Crop Mapping Using Sentinel Full-Year Dual-Polarized SAR Data and a CPU-Optimized Convolutional Neural Network With Two Sampling Strategies

Shengjie Liu, *Member, IEEE*, Zhize Zhou, Huaxiang Ding, Yuanjun Zhong, and Qian Shi, *Senior Member, IEEE*

***Abstract*—Although optical remote sensing can capture the Earth's environment with visible and infrared sensors, it is limited by weather conditions. Often, only a few sets of cloud-free optical imagery are available in cloudy regions, where many agricultural towns are located. On the other hand, radar remote sensing can capture imagery under cloudy conditions. In this study, we examined the capability of Sentinel-1 multitemporal dual-polarized synthetic aperture radar (SAR) imagery in a whole year from Google Earth Engine in crop mapping in two study sites in Chongqing, China, and Landivisiau, France. Results show that it is possible to produce better crop classification maps using multitemporal SAR imagery, but the performance is limited by local terrain. Flat agricultural regions, such as Western Europe, are expected to benefit from the multitemporal SAR information. Mountain agricultural regions, such as Southwestern China, will encounter difficulties due to the undulate terrain. We also tested two sampling strategies, i.e., random sampling and regional sampling, and observed high variation in overall accuracy: the former led to a higher accuracy. The gap is caused by the diversity of training sets examined using tSNE visualization. The importance of SAR channels in each month are correlated with their entropy. Data from the growing season are important in distinguishing crop types. The 3-D convolutional neural network (CNN) achieved similar results under a huge computation cost compared with 2-D CNNs. Based on the experiments, we recommend to use a lightweight 2-D CNN that can run on the CPU for real-world crop mapping with SAR data.**

***Index Terms*—Convolutional neural network (CNN), crop mapping, multitemporal, radar remote sensing, sampling strategy, synthetic aperture radar (SAR).**

Manuscript received February 9, 2021; revised April 13, 2021 and June 11, 2021; accepted June 27, 2021. Date of publication July 7, 2021; date of current version July 23, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61976234 and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515011057. *(Corresponding authors: Huaxiang Ding; Yuanjun Zhong.)*

Shengjie Liu is with the Department of Physics, The University of Hong Kong, Pokfulam, Hong Kong (e-mail: sjliu.me@gmail.com).

Zhize Zhou is with the State Key Laboratory of Computer-Aided Design and Computer Graphics, Zhejiang University, Hangzhou 310058, China (e-mail: zhouzhize@zju.edu.cn).

Huaxiang Ding and Yuanjun Zhong are with the Institute of Surveying and Mapping, Department of Natural Resources of Guangdong Province, Guangzhou 510500, China (e-mail: wydinghx@163.com; gdzyj1222@163.com).

Qian Shi is with the School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: shixi5@mail.sysu.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/JSTARS.2021.3094973, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2021.3094973

I. Introduction

With more than one billion people at the brink of starvation, the United Nations has announced zero hunger as one of the sustainable development goals [1]. To achieve the goal, timely and accurate monitoring of food production is essential. Traditionally, governments rely on specialists to collect food production data, which is time-consuming and expensive. Thanks to the development of remote sensing technologies, we can now estimate food production from satellite imagery with minimal laborious works.

Optical remote sensing relies on the visible and infrared information to distinguish crop types. This technique has shown its great potential in crop mapping. For example, Wardlow and Egbert [2] used time-series MODIS NDVI data to create crop type maps with a hierarchical classification approach. Kussul *et al.* [3] conducted crop classification in Kyiv, Ukraine with multitemporal Landsat-8 data using an MLP model. Shelestov *et al.* [4] explored the efficiency of using the Google Earth Engine platform to create large-area crop mapping in Ukraine using multitemporal Landsat-8 and Sentinel-2 imagery. Pan *et al.* [5] derived the NDVI time series from China's HJ-1 A/B satellite data to map small-scale crop seasonality. Gao *et al.* [6] generated 30-m crop phrenological metrics using fusion of Landsat and MODIS data. Skakun *et al.* [7] combined Landsat-8 and Sentinel-2A images to create winter crop maps and estimated wheat yield. Wang *et al.* [8] used Fourier transform of Landsat time-series images to distinguish crop types and achieved over 80% overall accuracy without in-season field data in some regions using a random forest transfer technique. De Castro *et al.* [9] developed an automatic random forest with object-based image analysis for weed mapping using UAV imagery. With the literature going deep, crop type classification is generated from low-resolution satellite images to from high-resolution satellite and even UAV images, but most of the studies focus on the usage of random forest because of its robustness [10].

