Article

Two-Step Deep Learning Approach for Estimating Vegetation Backscatter: A Case Study of Soybean Fields

Dong Zhu [1], Peng Zhao [2], Qiang Zhao [1], Qingliang Li [2], Jinpeng Zhang [2] and Lixia Yang [1],*

[1] Information Materials and Intelligent Sensing Laboratory of Anhui Province, Anhui University, Hefei 230601, China; 24721@ahu.edu.cn (D.Z.); p20101003@stu.ahu.edu.cn (Q.Z.)
[2] National Key Laboratory of Electromagnetic Environment, China Research Institute of Radiowave Propagation, Qingdao 266108, China; zhaop1@crirp.ac.cn (P.Z.); liql@crirp.ac.cn (Q.L.); zhangjp@crirp.ac.cn (J.Z.)
* Correspondence: 19002@ahu.edu.cn

Abstract: Precisely predicting vegetation backscatter involves various challenges, such as complex vegetation structure, soil–vegetation interaction, and data availability. Deep learning (DL) works as a powerful tool to analyze complex data and approximate the non-linear relationship between variables, thus exhibiting potential applications in microwave scattering problems. However, few DL-based approaches have been developed to reproduce vegetation backscatters owing to the lack of acquiring a large amount of training data. Motivated by a relatively accurate single-scattering radiative transfer model (SS-RTM) and radar measurements, we, for the first time to our knowledge, introduce a transfer learning (TL)-based approach to estimate the radar backscatter of vegetation canopy in the case of soybean fields. The proposed approach consists of two steps. In the first step, a simulated dataset was generated by the SS-RTM. Then, we pre-trained two baseline networks, namely, a deep neural network (DNN) and long short-term memory network (LSTM), using the simulated dataset. In the second step, limited measured data were utilized to fine-tune the previously pre-trained networks on the basis of TL strategy. Extensive experiments, conducted on both simulated data and in situ measurements, revealed that the proposed two-step TL-based approach yields a significantly better and more robust performance than SS-RTM and other DL schemes, indicating the feasibility of such an approach in estimating vegetation backscatters. All these outcomes provide a new path for addressing complex microwave scattering problems.

Keywords: backscattering coefficient; radiative transfer model; soybean field; transfer learning; deep neural network; long short-term memory network

1. Introduction

The accurate estimation of vegetation backscatter plays a crucial role in various applications. For instance, a better understanding of the microwave scattering mechanisms of tree components contributes to the study of communication channel sensitivity in forested areas and the detection of targets under the trees [1]. In addition, the applicability to predict variation in backscattering coefficients caused by different moisture conditions at different times could offer valuable information for crop yield estimation [2]. Therefore, accurately modeling the backscatter from a vegetated surface has attracted widespread interest.

In contrast with surface scattering, the interactions between radar signals and vegetation surfaces are rather complicated to interpret because the scattering components in a vegetation canopy usually have various sizes, shapes, and orientations. Over the past

decades, extensive efforts were devoted to better analyzing and characterizing the backscatter behavior of vegetation canopies [3–13]. In consequence, there are two main categories of those methods, including empirical models and theoretical models. Concerning the empirical models, the widely used one is the Water Cloud Model (WCM) [3,7,11]. In the original WCM, Attema et al. [3] treated the canopy as an equivalent water cloud comprising identical scatterers that are uniformly distributed within the canopy. Considering the effect of leaf angle orientation, Kweon et al. [7] added two parameters related to the leaf angle distribution, thus developing a modified WCM. Experimental results demonstrated that the modified WCM accurately estimates the radar backscatters for vegetation canopies. In [11], Mandal et al. appraised the WCM inversion capability using the multi-output support vector regression (MSVR) technique for the simultaneous retrieval of biophysical parameters and the performance of the proposed method was evaluated with in situ measurements. Reportedly, empirical models were only applicable to the conditions under which those radar data were measured [14]. Therefore, such models involved significant limitations.

With all of the theoretical methods, the vector radiative transfer model (RTM) is considered to be the typical one. Many RTM-based models have been reported in the literature [4,5]. In 1990, Ulaby et al. [4] proposed the Michigan Microwave Canopy Scattering Model (MIMICS) to predict the backscatter from a forest. According to MIMICS, a vegetation canopy is geometrically divided into three regions: the crown region, the trunk region, and the underlying soil surface. Each scattering component in MIMICS has a well-defined physical meaning, thus achieving a high degree of precision. However, it is reported that those RTM-based models are relatively accurate but quite complicated and inconvenient since they involve many input parameters. To reduce the complexity, Ulaby et al. [13] treated the canopy elements as equivalent spherical Rayleigh scatterers and limited the solution to the first-order scattering, thereby simplifying the forms of the extinction and phase matrices in RTM derivation, and obtained a simple form, namely, the single-scattering RTM (hereafter called SS-RTM). As a result, this model decreased the input parameters and computational complexity. In [13], the predicted values of SS-RTM were simulated and the results showed that this model can provide good agreement with radar observations for different vegetation canopies. It is, therefore, suitable for generating a simulated dataset.

Essentially a research tool, deep learning (DL) approaches have been widely used in many fields, especially for semantic segmentation [15–17], object detection [18,19], and image recognition [20,21]. On the basis of their impressive progress, extensive efforts have been devoted to exploring the application of DL approaches to remote sensing [22–27] and microwave scattering problems [28–38]. With respect to the field of remote sensing, Zhang et al. [22] proposed a DL-based method to address the limitations in extracting cropland information from large-scale remote sensing imagery. Compared with the existing state-of-the-art methods, the proposed network can achieve high precision and practicality in segmenting large-area croplands. Amin et al. [23] introduced a novel DL methodology for a tree census classification system. By means of this multi-modal training approach, this method achieved good precision in classifying the dominant tree species of Cyprus. Furthermore, to tackle the challenges in handling complex architectural structures and shadow occlusions, Li et al. [25] developed a novel encoder–decoder network, and experimental results demonstrated that the proposed network can achieve superior performance. Furthermore, Paolanti et al. [27] introduced a comprehensive ethical framework designed to assess and quantify the trustworthiness of DL methods in the field of remote sensing. This work provides a practical tool for developers in remote sensing to ensure the responsible deployment of DL. Obviously, the above-mentioned findings proved that DL-based methods are able to obtain a better performance in remote sensing.

In the case of microwave scattering problems, Li et al. [28], for the first time, investigated the connection between DL networks and the iterative method of nonlinear electromagnetic inverse scattering problems. The results demonstrated that the proposed DL-based approach outperforms remarkably conventional inverse methods in terms of quality and computational time. To quantitatively reconstruct the unknown scatterers from phaseless data, Xu et al. [29] proposed a learning-based inversion approach in the frame of U-net. Both numerical and experimental tests were conducted with different DL schemes. Zhou et al. [30] proposed a modified contrast scheme (MCS) to reconstruct both the target objects with high contrast. Numerical results showed that, compared with other schemes, MCS performed well in two-dimensional and three-dimensional examples in real time after an offline training process. In our previous work [31], a multi-model fusion network was introduced to solve the inverse problem. In this method, amplitude and phase of the measured scattering data were applied to train the proposed scheme for the first time, and numerical simulations showed that the proposed method can achieve a better performance in reconstructing homogeneous and heterogeneous scatterers. In addition to the above-mentioned reports, Xiao et al. [36] developed a hybrid approach that combines DL and an experimental design to efficiently and accurately predict the monostatic radar cross section (RCS) of a conducting target. Numerical results showed that the DL-based method can reduce the predictive RMSE compared with traditional methods. Another representative research was reported by Cao et al. [37]. They designed a U-net-based network to perform an efficient scattering center (SC) prediction for targets with coating defects from the input geometric image, and the results proved that the proposed method is promising in providing efficient SC prediction in real-time scenarios. Throughout all this research, most of the research in this field focused on developing an inversion algorithm (namely, inverse problems) to retrieve the parameters of interest [28–34], and few studies looked into the feasibility of applying those approaches in forward problems [35,36], e.g., predicting the radar backscatter of vegetation canopies. One crucial reason is that the performance of DL-based approaches relies on a large amount of training samples. In practice, radar measurements are the most straightforward way to obtain training data. However, to build a ground-truth dataset that is representative of all available conditions of vegetation canopies involving different radar configurations would require a huge time and manpower investment. Furthermore, traditional DL-based methods, particularly those with complex architectures, are often considered a "black box," making it difficult to understand how they make predictions. This lack of transparency can be a significant drawback for scientific applications where understanding the process is as important as the prediction itself.

Alternatively, transfer learning (TL) is regarded as a reliable DL approach to the tasks in which the training data are insufficient [39–45]. Pan et al. [39] focused on categorizing and reviewing the current progress on transfer learning for classification, regression, and clustering problems that are related more closely to data mining tasks. This survey provided a comprehensive analysis for the data mining and DL community. Considering the "black box" nature of DL models, Yosinski et al. [40] experimentally quantified the generality versus specificity of neurons in DL models and provided an insight in transferability. The above research demonstrated the superiority of the TL method. Meanwhile, fine-tuning works as an effective TL strategy provides the ability to leverage the knowledge learned from a source domain and transfer it to the target domain [39]. For instance, in research work [46], Yu et al. presented a novel method by means of deep transfer learning to address the estimation of fractional vegetation cover (FVC). Experiments conducted on Sentinel-2 multispectral satellite images revealed that the proposed method outperformed the traditional method and two other machine learning approaches. Such results

proved the practicability of applying a TL-based method in the field of remote sensing. Furthermore, in order to address microwave scattering problems, some scholars have made great contributions in this field [41–45]. To achieve target classification by resonant scattering electromagnetic signals, Selver et al. [42] presented a convolutional neural network (CNN)-based strategy. By extending such a strategy to the measured data via modern data augmentation and TL techniques, they obtained an improved classification performance for complex targets. In research work [44], Zhu et al. designed a novel pre-training procedure method based on TL, combining the dual advantages of data-driven and physical-driven, to analyze spatial object optical scattering characteristics and an integrating feature. The results demonstrated that the proposed method outperformed other methods. Moreover, Dai et al. [45] proposed a TL-based encoder–decoder network to break the bottleneck in the application of DL methods in the field of electromagnetic compatibility, and the results confirmed the effectiveness of TL. As we know, currently no report has been presented to apply the TL approach to predict the radar backscatter of vegetation canopies. Apparently, a comprehensive investigation of the feasibility and performance of such an approach should be undertaken, which can lead to the potential applications for microwave scattering problems.

The objective of this work was to develop a new backscatter model of vegetation canopies based on limited measured data, aiming to estimate vegetation backscatters more accurately than the theoretical models under real conditions. Motivated by the radar measurements, we proposed a two-step TL-based approach to reproduce the radar backscatter from vegetation canopies. In the first step, a large amount of simulated data was generated based on the SS-RTM. In addition, a zero mean Gaussian random noise with a standard deviation of ± 0.5 dB was added to the computed backscattering coefficients to make the dataset more robust. Then, two baseline networks, namely, a deep neural network (DNN) and long short-term memory network (LSTM), were pre-trained using the simulated dataset. In the second step, a limited number of the measured data was used to fine-tune the weight matrices of the previously pre-trained networks. To facilitate the assessment of the proposed approach, three research questions were posed concerning the capability, effectiveness, and precision. Extensive experiments were conducted on both simulated data and in situ measurements to answer those questions. The results revealed that the proposed novel idea opens a new gateway between conventional scattering methods and DL approaches with various application potentials.

The structure of this paper is organized as follows. Section 2 gives the S- and C-bands' radar data. Section 3 introduces the proposed two-step approach, and then Section 4 illustrates the detailed experimental results. Section 5 presents the discussion. Finally, Section 6 gives the conclusion.

2. Materials

2.1. S-Band Radar Data

CRIRP 2018: As demonstrated in Figure 1a, the China Research Institute of Radiowave Propagation (CRIRP) developed a truck-mounted scatterometer system, namely, CRIRP-SCAT, to measure the hh polarization radar backscatters of the soybean fields at S-band (i.e., 3.2 GHz) at a wide range of incidence angles (i.e., 20° to 60° with 5° interval). CRIRPSCAT mainly consists of three parts, including an antenna module, a vector network analyzer (VNA), and a servo module. Particularly, the antenna module was equipped with six sets of antennas in L-, S-, C-, X-, Ku-, and Ka-bands, respectively. The computer unit was used to send commands to the VNA to transmit signals through the antenna and then receive the echoes from the receiving antenna for the desired data. The servo module, involving a servo control unit and a servo motor, was used to collect the scattering echoes of point (or

distributed) targets with different incidence angles by controlling the elevation angles of the antenna [14]. Table 1 lists the basic configurations of the CRIRPSCAT, correlated to the center frequency, bandwidth, antenna gain, etc.

