availability of this dataset provided important support for evaluating the simulation and refinement method proposed in this paper. The authors express their sincere gratitude.

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

References

1. Kang, C.; He, C. SAR image classification based on the multi-layer network and transfer learning of mid-level representations. In *Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*; IEEE: Piscataway, NJ, USA, 2016; pp. 1146–1149.
2. Toizumi, T.; Sagi, K.; Senda, Y. Automatic association between SAR and optical images based on zero-shot learning. In *Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Piscataway, NJ, USA, 2018; pp. 17–20.
3. Diemunsch, J.R.; Wissinger, J. Moving and stationary target acquisition and recognition (MSTAR) model-based automatic target recognition: Search technology for a robust ATR. In *Proceedings of the Algorithms for Synthetic Aperture Radar Imagery V*; SPIE: Bellingham, WA, USA, 1998; Volume 3370, pp. 481–492.
4. Xu, F.; Jin, Y.Q. Bidirectional analytic ray tracing for fast computation of composite scattering from electric-large target over a randomly rough surface. *IEEE Trans. Antennas Propag.* **2009**, *57*, 1495–1505. [CrossRef]
5. Yue, D.X.; Xu, F. A coherent generative scheme for SAR image representation. In *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*; IEEE: Piscataway, NJ, USA, 2021; pp. 427–430.
6. Zhang, M.; Zhao, Y.; Li, J.X.; Wei, P.B. Reliable approach for composite scattering calculation from ship over a sea surface based on FBAM and GO-PO models. *IEEE Trans. Antennas Propag.* **2016**, *65*, 775–784. [CrossRef]
7. Xu, F.; Jin, Y.Q. Imaging simulation of polarimetric SAR for a comprehensive terrain scene using the mapping and projection algorithm. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 3219–3234. [CrossRef]
8. Balz, T.; Stilla, U. Hybrid GPU-based single-and double-bounce SAR simulation. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3519–3529. [CrossRef]
9. Houret, T.; Lévêque, O.; Talibart, B.; Roussel, S.; Everaere, E.; Unger, K.; Trouvé, N. Large-Scale SAR Image Simulation using EMPRISE® Software Compared to Real Image acquisitions with SETHI airborne Platform over the Sainte-Marie-la-Mer countryside. In *Proceedings of the 2024 International Radar Conference (RADAR)*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6.
10. Meng, H.; Zhang, L.; Hu, H.; Rao, S.; Gao, B. Research on SAR image simulation of marine geographical feature based on FEKO. In *Proceedings of the International Conference on Advanced Imaging, Detection, and Signal Processing (AIDSP 2024)*; SPIE: Bellingham, WA, USA, 2025; Volume 13693, pp. 69–80.
11. Ding, J.; Chen, B.; Liu, H.; Huang, M. Convolutional neural network with data augmentation for SAR target recognition. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 364–368. [CrossRef]
12. Cui, Z.; Zhang, M.; Cao, Z.; Cao, C. Image data augmentation for SAR sensor via generative adversarial nets. *IEEE Access* **2019**, *7*, 42255–42268. [CrossRef]
13. Xie, D.; Ma, J.; Li, Y.; Liu, X. Data augmentation of sar sensor image via information maximizing generative adversarial net. In *Proceedings of the 2021 IEEE 4th International Conference on Electronic Information and Communication Technology (ICEICT)*; IEEE: Piscataway, NJ, USA, 2021; pp. 454–458.
14. Zheng, C.; Jiang, X.; Liu, X. Semi-supervised SAR ATR via multi-discriminator generative adversarial network. *IEEE Sens. J.* **2019**, *19*, 7525–7533. [CrossRef]
15. Lu, Q.; Jiang, H.; Li, G.; Ye, W. Data augmentation method of sar image dataset based on wasserstein generative adversarial networks. In *Proceedings of the 2019 International Conference on Electronic Engineering and Informatics (EEI)*; IEEE: Piscataway, NJ, USA, 2019; pp. 488–490.
16. Liu, L.; Pan, Z.; Qiu, X.; Peng, L. SAR target classification with CycleGAN transferred simulated samples. In *Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Piscataway, NJ, USA, 2018; pp. 4411–4414.
17. Zhang, M.; Cui, Z.; Wang, X.; Cao, Z. Data augmentation method of SAR image dataset. In *Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Piscataway, NJ, USA, 2018; pp. 5292–5295.