Although multitemporal optical remote sensing has shown great potentials in crop mapping, it is limited by weather conditions. In the rainy season, it cannot capture the ground information because of the existence of dense clouds. Instead, radar remote sensing is insensitive to clouds due to its long wavelength nature. Radar remote sensing can penetrate heavy clouds and

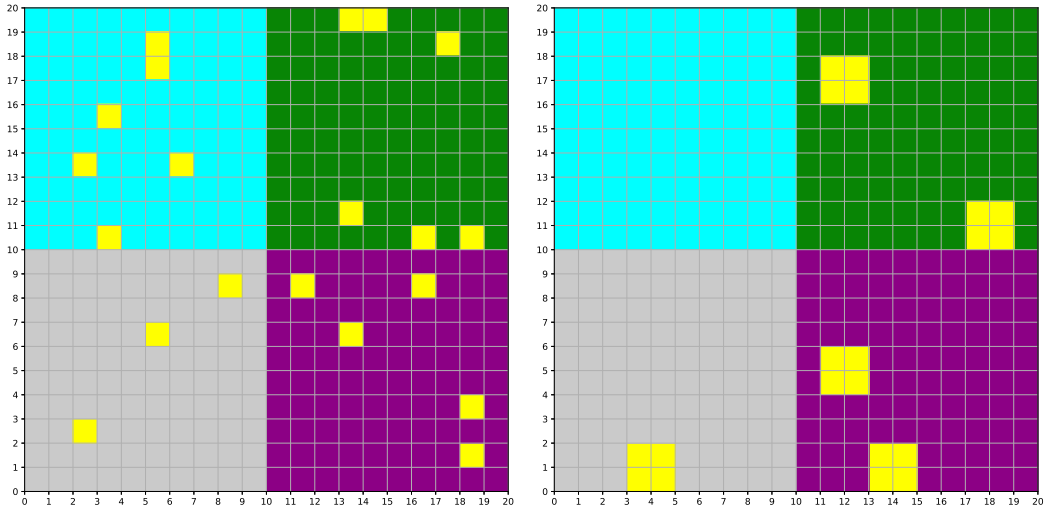
sense the solid ground, where the crops grow. Hence, it provides us an opportunity to monitor crop areas in cloudy regions in the rainy season. For instance, Le Toan *et al.* [11] assessed the use of ERS-1 synthetic aperture radar (SAR) data to map rice-growing areas and to retrieve rice parameters at a tropical site in Indonesia and a temperate site in Japan. Jiao *et al.* [12] assessed the use of object-oriented classification of 19 sets of multitemporal RADARSAT-2 PolSAR imagery in Canada. In a nonmachine learning fashion, Chen *et al.* [13] discriminated crop types based on polarimetric correlation coefficients of PolSAR data. Furthermore, a particle-swarm-optimized kernel-based clustering method was developed using multitemporal L-band PolSAR data for crop mapping [14]. Skriver *et al.* [15] assessed the performance of the airborne AgriSAR 2006 data in crop classification and concluded that multitemporal information was important in distinguishing, for both single- and dual-polarized modes. The tradeoff between polarimetric information and the multitemporal information was further discussed in a crop mapping study using the SAR data from the EMISAR system, and multitemporal information was again confirmed to be more important in crop mapping [16]. Bargiel [17] proposed a novel phenological sequence pattern approach using dense time series of Sentinel-1 images, which outperformed standard methods such as random forest and maximum likelihood in distinguishing cereal crops, oat, winter barley, and rye. A parcel-based method was tested with multitemporal TerraSAR-X dual-polarimetric data using four machine learning algorithms and two kernel-based methods, where the multiple kernel learning achieved the highest overall accuracy of 92.1% [18]. Using six spotlight TerraSAR-X SAR images, Busquier *et al.* [19] concluded that an increase of up to 10% in overall accuracy could be achieved with the addition of coherent copolar polarimetry. A recent study by Li *et al.* [20] has made use of the full-year L-band UAVSAR time series for crop mapping with the random forest method. The obtained result is promising. However, UAVSAR time series is rare in many regions, and more studies should be conducted on public popular datasets, e.g., the spaceborne Sentinel-1 data.