(a)   (b)

Figure 1. Photograph of the (**a**) truck-mounted scatterometer system; (**b**) test site on 9/30.

Table 1. Configuration of the CRIRPSCAT.

Configuration	Value
Frequency (GHz)	3.2
Bandwidth (GHz)	0.2
Antenna Gain (dB)	$\geq$18
Antenna Beamwidth ($^\circ$)	$\geq$10
Polarization	hh
Platform Height (m)	14

The operation principle of CRIRPSCAT is as follows. First, given a frequency, the VNA transmits the generated signals to the transmission antenna via a feeder. The transmission antenna controls the polarization mode of the signals and delivers this signal to the feeder. Then, the antenna receives the scattering signals from the target and transmits the signals to the VNA, which converts the received signals into the time domain and filters out edge clutter signals to retain the valid signals, finally converting such signals to the frequency domain to obtain the reflected power over the irradiated area. It should be noted that CRIRPSCAT is required to be calibrated before conducting radar measurement. For details of the calibration method, one can refer to [14].

To obtain the backscatter data from soybean fields, a radar campaign was conducted by the CRIRP, namely, CRIRP 2018. The test site is located at 36.349°N, 120.378°E, in Qingdao, Shandong province, China. This campaign measured the backscattering coefficients of the test fields four times (11–13 and 30 September 2018). For each test time, measurements were repeated at various azimuth angles (>16) to increase the number of independent samples. Simultaneously to radar acquisitions, the ground-truth data related to the canopy and surface parameters were measured and are summarized in Table 2. The first column (i.e., Date) represents the time when the data were measured, e.g., 9/11 denotes 11 September 2018. The dielectric constants of soil and vegetation were calculated using the formulas reported in [47,48], respectively. The soybeans were in the fully grown stage from 9/11 to 9/13, with canopy water content all around 0.77 g/cm^3. As can be seen in Figure 1b, on 9/30, the soybeans were in the late stage of growth and the leaves were beginning to

yellow. Furthermore, it can also be found from Table 1 that the water content dropped to 0.5 g/cm^3 at 9/30.

Table 2. Ground measurements of CRIRP 2018.

Date	Canopy Height (cm)	Canopy Water Content (g/cm^3)	Volume Fraction (m^3/m^3)	Particle Radius (cm)	Soil Moisture (cm^3/cm^3)	Surface rms Height (cm)	Correlation Length (cm)
9/11	70	0.77	0.68	0.18	0.16		
9/12	79	0.74	0.69	0.20	0.17		
9/13	80	0.76	0.69	0.21	0.17	2.04	36.7
9/30	83	0.50	0.72	0.19	0.14		

2.2. C-Band Radar Data

Yueh 1992: The backscattering coefficients from soybean fields were measured at C-band (5.3 GHz) by the Centre National d'Etudes Spatiales (CNES) in France over a wide period in 1986 [49]. The radar measurements were conducted at different times (17, 25 July; 7, 14, 20, 28 August; 3, 12 September 1986), and each point of the dataset was measured at three polarizations (hh, hv, and vv) for six incidence angles (i.e., 0° to 50°, 10° interval). Furthermore, the height of radar was fixed at 14.5 m, and the 3 dB beamwidth of radar was 9.2°. Along with the radar measurements, extensive ground-truth data related to canopy characteristics and ground parameters were collected. The underlying soil surface at the test site consisted of 11.1% sand, 61.7% loam, and 27.2% clay.

Wigneron 1999: From July to October 1989, Wigneron [50] made the radar measurements during the growing season of the soybean field in Avignon. The backscatter data from the soybean fields were collected by the scatterometer designed by CNES, at C-band (5.3 GHz) with an incidence angle of 23° for hh polarizations. The soil surface and vegetation characteristics were measured and are given in [50].

Figure 2 shows the range of vegetation parameters for different experiments. It can be observed that, compared with S-band radar data, measurements for C-band cover a longer period and a wider range of conditions. For example, the canopy height varied from 0.7 to 0.83 m for CRIRP 2018 (S-band), while that for Yueh 1992 (C-band) changed from 0.2 to 0.825 m. Furthermore, Figure 3 illustrates the variation in C-band radar data with incidence angles and vegetation parameters. It can be clearly seen in Figure 3a that the backscattering coefficients decreased with the increase in incidence angles. As shown in Figure 3b, the backscattering coefficients also exhibited certain relationship with vegetation parameters. Concretely, the backscattering coefficients gradually increased with increasing canopy water content, demonstrating a trend of positive correlation, while the opposite relationship can be observed for canopy height.
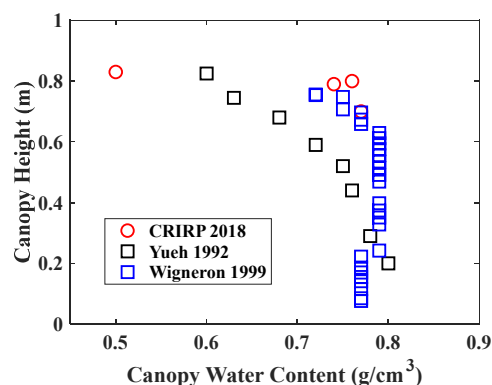

Figure 2. Vegetation parameters for CRIRP 2018, Yueh 1992 [49], and Wigneron 1999 [50].
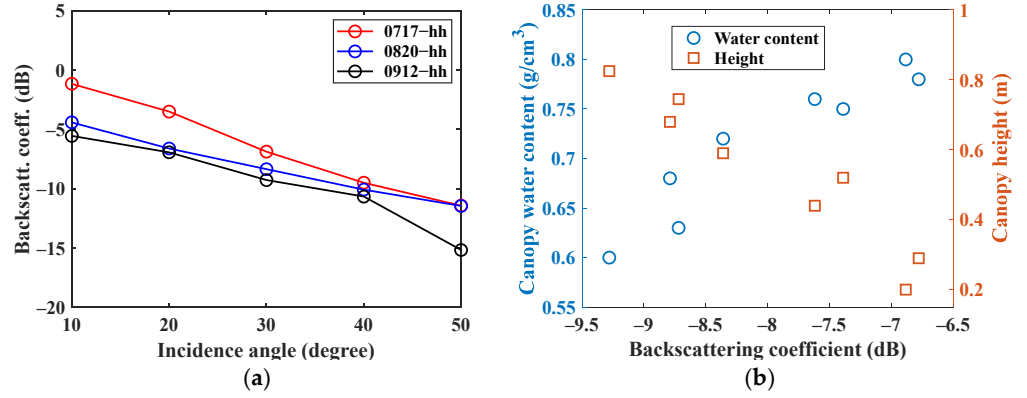
Figure 3. Variation in C-band radar data with (**a**) incidence angles and (**b**) vegetation parameters.

3. Methods

The designed architecture of the proposed two-step TL-based approach is illustrated in Figure 4. As the name implies, this approach was formulated into the following two steps: model pre-training and model fine-tuning. In the first step, an extensive simulated dataset with Gaussian random noise was generated by the SS-RTM in the source domain. Then, two baseline networks were pre-trained with the simulated dataset. In the second step, the previously pre-trained networks were fine-tuned in the target domain with a limited number of radar data introduced in Section 2. Finally, the TL-based predictor for the soybean canopy was obtained.
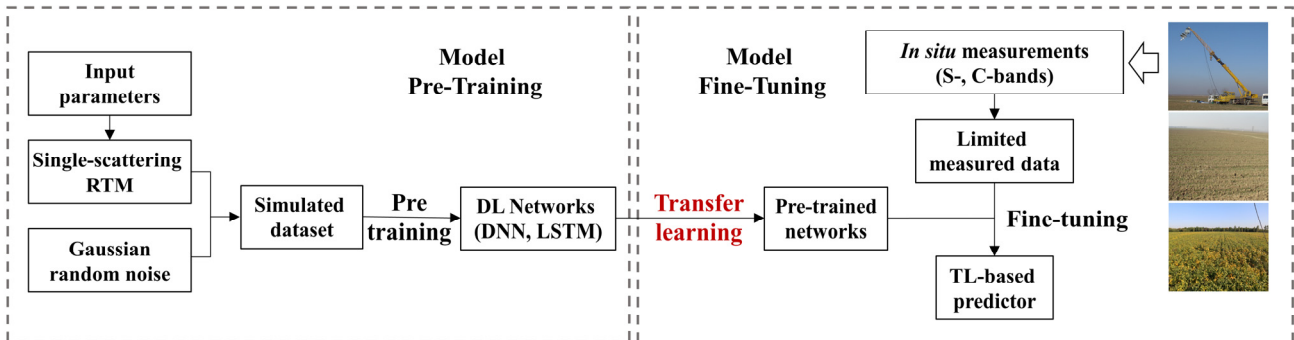

Figure 4. The designed architecture of the proposed two-step TL-based approach for estimating vegetation backscatters.

3.1. SS-RTM-Based Dataset

In research work [13], Ulaby et al. considered the canopy components as equivalent spherical Rayleigh scatterers; thus, the forms of the extinction and phase matrices in RTM were simplified. By limiting the solution to the first-order scattering, the total backscattering coefficients from vegetation canopy can be introduced as follows:

$$
\begin{aligned}
\sigma^0_{pq} &= \sigma^0_{pq1} + \sigma^0_{pq2} + \sigma^0_{pq3} + \sigma^0_{pq4} \\
&= Y_p Y_q \sigma^0_{s_{pq}}(\theta_i) \\
&\quad + \tfrac{3a}{4}\cos\theta_i \left(1 - Y_p Y_q\right)\left(1 + \Gamma^2_{pq} Y_p Y_q\right) \\
&\quad + 6\kappa_s d \Gamma_{pq} Y_p Y_q
\end{aligned}
\tag{1}
$$

where θ_i denotes the incidence angle, $\sigma^0_{s_{pq}}$ is the direct backscattering from the underlying soil surface, Y_p denotes the p-polarized one-way transmittivity of the canopy, with $\exp\left(-\kappa^p_e h \sec\theta_i\right)$, κ^p_e is the p-polarized extinction coefficient of the canopy, h denotes the canopy height, Γ^p represents the p-polarized reflectivity of the underlying surface, and a

represents the single-scattering albedo, with κ_s^p / κ_e^p, where κ_s^p is the scattering component of the p-polarized extinction coefficient.

To compute $\sigma_{s_{pq}}^0$ in Equation (1), we adopt the advanced integral equation model (AIEM) [51], which is considered as a new contribution to the surface scattering computation:

$$\sigma_{s_{pq}}^0 = \frac{k^2}{2} \exp[-s^2(k_z^2 + k_{sz}^2)] \times \sum_{n=1}^{\infty} \frac{s^{2n}}{n!} \left| I_{pq}^n \right|^2 W^{(n)}(k_{sx} - k_x, k_{sy} - k_y) \qquad (2)$$

where $W^{(n)}$ is the Fourier transform of the nth power of the normalized surface correlation function. It should be noted that the exponential correlation function was performed for this paper. For a complete description of the AIEM, one can refer to [51].