18. Oh, J.; Kim, M. PeaceGAN: A GAN-based multi-task learning method for SAR target image generation with a pose estimator and an auxiliary classifier. *Remote Sens.* **2021**, *13*, 3939. [CrossRef]
19. Guo, Q.; Xu, F. A deep feature transformation method based on differential vector for few-shot learning. In *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*; IEEE: Piscataway, NJ, USA, 2021; pp. 407–410.
20. Song, Q.; Xu, F.; Zhu, X.X.; Jin, Y.Q. Learning to generate SAR images with adversarial autoencoder. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–15. [CrossRef]
21. Dong, G.; Liu, H. A new recurrent architecture to predict SAR image of target in the absent azimuth. In *Proceedings of the IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Piscataway, NJ, USA, 2022; pp. 2610–2613.
22. Agarwal, T.; Sugavanam, N.; Ertin, E. Sparse signal models for data augmentation in deep learning ATR. *Remote Sens.* **2023**, *15*, 4109. [CrossRef]
23. Liu, Z.; Niu, S.; Qiu, X.; Peng, L.; Ding, C. A Method for Analyzing Strong Scattering in SAR Images Based on a Differentiable Simulator. In *Proceedings of the IGARSS 2024–2024 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Piscataway, NJ, USA, 2024; pp. 3327–3331.
24. Cha, M.; Majumdar, A.; Kung, H.; Barber, J. Improving SAR automatic target recognition using simulated images under deep residual refinements. In *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2018; pp. 2606–2610.
25. Loper, M.M.; Black, M.J. OpenDR: An approximate differentiable renderer. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 154–169.
26. Kato, H.; Ushiku, Y.; Harada, T. Neural 3d mesh renderer. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2018; pp. 3907–3916.
27. Liu, S.; Li, T.; Chen, W.; Li, H. Soft rasterizer: A differentiable renderer for image-based 3d reasoning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 2019; pp. 7708–7717.
28. Li, T.M.; Aittala, M.; Durand, F.; Lehtinen, J. Differentiable monte carlo ray tracing through edge sampling. *ACM Trans. Graph. (TOG)* **2018**, *37*, 1–11. [CrossRef]
29. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **2021**, *65*, 99–106.
30. Fu, S.; Xu, F. Differentiable SAR renderer and image-based target reconstruction. *IEEE Trans. Image Process.* **2022**, *31*, 6679–6693. [CrossRef] [PubMed]
31. Jia, H.; Pu, X.; Wang, Y.; Fu, S.; Xu, F. Multi-view Sample Augumentation for SAR Based onDifferentiable SAR Renderer. *J. Radars* **2024**, *13*, 457–470.
32. Wilmanski, M.C.; Tamir, J.I. Differentiable rendering for synthetic aperture radar imagery. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *60*, 82–93. [CrossRef]
33. Lei, Z.; Xu, F.; Wei, J.; Cai, F.; Wang, F.; Jin, Y.Q. SAR-NeRF: Neural radiance fields for synthetic aperture radar multiview representation. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5221415. [CrossRef]
34. Ehret, T.; Marí, R.; Derksen, D.; Gasnier, N.; Facciolo, G. Radar fields: An extension of radiance fields to sar. In *Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: Piscataway, NJ, USA, 2024; pp. 564–574.
35. Li, A.; Lei, Z.; Wei, J.; Xu, F. Sar-gs: Gaussian splatting-based sar image rendering and target reconstruction. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5224913. [CrossRef]
36. Wei, J.; Luomei, Y.; Zhang, X.; Xu, F. Learning surface scattering parameters from SAR images using differentiable ray tracing. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 2006415. [CrossRef]
37. Besl, P.J.; McKay, N.D. Method for registration of 3-D shapes. In *Proceedings of the Sensor Fusion IV: Control Paradigms and Data Structures*; SPIE: Bellingham, WA, USA, 1992; Volume 1611, pp. 586–606.
38. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
40. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19.
41. Qiu, X.; Jiao, Z.; Peng, L.; Chen, J.; Guo, J.; Zhou, L.; Chen, L.; Ding, C.; Xu, F.; Dong, Q.; et al. SARMV3D-1.0: Synthetic aperture radar microwave vision 3D imaging dataset. *J. Radars* **2021**, *10*, 485–498.

42. Qiu, X.; Jiao, Z.; Yang, Z.; Cheng, Y.; Lin, B.; Luo, Y.; Wang, W.; Dong, Y.; Zhou, L.; Ding, C. Key technology and preliminary progress of microwave vision 3D SAR experimental system. *J. Radars* **2022**, *11*, 1–19.
43. Qiu, X.; Luo, Y.; Song, S.; Peng, L.; Cheng, Y.; Yan, Q.; Shangguan, S.; Jiao, Z.; Zhang, Z.; Ding, C. Microwave vision three-dimensional SAR experimental system and full-polarimetric data processing method. *J. Radars* **2024**, *13*, 941–954.

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