The literature works about multitemporal SAR information for crop mapping are extensive, but most of them used either rule-based methods or tree-based supervised learning (e.g., random forest), and few explored the usage of deep learning methods. Tomppo *et al.* [21] applied an improved k nearest neighbor (kNN) algorithm on 17 Sentinel-1 multitemporal scenes and achieved a 72% overall accuracy on crop mapping in Eura, Finland. Xiao and Lu [22] tested a subspace kNN and a bagged tree method using 26 SAR images for crop mapping in Bengbu, China, and found that the subspace kNN could achieve overall accuracy of 97%. Arias *et al.* [23] used a rule-based supervised method and 15-month Sentinel-1 SAR images to classify 14 crop types in Navarre, Spain, and concluded that the field size of croplands significantly affected the classification performance (14%), where a smaller field was more difficult to classify ($<$0.5 ha). Small croplands are signature in China in the form of terraces [24], leading to a difficult crop mapping scenario. Apart from backscatter, multitemporal InSAR coherence was examined using multiple feature-based classifiers (e.g., random forest, SVM, and kNN) by Jacob *et al.* [25]. A main conclusion was that classification could achieve over 90% accuracy in flat area such as Doñana wetlands National Park in Spain, but only achieved 77% accuracy in challenging alpine terrain such as the northern Italy. Full-year time-series interferometric coherence and backscatter for crop mapping were also examined using random forest by Mestre-Quereda *et al.* [26], where the joint use of coherence and backscattering coefficient benefited classification the most. Random forest was found the most-often-used classifier for multitemporal SAR crop mapping. However, it was rare to include spatial information [either with spatial filters or convolutional neural networks (CNNs)] when the multitemporal data were used, though they have already been proven to be powerful tools to increase classification accuracy.

Deep learning pixelwise classification, or patch-based classification, opens a small window from the pixel to be classified to extract spatial information. Although this process saves us the time to manually design spatial features, such as the GLCM textures [27], the local binary patterns [28], and the Gobar filters [29], criticisms have been raised about whether this process leads to overestimating the classification accuracy [30]–[36]. The main argument is that some of the reference data are leaked when we open a small window [35]. In real-world applications, the reference data are often collected in a parcel (a small region). For example, farmland might be 50×50 m^2 large; on a standard Sentinel-2 scene with 10-m ground sampling distance, we will collect a small region of 25 pixels as the reference data. As a result, collected samples are not randomly uniformly distributed in the agricultural region. This is contradictory to many literature works developing state-of-the-art CNNs for land cover mapping with SAR and hyperspectral data. Many advanced CNNs are developed with the random sampling strategy, instead of using the above regional sampling strategy. In these literature works, the available training and testing samples are randomly uniformly distributed in the entire study area or called from the same domain in technical terms [37], [38].

We show a toy example in Fig. 1 illustrating the difference of the two strategies. The Earth environment is diverse in different regions; in a satellite image, a region on the upper left may be different from a region on the lower right in terms of atmosphere, terrain, illumination, and satellite angle. As a result, two identical objects on the ground can have two different spectral profiles in an image. The variation inside a land cover class is better captured in random sampling than regional sampling, as the former covers a more diverse region. From the perspective of supervised machine learning, random sampling has a better estimation of the data's real distribution [39], [40]. If the same number of training samples is used, the sample set from random sampling will be more diverse than that from regional sampling [41], which may be one of the reasons why random sampling leads to higher classification accuracy. However, the problem is: how would this affect real-world crop mapping using multitemporal SAR data?