Before applying the SS-RTM for generating the simulated dataset, we attempted to validate the applicability of such a model with the measured radar data and the ground-truth data shown in Section 2. Figures 5 and 6 illustrate the comparison between the SS-RTM simulations and the radar data at hh-polarization for the S- and C-bands, respectively. It was clearly observed that the SS-RTM simulations well captured the angular behavior of the measured data for 9/11 (S-band) and 9/12 (C-band) and the values of RMSE were 2.17 dB and 1.99 dB, respectively. As for 9/30 (S-band) and 8/14 (C-band), slight discrepancies were observed, with higher RMSEs of 2.86 dB and 2.40 dB, respectively. Furthermore, it also can be seen that SS-RTM simulations seemed to somewhat overestimate the measured data for all the samples. These may be explained by the fact that: (1) SS-RTM performed the single-scattering mechanism for this paper; however, multiple scattering mechanisms also existed within the canopy particles. (2) There were many difficulties in the actual measurement process, resulting in potentially inaccurate results of vegetation parameters.
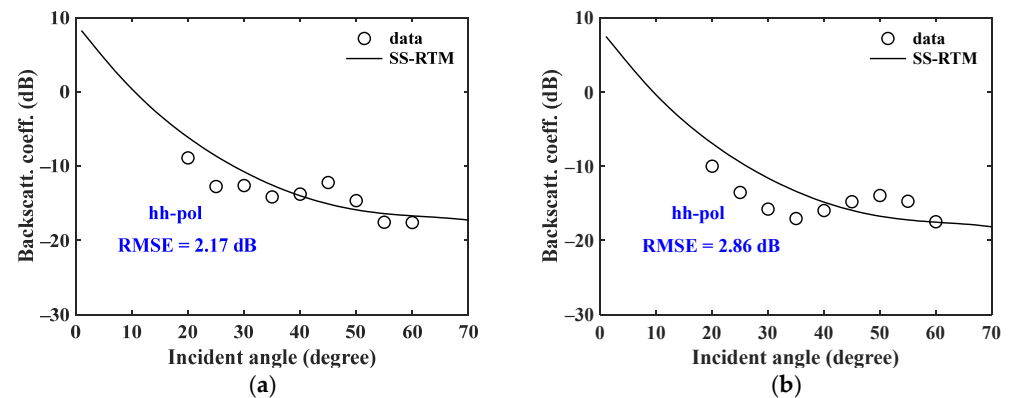

Figure 5. Comparison between the values of backscattering coefficients simulated by the SS-RTM and the radar data for (**a**) 9/11, (**b**) 9/30 of S-band.
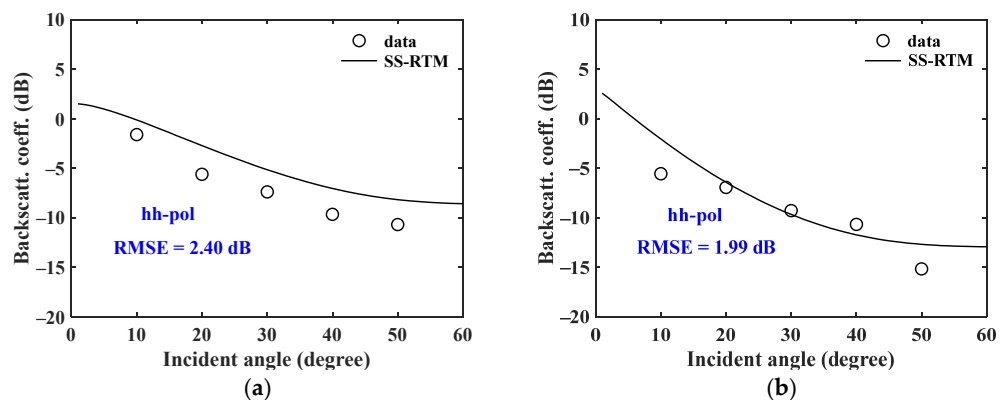

Figure 6. Comparison between the values of backscattering coefficients simulated by the SS-RTM and the radar data for (**a**) 8/14, (**b**) 9/12 of C-band.

On the whole, SS-RTM was reliable in reproducing vegetation backscatter. Other studies also verified that the SS-RTM predictions can yield excellent agreement with the experimental observations for many scene conditions [52]. It is, therefore, suitable for creating a simulated training dataset that covers a wide range of canopy conditions. It should be noted that, although RTM-based methods can also be adopted to general vegetation cover, such as wheat and corn [13,52], due to the limited radar data available to us, we only took into account the capability in the soybean fields.

Since there were no explicit patterns related to all the variables, for the convenience of presentation, we treat Equation (1) as the following form:

$$\sigma_{hh}^0 = \text{SS-RTM}(f, \theta_i; s, l, mv; h, v, r, mg) \tag{3}$$

where f is the frequency, θ_i represents the incidence angles, s and l denote the rms height and correlation length of the soil surface, respectively, mv is the soil moisture content, and h, v, r, and mg denote the canopy height, volume fraction, particle radius, and canopy water content, respectively. According to Equation (3), one can generate a simulated dataset. Table 3 illustrates the ranges and sampling intervals of those parameters. To maintain consistency with the radar data, we restricted f at 3.2 and 5.3 GHz. Other parameters were fixed in a reasonable range to cover different vegetation and soil surface conditions based on the radar measurements.

Table 3. Input parameters of SS-RTM.

Frequency	3.2 GHz		5.3 GHz	
	Range	Interval	Range	Interval
θ_i (°)	20–60	5	20–60	10
s (cm)	1.8–2.2	0.2	1–1.8	0.4
l (cm)	32–36	2	6–14	4
mv (%)	0.11–0.2	0.03	0.1–0.3	0.1
h (m)	0.7–0.8	0.05	0.3–0.7	0.2
v (%)	0.65–0.75	0.05	0.2–0.28	0.04
r (cm)	0.2	-	0.1–0.25	0.05
mg (%)	0.5–0.8	0.1	0.6–0.8	0.1

According to Equation (3) and Table 3, an extensive dataset can be simulated by using SS-RTM. Furthermore, we added a zero mean Gaussian random noise with a standard deviation of ±0.5 dB to the computed values of the backscattering coefficient to make the dataset more robust. After calculations, we obtained 11,664 sets of data for S-band and 14,580 sets for C-band, respectively. These data consisted of input parameters and corresponding backscattering coefficients simulated by SS-RTM, as shown in Table 4. It should be noted that, for this study, 70% of the data was randomly selected as the pre-training dataset, 20% as the validation dataset, and the rest as the test set.

Table 4. Illustration of the dataset.

Data Number	Input Parameters									Labels
	f	θ_i	s	l	mv	mg	h	v	r	σ_h^0
1	3.2	20	0.018	0.34	0.11	0.50	0.70	0.007	0.002	−7.86
2	3.2	20	0.022	0.32	0.11	0.70	0.80	0.0065	0.002	−5.49
3	3.2	25	0.020	0.32	0.20	0.60	0.80	0.007	0.002	−7.72

Table 4. *Cont.*

Data Number	Input Parameters									Labels
	f	θ_i	s	l	mv	mg	h	v	r	σ_h^0
					$\cdots$					
11,665	5.3	30	0.018	0.14	0.20	0.60	0.70	0.002	0.0025	−7.75
11,666	5.3	20	0.01	0.10	0.30	0.60	0.70	0.0028	0.002	−3.69
11,667	5.3	20	0.014	0.06	0.30	0.70	0.30	0.0024	0.002	−5.42
					$\cdots$					
26,244	5.3	50	0.01	0.14	0.30	0.70	0.30	0.0022	0.002	−9.45

3.2. DL Networks

DL can be widely used owing to its powerful capabilities in nonlinear relationship processing and end-to-end feature extraction. More importantly, it is reported that electromagnetic models, combined with the use of DL networks, were regarded as a powerful tool for addressing microwave scattering problems [53]. In this context, two baseline networks were adopted to perform the proposed two-step approach, namely, a deep neural network (DNN) and long short-term memory network (LSTM). The reasons we utilized a DNN and LSTM lie in the fact that:

1. DNNs are highly effective at learning complex, nonlinear patterns from data. In microwave scattering, the underlying relationships among the surface properties, dielectric constants, and scattering behavior can be intricate and difficult to model explicitly. Combined with TL, DNNs trained on an SS-RTM-based dataset can be fine-tuned for the measured data. This allows the network to reuse its learned feature representations, which can generalize well across different scattering conditions.

2. LSTM networks are designed to handle sequential data and capture long-term dependencies. In the proposed method, we treated the input parameters as a sequence in Equation (3). Therefore, LSTMs not only learned the nonlinear relationship between the input parameters and the radar data but also extracted the coupling relationship between the individual parameters, thus boosting the performance of the proposed method.

3. Typically, both DNNs and LSTMs require large datasets and significant training time to converge effectively. Combining with TL can make such models more suitable for applications with limited data availability (i.e., vegetation backscatter problem) and reduce the time needed to train the models.

3.2.1. DNN

As shown in Equations (1)–(3), we observed that the soybean backscatter was related to the parameters on the right side, exhibiting a complex nonlinear relationship. DNN is comprised of many computational units (called hidden neurons) working in parallel and related to each other through connections characterized by multiplying factors. With the presence of the activation functions, DNN is well suited for solving nonlinear problems.

Figure 7a shows the diagram of a DNN with an L-layer and a neuron, respectively. In the forward-propagation process, we assume that ψ represents the set of all the parameters of a DNN, with $\psi = \{\psi_1, \psi_2, ..., \psi_L\}$. In the l-th ($l \in \{1, 2, ..., L\}$) layer, we have n_l nodes and the set of parameters can be represented by $\psi_l = \{W^{(l)}, b_l\}$, where $W^{(l)} \in R^{n_{l-1} \times n_l}$ and $b^l \in R^{n_l \times 1}$ denote the weight matrix and the bias vector, respectively. Taking the first layer as an example, the output of the first layer is given by

$$y_1 = \delta(W^{(1)}x + b_1) \tag{4}$$

where δ represents the activation function. To perform the nonlinear mapping of a DNN, we took ReLU as the activation function, with $\delta(x) = \max(0, x)$. ReLU is a widely used function that learns quickly in different neural networks. Furthermore, the gradient of this function is always a single value, either 0 or 1, which prevents the size of the gradients from reducing exponentially during back-propagation, thus alleviating the gradient vanishing problem.
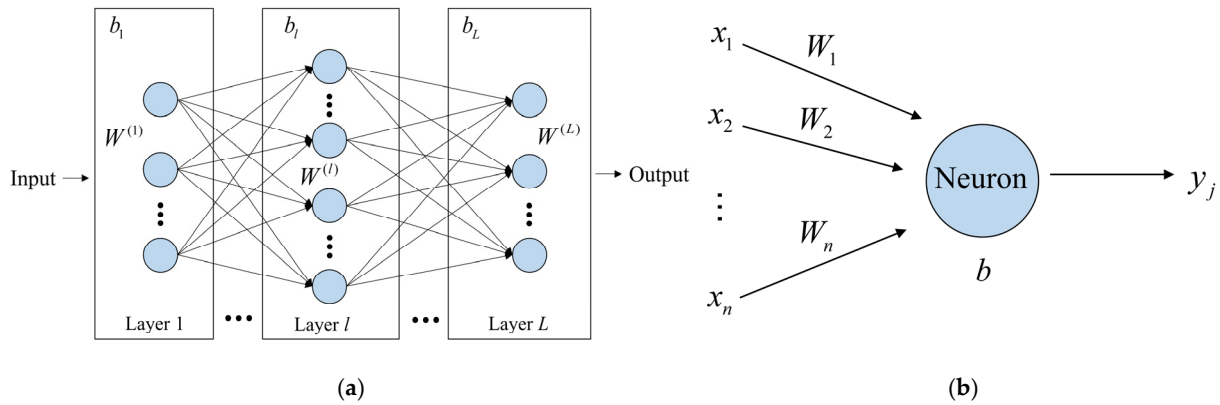
Figure 7. Schematic diagram of (**a**) a DNN, (**b**) a neuron.

More specifically, we assumed that the neuron shown in Figure 7b was the *j*-th neuron in the first layer. In Figure 7b, x_n and y_j are the input and output, respectively. According to Equation (3) and Table 4, we obtained that $n = 9$. And the calculation of the *j*-th neuron can be expressed as

$$
\begin{aligned}
y_j = \delta(W_j^1 * f + W_j^2 * \theta_i + W_j^3 * s + W_j^4 * l + \\
W_j^5 * mv + W_j^6 * mg + W_j^7 * h + W_j^8 * v + W_j^9 * r) + b_j
\end{aligned} \tag{5}
$$

where W_j^n and b_j are the weight and bias of the *j*-th neuron, respectively. Therefore, the output of the first layer is given by

$$
y_1 = \sum_{j=1}^{p} y_j \tag{6}
$$

where p is the number of neurons of the first layer. Then, y_1 can be fed into the second layer and so on. Through layers of abstraction, we can obtain the final output of a DNN, with

$$
y_L = \delta(W^{(L)}(\delta(W^{(L-1)}(\cdots \delta(W^{(1)}x + b_1)\cdots) + b_{L-1})) + b_L) \tag{7}
$$

where x is equal to the input vector and can be found in Equation (3).

3.2.2. LSTM

An LSTM network is designed to prevent gradient explosion and disappearance during the training process [54]. An LSTM is very good at handling sequence data, since its unique internal structure makes it a DL network with nonlinear fitting ability and robustness stronger than other methods [54–56].

In addition to the nonlinear relationship, soybean backscatter also depends on many parameters in a coupled way, as listed in Equation (3). Accordingly, we would like to treat the input variables in Equation (3) as a set of orderly and mutually correlated sequences. With reference to Figure 8, the LSTM block consisted of four units, namely, an input gate, a forget gate, a memory cell, and an output gate. The general formulas of these four units are given by:

$$
i_t = \delta(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{8}
$$

$$f_t = \delta(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{9}$$

$$o_t = \delta(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{10}$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{11}$$

$$h_t = o_t \cdot \tanh(c_t) \tag{12}$$

where x_t and h_t denote the input and final output, respectively, tanh represents the hyperbolic tangent activation function, W is the weight matrix, while b is the bias (for convenience, we ignore the subscripts), and $\odot$ denotes the element-wise product.
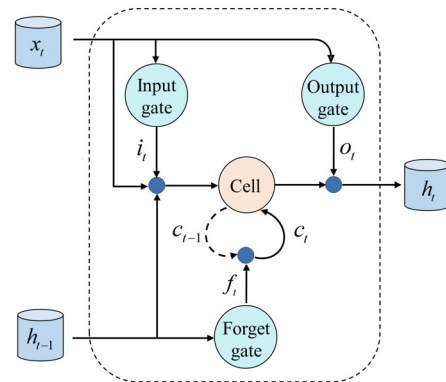
Figure 8. Schematic diagram of LSTM.