In this study, we are going to use multitemporal SAR data from a whole year to produce crop maps, with both random sampling training set and regional sampling training set. We select three advanced CNNs, namely, the wide contextual residual network (WCRN) [42], the HResNet [43], and the double-branch



(a) (b)

Fig. 1. Top example about the random sampling and regional sampling strategies. The example here includes four domains, and the same materials may have different radar signals and spectral responses due to the effects of atmosphere, terrain, illumination, satellite angle, etc. Selected training samples are highlighted in yellow. Although the two strategies obtained the same number of training samples, regional sampling fails to collect enough samples from the blue region. Compared to random sampling, the training set's diversity via regional sampling will be significantly lower, thereby limiting the performance of classifiers. (a) Random sampling. (b) Regional sampling.

multiattention mechanism (DBMA) network [44], and random forest [45] to test their performance. Among the three deep learning models, the WCRN can run on CPU, and the DBMA has the best performance in benchmark datasets with random sampling strategy. The WCRN is the representative of 1×1 convolutional 2-D CNNs, inspired by the 2-D deep contextual CNN [46]. The HResNet is the representative of ordinary 2-D CNNs (the most common version). DBMA is the representative of 3-D CNNs [47], [48]. Two study sites are selected: one in China and one in France. Based on the experiments, we are going to answer the following questions at the end of this article.

- Can multitemporal SAR data classify crop types as good as optical data?
- Will the two sampling strategies, i.e., random sampling and region sampling, affect the classification, and if so, what are the effects?
- Which classifier is the best for crop mapping using multitemporal SAR data?

The rest of this article is organized as follows. In Section II, we introduce the two study sites with the collected reference data and the multitemporal SAR data. In Section III, we describe the three CNNs and random forest in detail. The results and analysis are shown in Section IV. Finally, Section V concludes this article.

II. Study Area and Data

A. Chongqing, China

The first study site is located in the Jiangjin District, Chongqing, the Mountain City of China, as shown in Fig. 2. Chongqing lies in Sichuan Basin, southwestern China, and is one of the cloudiest cities in the world. It receives only about 1200 sunshine hours per year [49]. As a comparison, the northern China received about 2500 sunshine hours per year [50]. The study site is in the mountains, where the elevation ranges from 178 to 1709 m. The entire district has a population of 1.3 million and is one of the most active agricultural regions in China, famous for its massive production of Sichuan pepper.

TABLE I
Number of Training and Testing Samples of the Chongqing Data

Index	Name	#Train	#Test
1	Paddy	200	1943
2	Corn	200	867
3	Sichuan Pepper	200	1858
4	Orange Trees	200	921
5	Peanuts	32	101
6	Lotus	187	341
7	Buildings	200	3512
8	Water	197	1223
9	Other Plants	159	625
10	Barren	29	200

For this study site, the reference data were collected in a field survey on July 5, 2019, along the road network. Six crop types were recognized, plus four land cover classes to complete the classification system. The number of training and testing samples is shown in Table I. Photos of some crop types recorded in the field survey are shown in Fig. 3. In regional sampling, we randomly split the collected polygons into two parts: training and testing. The maximum number of training samples per class is limited to 200 to avoid imbalanced training. In random sampling, the same amount of training and testing data were selected randomly from all reference data to ensure a fair comparison between two sampling strategies.

The multitemporal SAR data were extracted from Google Earth Engine. Google Earth Engine restores a multipetabyte catalog of satellite imagery, including the Sentinel-1 images,