As illustrated previously, the simulated dataset did not contain an obvious temporal dimension. With reference to Table 4, we took the order of the samples as the temporal dimension to address this issue, i.e., the input at the first moment was the first set of the samples, the input at the second moment was the second set of the samples, and so on. This solution was proven in [14], and the details are not described here.

3.3. Model Pre-Training

A pre-training strategy is widely used in TL-based approaches and is a common way to improve the generalization performance of neural networks. In view of this, the previously obtained dataset was utilized to initialize the weight matrices and the bias vectors of the DL networks mentioned above.

All the baseline networks used in this paper were mainly implemented with the Keras framework and end-to-end pre-trained on an NVIDIA GTX 1650Ti GPU laptop. We adopted three hidden layers for the DNN, and the number of hidden neurons was set to 32, 64, and 128, respectively. In the case of the LSTM, we implemented three blocks, and the hidden units of these blocks were also set to 32, 64, and 128. Moreover, dropout with a probability of 0.2 was adopted to avoid over-fitting on the simulated datasets. According to Table 4, the dimensions of the inputs and outputs were set to 9 (i.e., columns 1–9) and 1 (i.e., columns 10), respectively. In the pre-training phase, the mean-squared error loss function was optimized by the Stochastic Gradient Descent (SGD) [57] with momentum. To reduce the computational complexity, the backscattering coefficients were normalized to [0, 1]. The initial learning rate during the pre-training phase was set to 0.001. The number of epochs was 500, and the batch size was equal to 32. The momentum was fixed to 0.9, and weight decay was set to 1×10^{-4}.

Furthermore, two error metrics were used to quantitatively verify the performance of the proposed approach, namely, the RMSE and Bias. Their expressions are shown as follows:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_i^p - y_i^t)^2} \tag{13}$$

$$\text{Bias} = \frac{1}{m} \sum_{i=1}^{m} (y_i^p - y_i^t) \tag{14}$$

where y_i^p and y_i^t are the *i*-th predicted result and ground truth, respectively, and *m* is the number of data points.

3.4. Model Fine-Tuning

As described in the pre-training phase, the forward mapping and the input–output discriminant relations were generated on the basis of the dataset computed by the SS-RTM, namely, in the source domain. Since the radar data from soybean fields are rather noisy, application of the pre-trained model for the abstract representation of various input parameters provided robustness to the overall pre-training phase.

Concerning the fine-tuning phase, the implementation details are shown in Algorithm 1. Concretely:

(1)   Constructing Dataset for Fine-Tuning

As illustrated in Algorithm 1, the first step of the fine-tuning phase was Constructing Dataset in the target domain, i.e., pairing the ground-truth data (denoted as X_t) with the corresponding radar backscatter data (denoted as Y_t). Mathematically, the target domain data can be represented as

$$D_t = \left\{ (x_t^1, y_t^1), (x_t^2, y_t^2), \ldots, (x_t^n, y_t^n) \right\} \tag{15}$$

where D_t denotes the target domain, $x_t^i \in X_t$ is the input parameters, while $y_t^i \in Y_t$ is the corresponding output, *i* is the number of the target domain data, and $i = 1, 2, \ldots, n$.

(2)   Initializing the Pre-Trained Network

The overall structures of the pre-trained networks were frozen, and the learned parameters (i.e., weight matrices and bias vectors) among each layer were fine-tuned in the target domain with the limited measured data introduced in Section 2. For a better understanding, we took the DNN as an example and redefined the related parameters. In particular, ψ_s represented the pre-trained parameters of the DNN in the source domain, and it can be written as follows:

$$\psi_s = \left\{ \psi_s^1, \psi_s^2, \ldots, \psi_s^L \right\} \tag{16}$$

where $\psi_s^l = \{W_s^l, b_s^l\}$ is the parameters of the *l*-th layer.

(3)   Fine-Tuning the Networks

In this step, the target domain data shown in Equation (15) were fed into the pre-trained networks. After the nonlinear mapping in the target domain, the predicted values y_s^L for the *i*-th set of the inputs x_t^i can be expressed as:

$$y_s^L = \delta(W_s^{(L)}(\delta(W_s^{(L-1)}(\cdots \delta(W_s^{(1)} x_t^i + b_s^1) \cdots) + b_s^{L-1})) + b_s^L) \tag{17}$$

Then, the discrepancies between the predicted values y_s^L and the radar data y_t^i were used to calibrate the learned parameters of the pre-trained networks based on the backward-propagation algorithm. Note that we adopted the Huber loss function [58] in the fine-tuning

phase, which combined the advantages of the mean-square error and mean-absolute error loss function. With reference to the expression of Huber loss, we have:

$$J(\psi_t; x_t^i; y_t^i) = L_g(y_t^i, y_t^L) = \begin{cases} \frac{1}{2}(y_t^i - y_t^L)^2, & \text{for } |y_t^i - y_t^L| \le g \\ g|y_t^i - y_t^L| - \frac{1}{2}g^2, & \text{otherwise.} \end{cases} \tag{18}$$

where g denotes the hyper-parameter and is equal to 1 in this paper. Then, SGD is used to optimize the objective function $J(\psi_t)$ by updating the parameters ψ_T in the opposite direction of the gradient of the objective function $\nabla_{\psi_t} J(\psi_t; x_t^i; y_t^i)$ related to the parameters [57]. That is,

$$\psi_t = \psi_t - \eta_t \cdot \nabla_{\psi_t} J(\psi_t; x_t^i; y_t^i) \tag{19}$$

where η_t denotes the learning rate in the target domain. Since the radar data were limited, we used a smaller learning rate, i.e., $\eta_t = 1 \times 10-4$. Furthermore, the fine-tuning epoch was set to 100, repeating this loop until the objective function converged to a global minimum. Finally, the fine-tuned TL-based predictors were obtained. This was a desirable feature for our approach, as later we illustrate that pre-trained and fine-tuned networks yielded an improved performance with respect to the theoretical model (i.e., SS-RTM). It should be clarified that, to make the fine-tuning strategy applicable to LSTM, one can simply replace Equations (16) and (17) with their forward-propagation formulas, i.e., Equations (8)–(12).

Algorithm 1: Fine-Tuning Strategy

1: **Input:** The pre-trained weights, the fine-tuning epoch, the learning rate, the loss function;

2: **Step 1:** *Constructing Dataset for Fine-Tuning*

3: Construct the ground-truth data and backscattering coefficients from in situ measurements according to Equation (3);

4: **Step 2:** *Initializing the Pre-Trained Network*

5: Freeze the network structure;

6: Load the pre-trained weights;

7: **Step 3:** *Fine-Tuning the Networks*

8: **while** *fine-tuning epoch is less than num_of_epoch* **do**

9:     Forward-propagate;

10:     **Step 3.1:** *Calculating the Loss Function*

11:     Calculate the Huber loss;

12:     Backward-propagate;

13:     **Step 3.2:** *Updating the Parameters*

14:     Loss gradient for updating the pre-trained weights;

15:     **if** convergence **then**

16:         break loop;

17:     **end if**

18: **end**

19: **Output:** Fine-tuned networks.

---

4. Results

As the key objective of this work was to develop an accuracy vegetation backscatter model with sufficient training data and limited measured data, we evaluated the proposed two-step TL-based approach on simulated data and in situ measurements and conducted extensive experiments to answer the following research questions, namely:

RQ1: Can the pre-trained networks capture the complex relationship well between input parameters and the backscatter from soybean canopy? If yes, which model gives a better representation?

RQ2: Fine-tuning is an efficient strategy to enhance the case of small samples, so how much can it improve in terms of accuracy and can the DL networks stay robust after fine-tuning? RQ3: Does the proposed two-step approach perform better than the pre-trained networks and the SS-RTM under real conditions?

Before presenting the answers, it is important to note that PT- and FT-, which were used in combination with the DNN or LSTM, represented the networks that were pre-trained or fine-tuned, i.e., PT-DNN and FT-DNN denote the pre-trained DNN and fine-tuned DNN, respectively.

4.1. Comparison with the Simulated Test Set (RQ1)

The objective of this section was not only to simply verify the feasibility of DL models for characterizing the relationship between input parameters and backscattering coefficients from a crop such as soybean but also to utilize the test dataset to estimate the performance of the two pre-trained networks. In addition, two other machine learning (ML) algorithms, i.e., Support Vector Machine (denoted as SVM) and Linear Regression (denoted as Linear) were also employed for this section to justify the choice of DNN and LSTM.

The correlation plot for both S- and C-bands are presented in Figures 9 and 10. From Figures 9 and 10, we can make the following observations. First, among these methods, LSTM provided the best performance in predicting the backscattering coefficients, with the RMSE reaching the values of 0.87 dB for S-band and 0.54 dB for C-band, respectively, thus demonstrating the capability of LSTM in predicting backscattering coefficients. Second, the linear model achieved the worst predictions, with the RMSE reaching the values of 1.70 dB for S-band and 2.07 dB for C-band. Such results demonstrate that there was a highly nonlinear relationship between the input parameters and backscattering coefficients and it is not sufficient to use a linear model. Third, concerning the performance of prediction, we can see that SVM can yield comparable results with DNN in terms of RMSE. However, the predicted values of SVM were more dispersed, while those for DNN were closer to the 1:1 line. This phenomenon proves that the capability of SVM is somewhat weaker than DNN in mapping outliers. Therefore, SVM is slightly inappropriate for the context of our case.

Figure 11 illustrates the box plot of the estimated differences between the DL model predictions and simulated results for S- and C-bands. The box plot is a statistical diagram that shows the dispersion of data, and the statistical representations are shown in blue font in Figure 11. Note that the labels on the horizontal coordinate in Figure 11 represent the S-band and C-band models, i.e., S-PT-DNN denotes the pre-trained DNN of S-band. It can be clearly observed from Figure 11 that the median of the estimated differences generally varied in the range of 0 to 2 dB. Compared with PT-LSTM, the performance of PT-DNN was slightly behind, especially for C-band, with the maximum difference reaching a value of 4.2 dB. Additionally, an increased number of outliers was found for C-band but not for S-band. These observations may involve the fact that, as the frequency increased, the wavelength was greater than the size of the scattering components within the soybean canopy, leading to inaccurate estimates of the SS-RTM and thus suppressing the accuracy of the DL networks.

Overall, the preceding results convincingly support the point that the pre-trained networks can not only process backscatter signals but also capture the intrinsic relationship among radar echoes and input parameters, justifying the ability of such DL networks to estimate backscattering coefficients of vegetation canopy. Moreover, the calculated error metrics revealed that the LSTM network was even more powerful than the DNN. That is to say, in addition to considering the nonlinear relationship among input parameters and backscattering coefficients, the interrelationship between input sequences was equally significant.
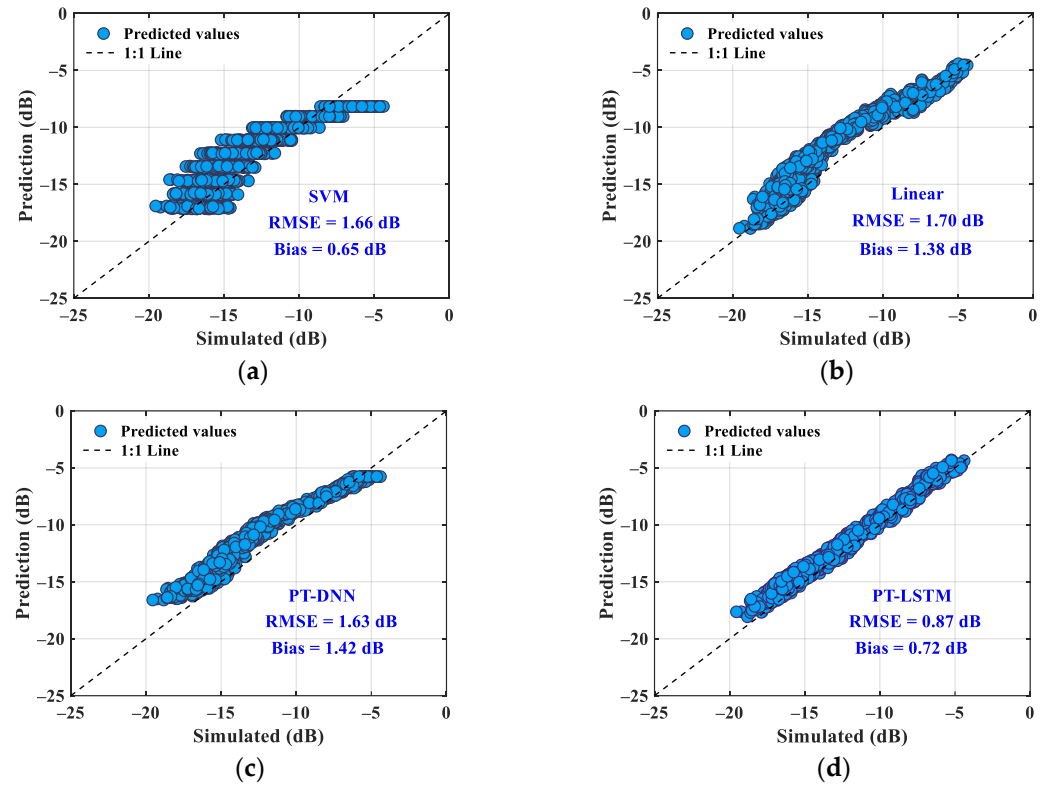
Figure 9. Correlation plot between predictions and validation data for S-band: (**a**) SVM; (**b**) linear; (**c**) PT-DNN; (**d**) PT-LSTM.
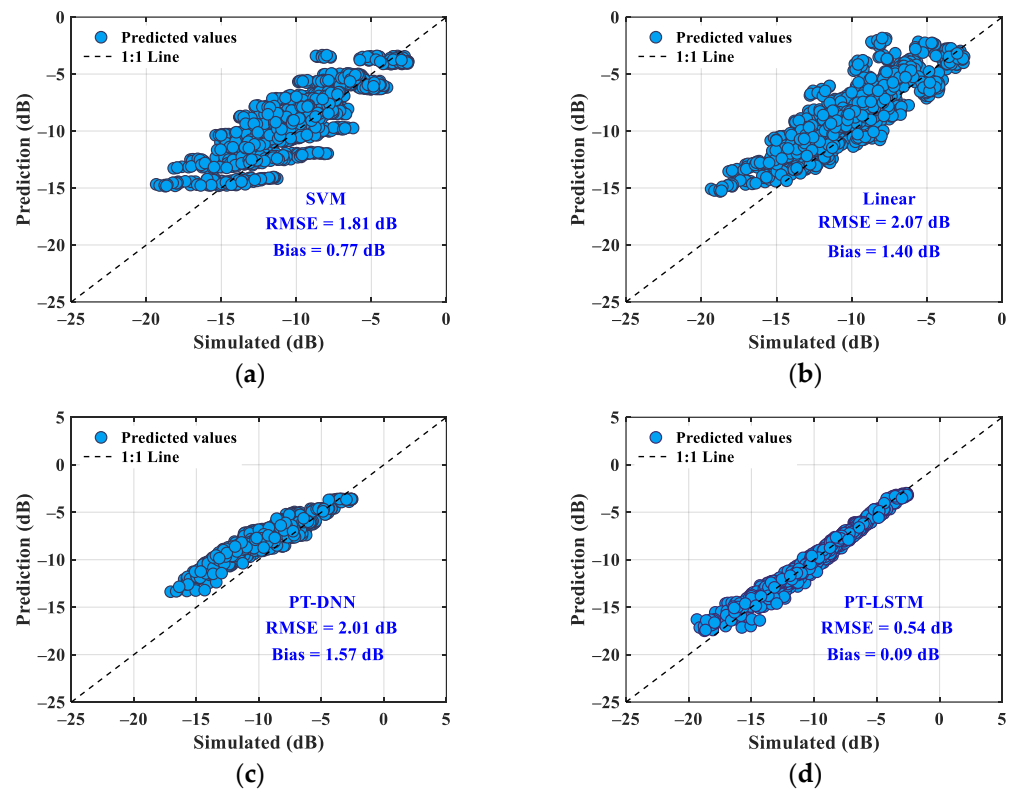

Figure 10. Correlation plot between predictions and validation data for C-band: (**a**) SVM; (**b**) linear; (**c**) PT-DNN; (**d**) PT-LSTM.
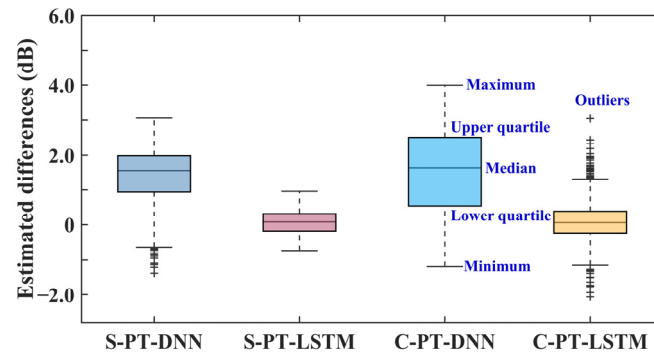
Figure 11. Box plot of the estimated differences between model predictions and simulated data.

4.2. Efficiency and Robustness of the Fine-Tuning Strategy (RQ2)

To answer RQ2, we built a dataset to verify the effectiveness of the fine-tuning strategy, namely, the fine-tuning dataset. This dataset included CRIRP 2018 for S-band and Yueh 1992 for C-band, which are illustrated in Section 2. It is worth mentioning at this point that the fine-tuning dataset only comprised 36 and 32 sets of the measured data for S- and C-bands, respectively. As a result, such datasets with insufficient data are well suited for evaluating the effectiveness of the fine-tuning strategy. In particular, 80% of the fine-tuning set was randomly selected to fine-tune the networks; the rest tested for validation. With reference to Algorithm 1, once the fine-tuning dataset was gradually fed into the pre-trained networks, the validation set was used to investigate the performance of the fine-tuned networks. Figure 12 shows the mean validation loss curves of the fine-tuning phase on the validation set.
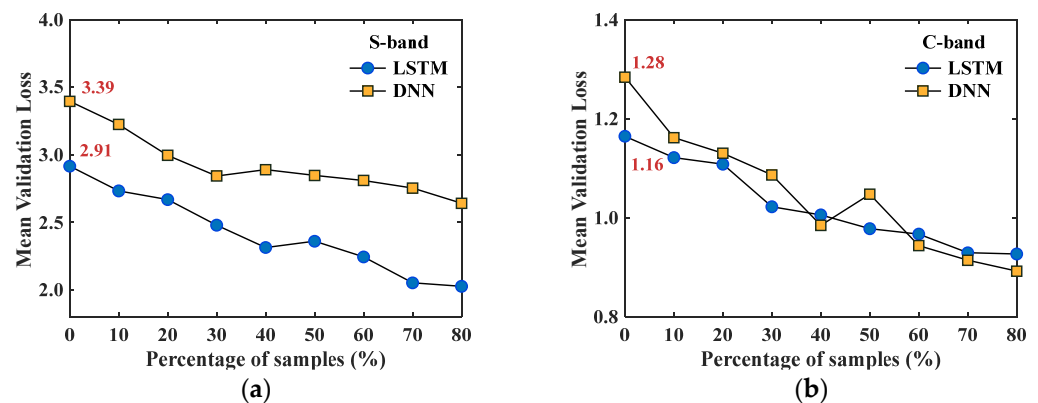

Figure 12. Validation loss on the radar data for (**a**) S-band and (**b**) C-band.

As depicted in Figure 12, the horizontal coordinate started from 0, which means we first evaluated the capabilities of the pre-trained networks with the validation dataset. The computed losses, regarding PT-DNN and PT-LSTM, were equal to 3.39 and 2.91 for S-band and 1.28 and 1.16 for C-band, respectively, which further proves that PT-LSTM outperformed PT-DNN. We can also observe from Figure 12 that, during the fine-tuning phase, the loss curves for S-band tended to drop dramatically compared to those for C-band. This may be explained by the fact that the estimations of the pre-trained networks significantly differed from S-band radar data, leading to a sustainable calibration of the weight matrices. Furthermore, as the fine-tuning samples proportionally increased, the loss curves remained relatively smooth, revealing that the weights matrices gradually converged towards the optimal configuration. On the whole, applying the fine-tuning strategy motivated the pre-trained networks to achieve better performance, even with only 10% of the fine-tuning samples.

To quantify the performance improvement, the improvement ratio is defined as the following from,

$$\text{ratio} = \frac{PT_{\text{RMSE}} - FT_{\text{RMSE}}}{PT_{\text{RMSE}}} \times 100\% \tag{20}$$

where PT_{RMSE} and FT_{RMSE} represent the RMSEs of the pre-trained and fine-tuned networks on the validation set. Small values of the ratio correspond to large values of FT_{RMSE}, and large values of the ratio correspond to small values of FT_{RMSE}. Table 5 lists the corresponding improvement ratio. Evidently, the fine-tuned strategy significantly reduced the RMSE on the validation set, especially for PT-LSTM at S-band, with the ratio values increasing from 8% to 31%. In comparison with S-band, the improvement ratio for C-band varied somewhat slowly, at approximately 15%. Consistently, this result was similar to the trend of the loss curves plotted in Figure 12b.

Table 5. Improvement ratio of the fine-tuning strategy.

Percentage (%)	S-Band		C-Band	
	FT-DNN	**FT-LSTM**	**FT-DNN**	**FT-LSTM**
10	9.39%	7.96%	7.02%	8.34%
20	7.77%	11.55%	8.33%	15.27%
30	15.89%	23.36%	5.90%	15.35%
40	17.09%	25.50%	4.84%	15.32%
50	19.71%	27.48%	6.70%	16.91%
60	20.62%	26.92%	8.91%	18.82%
70	21.34%	28.36%	12.92%	21.16%
80	24.69%	30.91%	18.01%	23.17%

For this section, we also evaluated the robustness of the fine-tuned networks based on Wigneron 1999 [50], since it was not used to fine-tune the pre-trained networks. As shown in Figure 13, a comparison of the estimated backscattering coefficients and the measured ones is depicted for different models. It can be clearly observed that fine-tuning worked as a robust calibration strategy, since the fine-tuned predictions were closer to the 1:1 line, demonstrating a quite strong correlation with the measured data. In Figure 13a,b, we can see that the RMSE values of SVM and Linear were, respectively, 2.79 dB and 3.06 dB, proving that such models are not effective. Furthermore, compared with the pre-trained networks, those that were fine-tuned reduced the RMSEs by 0.48, and 0.97 dB, respectively. Concerning the robustness, DNN, SVM, and Linear seemed to be slightly worse than that of LSTM.
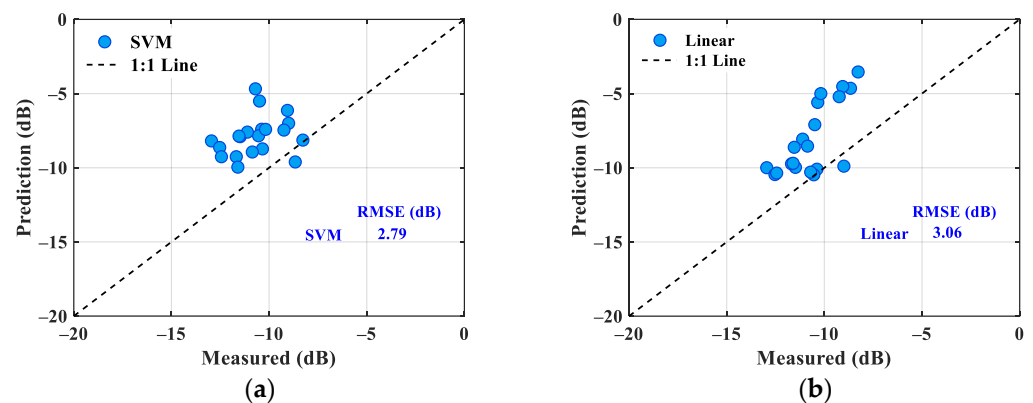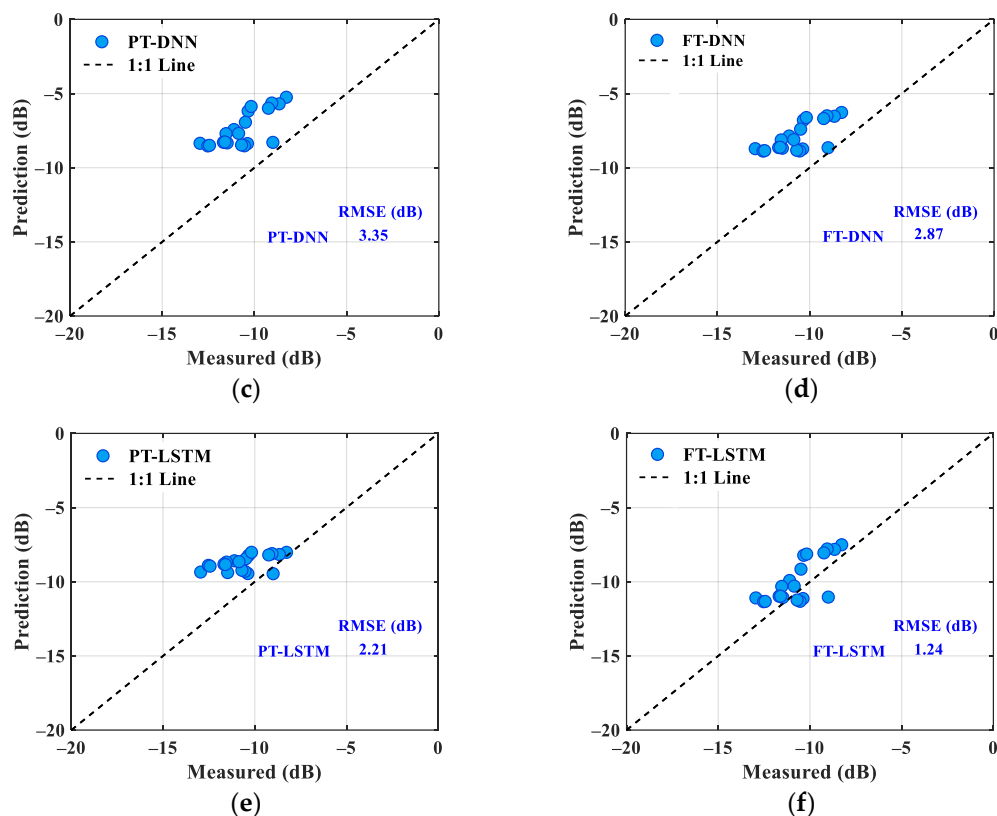