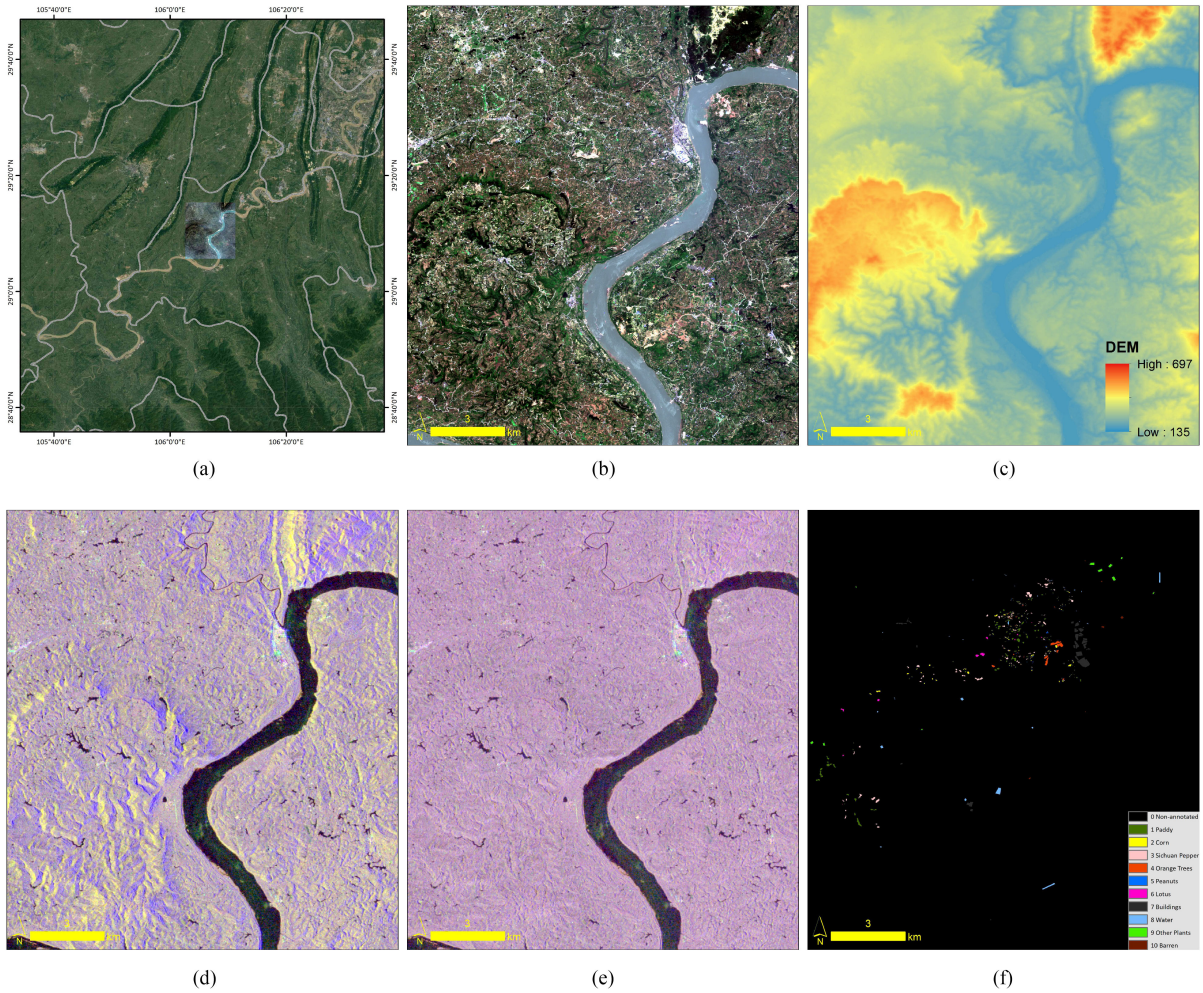


Fig. 2. Chongqing, China data. (a) Jiangjin, Chongqing, China. (b) Optical data (RGB: band 4, band 3, and band 2 of July). (c) DEM. (d) SAR data (RGB: ascending VH, ascending VV, and descending VH of July before terrain correction). (e) SAR data (RGB: ascending VH, ascending VV, and descending VH of July after terrain correction). (f) Reference data.

making it easy to use SAR imagery. By using the ready-to-use data available on Google Earth Engine, we can save the preprocessing time to handle raw SAR data: applying orbit file, noise removal, radiometric calibration, terrain correction etc. Here, we downloaded the monthly SAR composite from Google Earth Engine. Both ascending and descending images are used with dual polarization, i.e., VV and VH, resulting in a total number of 48 feature channels. Note that the Sentinel-1 SAR imagery was not despeckled by default in Google Earth Engine, which was the default setting in our experiment. We conducted an additional set of comparison to test the influence of the Refined Lee filter later. As a result, the SAR images in the main experiment did not contain information from the surrounding pixels. We collected two available cloudless scenes of Sentinel-2 multispectral data in this region for comparison purposes.