Figure 13. *Cont.*

Figure 13. Correlation of the backscattering coefficients between the measured data and other methods for (**a**) SVM; (**b**) Linear; (**c**) PT-DNN; (**d**) FT-DNN; (**e**) PT-LSTM; and (**f**) FT-LSTM.

To conclude this section, it was shown that fine-tuning is a powerful and flexible strategy to address the lack of training data in the target domain and it enables the pre-trained networks to be more robust as well as to further enhance the performance. This has a guiding significance for addressing microwave scattering problems without sufficient data.

4.3. Comparison with the Radar Measurements (RQ3)

For this section, we attempted to verify the performance of the proposed TL approach with the S- and C-bands' experimental data. Note that two sets of radar data (i.e., 9/13 for CRIRP 2018, 9/03 for Yueh 1992, respectively) were selected for testing, while the rest of the data were used to fine-tune the pre-trained networks.

Quantitatively, we compared the fine-tuned networks with the SS-RTM, ML algorithms, and the pre-trained networks. The detailed error metrics related to the backscatter predictions are computed in Table 6. We can make the following observations from Table 6. First, the performance of PT-DNN was somewhat worse than SS-RTM for S-band, whereas PT-LSTM yielded a slightly improved estimation with the lower values of RMSE and Bias (3.50 and 3.31 dB, respectively). Second, after fine-tuning with a limited amount of the radar data, the precision of such DL networks was enhanced, especially for FT-LSTM, with the RMSE value reduced from 3.50 to 2.32 dB for S-band and 1.28 to 0.99 dB for C-band. The result illustrates that using the fine-tuning strategy is an effective way to achieve higher accuracy. Third, as for the Bias, all of those networks overestimated the measured data. This could be explained by the fact that the fine-tuned data were insufficient for such networks to learn the angular behavior adequately. Fourth, concerning the performance of ML algorithms, SVM achieved better results than Linear for both S- and C-bands, however, slightly behind LSTM. Fifth, it was clearly observed that both PT-LSTM and FT-LSTM performed better than those for DNN, thus indicating that LSTM can not only extract

the interrelationships between input parameters and vegetation backscatters but also its performance was even better than the SS-RTM under real conditions.

Table 6. Error metrics for different models.

Models	9/13 (S-Band)		9/03 (C-Band)	
	RMSE (dB)	Bias (dB)	RMSE (dB)	Bias (dB)
SS-RTM	3.56	3.37	2.04	2.01
SVM	2.69	1.70	1.43	−0.14
Linear	3.08	1.73	2.23	0.92
PT-DNN	3.83	3.59	1.68	1.58
PT-LSTM	3.50	3.31	1.28	0.91
FT-DNN	3.24	2.76	1.23	1.20
FT-LSTM	2.32	1.98	0.99	0.08

To visually show the performance of the proposed two-step TL approach, we took LSTM predictions as an example, as shown in Figure 14. It can be seen in Figure 14a that, as the incidence angle increased, the deviation between the SS-RTM simulations and the measured data gradually increased. This may be explained by the fact that: (1) the SS-RTM performed a single-scattering mechanism, whereas the contribution of multiple scattering became more significant as the incidence angle increased, and (2) the ground measurements of radar backscatter from the vegetation canopy may not always be consistent. It was also observed that the PT-LSTM predictions almost reconstructed the values of SS-RTM, indicating that LSTM can learn the relationship between parameters and vegetation backscatter well. Furthermore, FT-LSTM yielded the best performance, with the lowest RMSE value of 2.32 dB. Compared with PT-LSTM and SS-RTM, FT-LSTM reduced the RMSEs by 1.18 and 1.24 dB. As expected, the predicted values of FT-LSTM, on the whole, captured the angular behavior of the radar measurements for C-band (see Figure 14b), with the lowest RMSE of 0.99 dB. Due to the fine-tuning strategy, FT-LSTM provided better accuracy than PT-LSTM among the largest part of the angular range under consideration. Concretely, compared with PT-LSTM and SS-RTM, the RMSE value in this case was decreased by 0.29 and 1.05 dB, respectively. Hence, the proposed two-step LSTM can better characterize the relationship between input parameters and vegetation backscatter for both S- and C-bands.
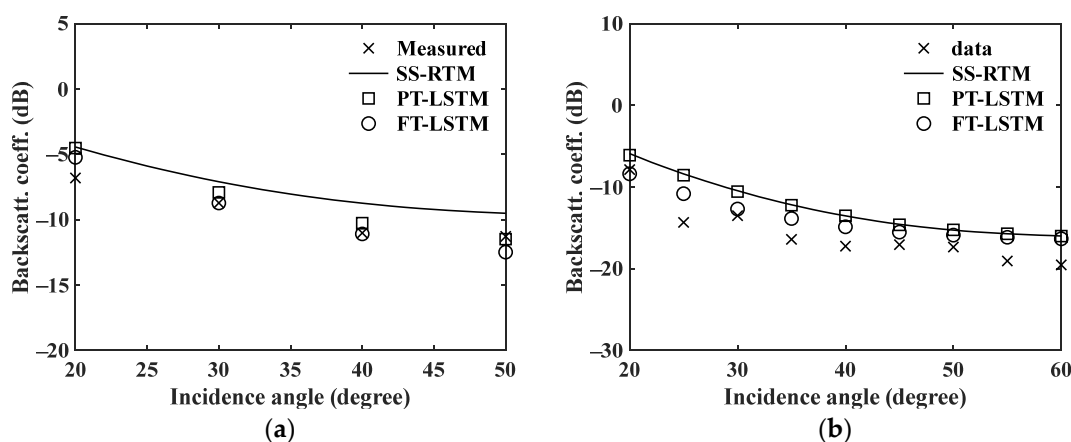

Figure 14. Comparison of backscattering coefficients among the measured data and SS-RTM simulations, PT-LSTM, and FT-LSTM for (**a**) S-band, (9/13, CRIRP 2018) and (**b**) C-band, (9/03, Yueh 1992).

In summary, the above-mentioned outcomes indicate that the proposed two-step approach yielded a better performance, in terms of quantitative and visual, for predicting the backscattering coefficients from soybean fields. Meanwhile, LSTM was more suitable for addressing the complex relationship compared with DNN, due to its distinctive structure. This gives LSTM the applicability to reproduce the radar backscatter from other types of vegetated cover.

5. Discussion

It is undeniable that measured data are of great importance for developing a new empirical or theoretical model as well as for characterizing and estimating canopy backscatters. However, to conduct radar campaigns involves huge resources of time and manpower, thus resulting in insufficient amounts of measured data. Aiming to develop a more accurate vegetation backscatter model based on limited measured data, we, for the first time, introduce a two-step TL-based approach, namely, model pre-training and model fine-tuning. Extensive experiments proved that the proposed approach performs better than other methods in estimating backscattering coefficients. Furthermore, a comprehensive assessment of such an approach concerning the advantages and limitations is illustrated in the following sections.

5.1. Advantages' Assessment of the TL-Based Method

5.1.1. Sustainability

As previously mentioned, the baseline DL models were pre-trained using the dataset simulated by SS-RTM in the pre-training process, which means those models already learned the relationship between input parameters and canopy backscatter and they can be reused. When the pre-trained models are fine-tuned using the measured data, the weight matrix and bias vectors of such models are modified to make them converge in a more accurate direction. Once additional measured data are available, the fine-tuning process is sustainable on the previously obtained models, thus significantly reducing the calculation expenses and time consumption. For other methods, they require end-to-end training. For example, in research work [59], Wang et al. developed a DNN network to reproduce the backscattering coefficients of a land surface based on AIEM. Although this method can give relatively favorable results, a major drawback is that such a model needs to be retrained once new data are acquired. On the contrary, the proposed TL-based method only requires fine-tuning, thus demonstrating its superiority.

For bettering clarification, we conducted a comparison experiment to prove the efficiency of the proposed method. Concretely, DNN and LSTM were selected as the end-to-end DL models for comparison. As stated above, the pre-trained networks can be reused. Therefore, for the proposed method, only the fine-tuning process is needed in this experiment. Taking S-band as an example, we mixed the simulated data and the measured data to obtain 11,700 sets of data for end-to-end training and 36 sets of the measured data for fine-tuning, as illustrated in Section 4.2. Furthermore, the experimental settings are the same as those in Sections 3.3 and 3.4. We listed the results of the proposed method and traditional end-to-end DL models in Table 7. It can be clearly seen that, compared with end-to-end DL models, the proposed method can significantly decrease the time consumption, which was reduced by 32 times for DNN and 38 times for LSTM, respectively. Concerning the memory, due to the sustainability of the pre-trained models, the proposed method only updates the learned parameters, and the memory remains constant with the pre-trained model. These findings demonstrate that the proposed method is more efficient than other end-to-end training methods in reducing the computational cost.

Table 7. Comparison of computational cost.

Method	End-To-End Training		Proposed Method	
	DNN	**LSTM**	**DNN**	**LSTM**
Time (s)	916.2	1414.1	28.3	37.8
Memory (KB)	176	858	-	-

5.1.2. Generalizability

The reason for the generalizability of the TL-based method lies in the following aspects. First, this paper considered two baseline networks (i.e., DNN and LSTM) and verified their precision. Alternatively, such baselines can be replaced once other models with better performance are available. Second, the proposed TL-based method acts more as a uniform framework. Researchers can flexibly modify the corresponding methods within this framework according to the problems they encounter. Taking surface scattering [59] as an example, one can use AIEM to generate the simulated dataset for pre-training DL models and fine-tune such models with the measured data illustrated in [59]. It is therefore entirely generalizable to other studies with insufficient experimental data. Furthermore, microwave scattering behavior varies significantly across different frequency bands (e.g., L, S, C, X bands) and polarizations (e.g., HH, VV, HV). Constructing separate models for each frequency and polarization can be computationally expensive and time consuming. TL can be used to generalize models across frequency bands and polarizations. For example, a model trained on L-band HH polarization can be adapted to C-band VV polarization with minimal additional training according to the TL method. This would enable the development of multi-frequency and multi-polarization scattering models that are more efficient and scalable, allowing for broader applications such as multi-sensor integration.

5.1.3. Reliability

Firstly, WCM [3] is another typical model to simulate the backscatter from vegetation canopy, as mentioned earlier. To provide justification for the choice of employing SS-RTM, this section evaluates the performance of WCM. Note that C-band data were used to fit the unknown parameters with $A_{hh} = -24.73$ and $B_{hh} = 0.072$. Figures 15 and 16 show the comparison results of SS-RTM and WCM on S-band data. It can be clearly seen that the predictions of SS-RTM outperformed those for WCM. Therefore, it is reasonable to believe that using the training dataset generated by SS-RTM can yield better estimations.
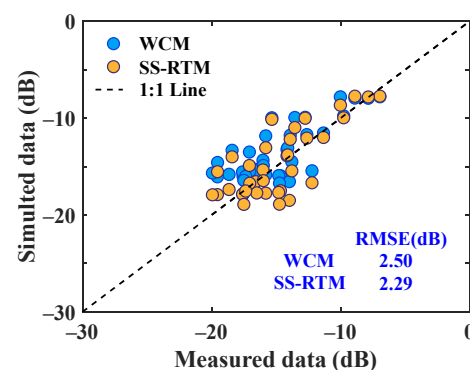
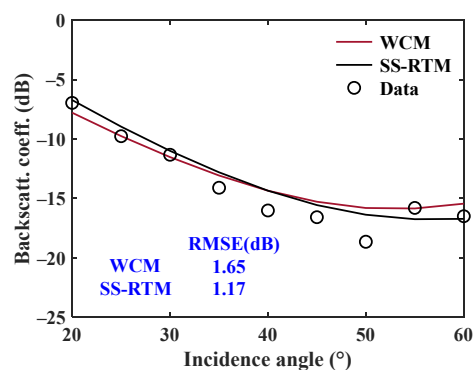**Figure 15.** Comparison of the backscattering coefficients between SS-RTM and WCM.

Figure 16. Comparison for S-band data (9/12, CRIRP 2018).

Secondly, we conducted an experiment to evaluate the time consumption of the proposed method for estimating vegetation backscatter on a large scale, and the results are listed in Table 8. From Table 8, we can see that the proposed method works more efficiently than other models. Furthermore, from the results illustrated in Section 4, it can be also observed that the proposed TL-based method can yield better performance in terms of capability, effectiveness, and precision. All the above outcomes, involving dataset generation, estimation accuracy, and efficiency of large-scale simulation, prove the reliability of the proposed TL-based method.

Table 8. Time consumption for different methods on a large scale.

| Method | WCM | SS-RTM | Proposed Method | |
			DNN	LSTM
Time (s)	5.37	4.15	2.86	2.95

5.2. Limitations' Assessment of the TL-Based Method

It is well known that DL networks rely heavily on the quality of training data. Therefore, a few additional limitations on the proposed method deserve to be clarified. First, the backscatter from the underlying surface contributes significantly to the total radar echoes from a canopy. Among the available methods for the computation of soil surface backscatter, it is unknown whether the AIEM simulations can be applied directly to the soil surface under a vegetation cover, especially under large roughness conditions (see Table 2). Actually, in our previous work [14], we found some non-negligible discrepancies simulated by AIEM with the measured radar data based on the radar measurements conducted by CRIRP, and the error can be up to 4.06 dB for HH polarization and 5.13 dB for VV polarization, respectively. Those can affect the quality of the precision to some extent. In our future work, efforts will be devoted to applying the method developed in [14] to the TL-based method to further improve the performance.

Second, it is worth mentioning at this point that SS-RTM could be suitable for developing an extensive training dataset. Such a model, however, is derived on the basis of several limiting assumption [13], e.g., to simplify the forms of the extinction and phase matrices, the scatterers within the canopy were treated as equivalent spherical Rayleigh scatterers, which means that the wavelength should be much larger than the particle size. In practice, the particle size may not always meet this assumption. Furthermore, as stated previously, SS-RTM performed the single-scattering mechanism for this paper. In research work [13], Ulaby et al. reported that second- and higher-order scattering contributions are significantly smaller than first-order (i.e., single-scattering) contributions within the RTM-based method, and experimental results proved that this assumption is reasonable [13]. However, in the case of dense vegetation canopies, the contribution of multiple scattering can be

significant. Moreover, concerning the overestimations (see Figure 4), the potential reason may lie in the inaccuracy of the measurement of vegetation parameters, and inevitably, such errors are unavoidable. Therefore, SS-RTM is somewhat inadequate for simulating the vegetation backscatter.

Third, it is noted that the pre-trained networks were trained with the simulated dataset computed by SS-RTM. As stated above, SS-RTM involves some drawbacks. Despite the fact that we added Gaussian random noise to the computed results, the simulated dataset is somewhat inaccurate for the backscatter characterization of vegetation in real-world conditions, thus degrading the performance of the proposed method to some extent. Future work will be devoted to developing more accurate algorithms for vegetation canopy backscatter in real-world conditions.

5.3. Broader Implications of the TL-Based Method

5.3.1. Increased Efficiency in Data-Limited Conditions

Many microwave scattering problems, especially in remote sensing or radar applications, require a significant amount of high-quality labeled data to develop accurate models. However, it is difficult to achieve a representative dataset. Fortunately, TL can be used to solve the above issue. TL-based methods can reduce the data requirements and computational costs for model training in new environments, allowing for faster deployment and improved model performance in regions or applications with limited labeled datasets. For instance, Yu et al. [46] found that there are few DL-based methods for estimating fractional vegetation cover (FVC) due to the difficulty of obtaining a large amount of training data. To address this issue, a novel method based on a TL and physical model was proposed, and the experimental results revealed that a TL-based method can overcome the limitations in determining the parameters of an empirical method and machine learning for FVC estimation. More importantly, the proposed TL-based method can significantly improve the FVC estimation accuracy with limited samples, thus demonstrating the superiority of such a method. Furthermore, other research works also proved this point [60,61] and will not be discussed here.

5.3.2. Adaptability to Inverse Problems

A TL-based method can be a powerful tool for solving inverse problems, particularly when data are limited, noisy, or difficult to collect. Typically, inverse problems involve inferring the underlying model or system from observed data, and one can range from tasks like inverse scattering [41,43], to inverse modeling [62]. For instance, full waveform inversion (FWI) can be highly nonlinear and underdetermined. Low-frequency information, which can seriously affect the FWI results, may be missing from obtained data due to the limitation of acquisition devices. Considering this issue, Jin et al. [43] developed a learning-based method to extrapolate the low-frequency data using TL, and the results revealed that the proposed method is efficient and effective. In research work [62], Cheng et al. proposed an inverse modeling method for nanophotonic structures, based on a mixture density network model enhanced by TL, and numerical results validated the proposed method. All these findings revealed that, by transferring knowledge learned from one problem to another, TL can reduce the amount of data needed and accelerate the process of solving inverse problems.

5.3.3. Cross-Domain Transfer for New Applications

Microwave scattering models are often specialized for specific applications, such as environmental monitoring or material analysis. When new applications arise, constructing models from scratch can be slow and resource intensive. Knowledge from one domain can be transferred to unrelated domains, including medical imaging [63–65] or structural

health monitoring [66,67]. For example, to address the limitation of data scarcity and costly annotation processes of medical image data, Dao et al. [65] introduced a novel TL-based method for medical image classification, and the results demonstrated that this approach significantly enhances image classification tasks while reducing the burden of annotation costs. In research work [66], Shao et al. proposed a damage quantification method combining TL in the field of aerospace structural health monitoring. The results showed that the model after transfer was superior to the model without transfer in both average error and maximum error, proving that a TL-based method can reduce training time and maintain high prediction accuracy. This cross-domain transfer significantly speeds up the development of models for emerging applications.

6. Conclusions

Accurately predicting vegetation backscatter using radar signals presents several challenges, primarily due to the complexity of interactions between radar waves and vegetation canopies as well as the influence of environmental factors. Motivated by DL techniques and radar measurements, in this paper, we introduced a new two-step TL approach to estimate the backscatter from vegetation canopy in the case of soybean fields. These two steps implemented the model pre-training and model fine-tuning processes, respectively. In the first step, two baseline networks were pre-trained by a simulated training set with Gaussian random noise, which was generated for various types of soybean canopies at different frequencies (S- and C-bands) based on the theoretical model (i.e., the SS-RTM). In the second step, the previously pre-trained networks were fine-tuned with the in situ measured ground-truth data and the radar data. On the basis of three research questions, we conducted extensive experiments to evaluate the performance of the proposed approach, and the results are listed as follows.

Firstly, we conducted a comparative experiment on the test data. The objective of this experiment was not only to simply verify the feasibility of DL models for characterizing the relationship between input parameters and backscattering coefficients from a crop such as soybean but also to utilize the test dataset to estimate the performance of the two pre-trained networks. For S-band, the correlation plot shows that the model predictions match well with the simulated values for both pre-trained models, with the RMSE reaching the values of 1.63 dB and 0.87 dB, respectively. In the case of C-band, deviation between prediction and simulation becomes apparent for PT-DNN, with an overall RMSE value of 2.01 dB, while PT-LSTM can still yield a good performance, with the RMSE reaching the value of 0.54 dB. We can conclude that the LSTM network is more powerful than DNN. That is to say, in addition to considering the nonlinear relationship among input parameters and backscattering coefficients, the interrelationship between input sequences is equally significant.

Secondly, we constructed a fine-tuned dataset to verify the effectiveness of the fine-tuning strategy. This dataset only comprised 36 and 32 sets of the measured data for S- and C-bands, respectively. It was, therefore, well-suited for evaluating the effectiveness of the fine-tuning strategy. The mean validation loss indicates that, during the fine-tuning phase, the loss curves for S-band tend to drop dramatically compared to those for C-band and remain relatively smooth as the fine-tuning samples proportionally increase. With respect to the improvement ratio, the fine-tuning strategy significantly reduces the RMSE on the validation set. Furthermore, compared with in situ measurement, fine-tuning was proved as a robust calibration strategy.

Thirdly, we verified the performance of the proposed TL approach with the S- and C-bands' experimental data in terms of quantitative and visual. The computed error indices reveal that the predictions of the proposed method yield excellent agreement with

the experimental data under real conditions. In particular, PT-LSTM performed better than those for DNN, thus indicating that LSTM can not only extract the interrelationships between input parameters and vegetation backscatters but also the performance is even better than the SS-RTM under real conditions. Similarly, in the visualized plot, FT-LSTM yields the best performance, with the lowest RMSE value of 2.32 dB for S-band and 0.99 dB for C-band, respectively.

The above-mentioned results revealed that the proposed TL-based approach outperforms the SS-RTM and other DL schemes, in terms of accuracy and robustness, demonstrating the good capabilities of such an approach in estimating vegetation backscatters. All these outcomes will serve a new path for further improving the TL-based method and broaden its applicability for other types of vegetation canopies. It is worth mentioning at this point that we only took into account capability in the soybean fields for this paper, due to the limited radar data available to us. More importantly, the TL-based method can be used for other vegetation canopies, such as wheat, corn, and carrot. The only required change is that one needs to build the fine-tuned dataset corresponding to the specific vegetation and then continue the fine-tuning process. Moreover, such a novel idea also can be guided to address complex microwave scattering problems.

References

1. Liang, P.; Pierce, L.E.; Moghaddam, M. Radiative transfer model for microwave bistatic scattering from forest canopies. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 2470–2483. [CrossRef]
2. Fontanelli, G.; Montomoli, F.; Azar, R.; Macelloni, G.; Villa, P. Assessing interactions between crop biophysical parameters and X-band backscattering using empirical data and model sensitivity analysis. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4405612. [CrossRef]
3. Attema, E.P.W.; Ulaby, F.T. Vegetation modeled as a water cloud. *Radio Sci.* **1978**, *13*, 357–364. [CrossRef]
4. Ulaby, F.T.; Sarabandi, K.; McDonald, K.; Whit, M.; Dobson, M.C. Michigan microwave canopy scattering model. *Int. J. Remote Sens.* **1990**, *11*, 1223–1253. [CrossRef]
5. Su, X.; Wu, J.; Huang, B.; Wu, Z. GPU-Accelerated computation for electromagnetic scattering of a double-layer vegetation model. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2013**, *6*, 1799–1806. [CrossRef]
6. Whitman, G.M.; Wu, M.Y.C.; Schwering, F.K. Propagation and scattering of spherical wave pulses in vegetation using scalar transport theory. *IEEE Trans. Antennas Propag.* **2010**, *58*, 1662–1676. [CrossRef]

7. Kweon, S.K.; Oh, Y. A modified water-cloud model with leaf angle parameters for microwave backscattering from agricultural fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2802–2809. [CrossRef]

8. Ardekani, M.R.; Jacques, D.C.; Lambot, S. A layered vegetation model for GPR full-wave inversion. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 18–28. [CrossRef]

9. Huertero, A.M.; Liu, P.; Judge, J. Phenology-based backscattering model for corn at L-band. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4989–5005. [CrossRef]

10. Sharma, A.; Lang, R.H.; Kurum, M.; O'neill, P.E.; Cosh, M.H. L-band radar experiment and modeling of a corn canopy over a full growing season. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 5821–5835. [CrossRef]

11. Mandal, D.; Kumar, V.; Lopez-Sanchez, J.M.; Bhattacharya, A.; McNairn, H.; Rao, Y.S. Crop biophysical parameter retrieval from Sentinel-1 SAR data with a multi-target inversion of Water Cloud Model. *Int. J. Remote Sens.* **2020**, *41*, 5503–5524. [CrossRef]

12. Meng, X.; Wei, Q.; Dong, C.; Guo, L.; Chen, Y. A composite scattering model of the vegetated ground with a target. *IEEE Antennas Wirel. Propag. Lett.* **2023**, *22*, 1587–1591. [CrossRef]

13. Ulaby, F.T.; Long, D.G. *Microwave Radar and Radiometric Remote Sensing*; The University of Michigan Press: Ann Arbor, MI, USA, 2014.

14. Zhu, D.; Zhao, P.; Zhao, Q.; Li, Q.L.; Zhang, Y.S.; Yang, L.X. A two-stream LSTM-based backscattering model at L-band and S-band for dry soil surfaces under large roughness conditions. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2024**, *17*, 3137–3150. [CrossRef]

15. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Ana. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef]

16. 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 (MICCAI), Munich, Germany, 5–9 October 2015; pp. 234–241.

17. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder–decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 833–851.

18. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef] [PubMed]

19. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

20. 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 (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.

21. Tan, M.X.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional neural network. *arXiv* **2017**, arXiv:1905.11946.

22. Zhang, J.; Li, Y.; Tong, Z.; He, L.; Zhang, M.; Niu, Z.; He, H. GLCANet: Global–local context aggregation network for cropland segmentation from multi-source remote sensing images. *Remote Sens.* **2024**, *16*, 4627. [CrossRef]

23. Amin, A.; Kamilaris, A.; Karatsiolis, S. A weakly supervised multimodal deep learning approach for large-scale tree classification: A case study in Cyprus. *Remote Sens.* **2024**, *16*, 4611. [CrossRef]

24. Wang, Z.; Zhao, L.; Meng, J.; Han, Y.; Li, X.; Jiang, R.; Chen, J.; Li, H. Deep learning-based cloud detection for optical remote sensing images: A Survey. *Remote Sens.* **2024**, *16*, 4583. [CrossRef]

25. Li, J.; Cheng, S. AFENet: An attention-focused feature enhancement network for the efficient semantic segmentation of remote sensing images. *Remote Sens.* **2024**, *16*, 4392. [CrossRef]

26. Li, Y.; Wang, X.; Zhang, C.; Zhang, Z.; Ren, F. High-fidelity infrared remote sensing image generation method coupled with the global radiation scattering mechanism and Pix2PixGAN. *Remote Sens.* **2024**, *16*, 4350. [CrossRef]

27. Paolanti, M.; Tiribelli, S.; Giovanola, B.; Mancini, A.; Frontoni, E.; Pierdicca, R. Ethical framework to assess and quantify the trustworthiness of artificial intelligence techniques: Application case in remote sensing. *Remote Sens.* **2024**, *16*, 4529. [CrossRef]

28. Li, L.; Wang, L.G.; Teixeira, F.L.; Liu, C.; Nehorai, A.; Cui, T.J. DeepNIS: Deep neural network for nonlinear electromagnetic inverse scattering. *IEEE Trans. Antennas Propag.* **2019**, *67*, 1819–1825. [CrossRef]

29. Xu, K.; Wu, L.; Ye, X.; Chen, X. Deep learning-based inversion methods for solving inverse scattering problems with phaseless data. *IEEE Trans. Antennas Propag.* **2020**, *68*, 7457–7470. [CrossRef]

30. Zhou, Y.; Zhong, Y.; Wei, Z.; Yin, T.; Chen, X. An improved deep learning scheme for solving 2-D and 3-D inverse scattering problems. *IEEE Trans. Antennas Propag.* **2021**, *69*, 2853–2863. [CrossRef]

31. Zhu, D.; Zhao, Q.; Yang, L.; Bo, Y.; Chen, W. Multi-model fusion approach for electromagnetic inverse scattering problems. *IET Microw. Antennas Propag.* **2022**, *16*, 610–616. [CrossRef]

32. Zhang, H.H.; Yao, H.M.; Jiang, L.; Ng, M. Enhanced two-step deep-learning approach for electromagnetic-inverse-scattering problems: Frequency extrapolation and scatterer reconstruction. *IEEE Trans. Antennas Propag.* **2023**, *71*, 1662–1672. [CrossRef]

33. Zhang, H.; Chen, Y.; Cui, T.J.; Teixeira, F.L.; Li, L. Probabilistic deep learning solutions to electromagnetic inverse scattering problems using conditional renormalization group flow. *IEEE Trans. Microw. Theory Tech.* **2022**, *70*, 4955–4965. [CrossRef]

34. Hu, Y.-D.; Wang, X.-H.; Zhou, H.; Wang, L. A priori knowledge-based physics-informed neural networks for electromagnetic inverse scattering. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5613109. [CrossRef]

35. Chen, H.; Yang, C.; Du, Y. Machine learning-assisted analysis of polarimetric scattering from cylindrical components of vegetation. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 155–165. [CrossRef]

36. Xiao, D.; Guo, L.; Liu, W.; Hou, M. Efficient RCS prediction of the conducting target based on physics-inspired machine learning and experimental design. *IEEE Trans. Antennas Propag.* **2021**, *69*, 2274–2289. [CrossRef]

37. Cao, J.; Ren, Q.; Dang, X.; Hou, Z.; Yan, H.; Li, L.; Yin, H. Efficient scattering center prediction method for targets with coating defects through deep learning. *IEEE Trans. Microw. Theory Tech.* **2022**, *70*, 5319–5331. [CrossRef]

38. Xue, B.W.; Guo, R.; Li, M.K.; Sun, S.; Pan, X.-M. Deep-learning-equipped iterative solution of electromagnetic scattering from dielectric objects. *IEEE Trans. Antennas Propag.* **2023**, *71*, 5954–5966. [CrossRef]

39. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]

40. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks. *arXiv* **2014**, arXiv:1411.1792.

41. Li, R.; Zhang, Y.; Tang, H.; Shi, Z.; Jiao, Y.; Xiao, L.; Wei, B.; Gong, S. Research on electromagnetic scattering and inverse scattering of target based on transfer learning physics-informed neural networks. In Proceedings of the 2024 International Applied Computational Electromagnetics Society Symposium, Xi'an, China, 16–19 August 2024; pp. 1–3.

42. Selver, M.A.; Toprak, T.; Secmen, M.; Zoral, E.Y. Utilizing resonant scattering signal characteristics of magnetic spheres via deep learning for improved target classification. In Proceedings of the International Symposium on Advanced Electrical and Communication Technologies, Rome, Italy, 27–29 November 2019; pp. 1–5.

43. Jin, Y.; Zi, Y.; Hu, W.; Hu, Y.; Wu, X.; Chen, J. Solving full waveform inversion enhanced by efficient progressive transfer learning. In Proceedings of the 2022 United States National Committee of URSI National Radio Science Meeting, Boulder, CO, USA, 4–8 January 2022; pp. 40–41.

44. Zhu, Q.; Xu, C.; Zhao, S.; Tao, X.; Zhang, Y.; Tao, H.; Wang, X.; Fang, Y. A space object optical scattering characteristics analysis model based on augmented implicit neural representation. *Remote Sens.* **2024**, *16*, 3316. [CrossRef]

45. Dai, R.; Xu, Y.; Qiu, J.; Ma, H.; Li, E. High-fidelity S-parameter prediction using transfer learning based encoder-decoder model. In Proceedings of the 2024 14th International Workshop on the Electromagnetic Compatibility of Integrated Circuits, Torino, Italy, 7–9 October 2024; pp. 80–83.

46. Yu, R.; Li, S.; Zhang, B.; Zhang, H. A Deep Transfer Learning Method for Estimating Fractional Vegetation Cover of Sentinel-2 Multispectral Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 6005605. [CrossRef]

47. Hallikainen, M.T.; Ulaby, F.T.; Dobson, M.C.; El-rayes, M.A.; Wu, L.-K. Microwave dielectric behavior of wet soil—Part II: Dielectric mixing models. *IEEE Trans. Geosci. Remote Sens.* **1985**, *GE-23*, 35–46. [CrossRef]

48. Ulaby, F.T.; El-rayes, M.A. Microwave dielectric spectrum of vegetation—Part II: Dual-dispersion model. *IEEE Trans. Geosci. Remote Sens.* **1987**, *GE-25*, 550–557. [CrossRef]

49. Yueh, S.H.; Kong, J.A.; Jao, J.K.; Shin, R.T.; Le Toan, T. Branching model for vegetation. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 390–402. [CrossRef]

50. Wigneron, J.-P.; Ferrazzoli, P.; Calvet, J.-C.; Bertuzzi, P. A parametric study on passive and active microwave observations over a soybean crop. *IEEE Trans. Geosci. Remote Sens.* **1999**, *37*, 2728–2733. [CrossRef]

51. Chen, K.S.; Wu, T.-D.; Tsang, L.; Li, Q.; Shi, J.; Fung, A.K. Emission of rough surfaces calculated by the integral equation method with comparison to three-dimensional moment method simulations. *IEEE Trans. Geosci. Remote Sens.* **2003**, *41*, 90–101. [CrossRef]

52. Fung, A.K.; Chen, K.-S. *Microwave Scattering and Emission Models for Users*; Artech House: Norwood, MA, USA, 2010.

53. Del Frate, F.; Ferrazzoli, P.; Schiavon, G. Retrieving soil moisture and agricultural variables by microwave radiometry using neural networks. *Remote Sens. Environ.* **2003**, *84*, 174–183. [CrossRef]

54. Lu, C.; Wang, Z.; Wu, Z.; Zheng, Y.; Liu, Y. Global ocean wind speed retrieval from GNSS reflectometry using CNN-LSTM network. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5801112. [CrossRef]

55. Gadiraju, K.K.; Vatsavai, R.R. Remote sensing based crop type classification via deep transfer learning. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2023**, *16*, 4699–4712. [CrossRef]

56. Shastry, A.; Carter, E.; Coltin, B.; Sleeter, R.; McMicheal, S.; Eggleston, J. Mapping floods from remote sensing data and quantifying the effects of surface obstruction by clouds and vegetation. *Remote Sens. Environ.* **2023**, *291*, 113556. [CrossRef]

57. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2017**, arXiv:1609.04747.

58. Huber, P.J.; Ronchetti, E.M. *Robust Statistics*; Wiley: New York, NY, USA, 2009.

59. Wang, Y.; He, Z.; Yang, Y.; Ding, D.Z.; Ding, F.; Dang, X.W. Multi-parameter inversion of AIEM by using bi-directional deep neural network. *Remote Sens.* **2022**, *14*, 3302. [CrossRef]

60. He, X.; Chen, Y. Transferring CNN ensemble for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2021**, *18*, 876–880. [CrossRef]

61. Zhai, Y.; Deng, W.; Lan, T.; Sun, B.; Ying, Z.; Gan, J.; Mai, C.; Li, J.; Labati, R.D.; Piuri, V.; et al. MFFA-SARNET: Deep transferred multi-level feature fusion attention network with dual optimized loss for small-sample SAR ATR. *Remote Sens.* **2020**, *12*, 1385. [CrossRef]

62. Cheng, L.; Singh, P.; Ferranti, F. Transfer learning-assisted inverse modeling in nanophotonics based on mixture density networks. *IEEE Access* **2024**, *12*, 55218–55224. [CrossRef]

63. Niu, S.; Liu, M.; Liu, Y.; Wang, J.; Song, H. Distant domain transfer learning for medical imaging. *IEEE J. Biomed. Health Inform.* **2010**, *25*, 3784–3793. [CrossRef] [PubMed]

64. Li, M.; Wang, J.; Chen, Y.; Tang, Y.; Wu, Z.; Qi, Y.; Jiang, H. Low-dose CT image synthesis for domain adaptation imaging using a generative adversarial network with noise encoding transfer learning. *IEEE Trans. Med. Imaging* **2023**, *42*, 2616–2630. [CrossRef] [PubMed]

65. Dao, H.; Nguyen, T.; Mugisha, C.; Paik, I. A multimodal transfer learning approach using PubMedCLIP for medical image classification. *IEEE Access* **2024**, *12*, 75496–75507. [CrossRef]

66. Shao, W.; Sun, H.; Wang, Y.; Qing, X. Damage Quantification Method for Aircraft Structures Based on Multitask CNN-LSTM and Transfer Learning. *IEEE Sens. J.* **2024**, *24*, 9217–9228. [CrossRef]

67. Zhang, B.; Huang, R.; Luo, Z.; Huang, L.; Hong, X.; Jin, G. Ring array scatter feature adaption deep transfer imaging method for composite plate structure health monitoring using guided waves. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 3533610. [CrossRef]