B. *Landivisiau, France*

The second study site is located in Landivisiau, northwestern France, as shown in Fig. 4. The elevation ranges from −8 to 131 m, a flat region compared to the Chinese site. The northwestern part is the cloudiest region in France. For example, Tours, a city between the Landivisiau and Paris, has about 1800 sunshine hours per year [51].

The reference data were collected by the National Institute of Forest and Geography Information in France based on farmers' self-report under the Common Agricultural Policy of the European Union in 2017 [52]. A total of 23 classes were recorded, where we selected 12 classes after adding the artificial class (impervious) and discarding small classes. The number of training and testing data samples is shown in Table II. We randomly split the collected polygons into two subsets: one as the training set and the other as the testing set. The maximum of training samples is limited to 10 000 to balance the samples, which were randomly chosen during training. The above is regional sampling. For random sampling, the same amount of training and testing data was selected from all reference data randomly.

Similar to the Chongqing data, a total of 48 SAR channels were collected to train the model. For comparison purposes, we collected seven available cloudless scenes of multispectral data from the Google Earth Engine.

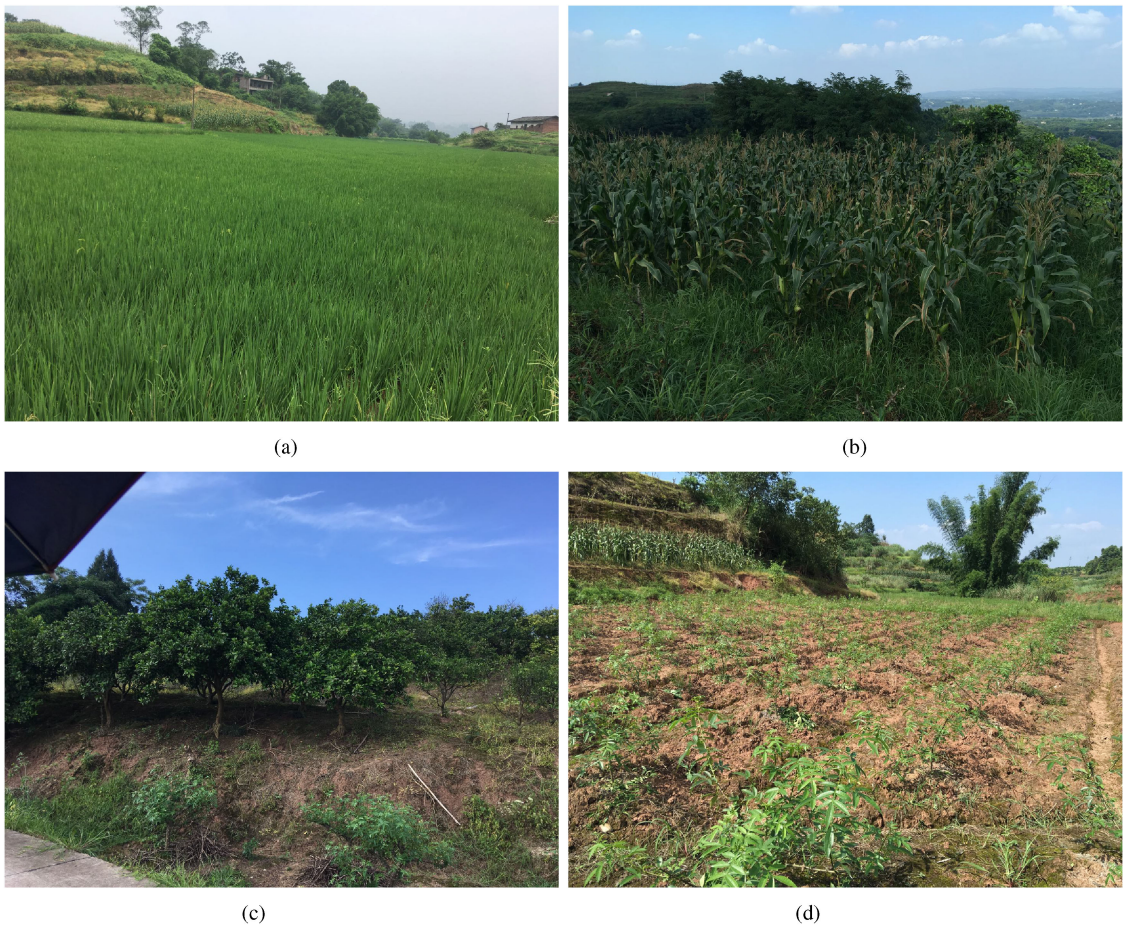


Fig. 3. Photos from site visit. (a) Paddy. (b) Corn. (c) Orange trees. (d) Sichuan pepper.

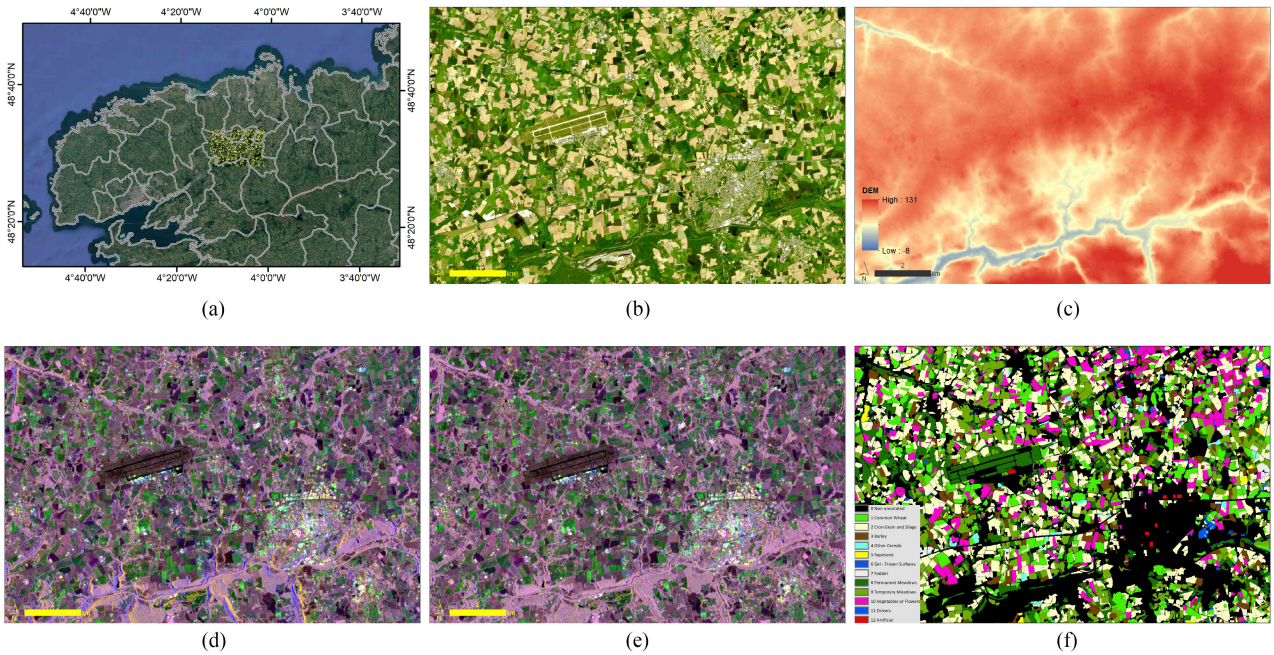


Fig. 4. Landivisiau, France data. (a) Landivisiau, France. (b) Optical data (RGB: band 4, band 3, and band 2 of May). (c) DEM. (d) SAR Data (RGB: ascending VH, ascending VV, and descending VH of May before terrain correction). (e) SAR data (RGB: ascending VH, ascending VV, and descending VH of May after terrain correction). (f) Reference data.

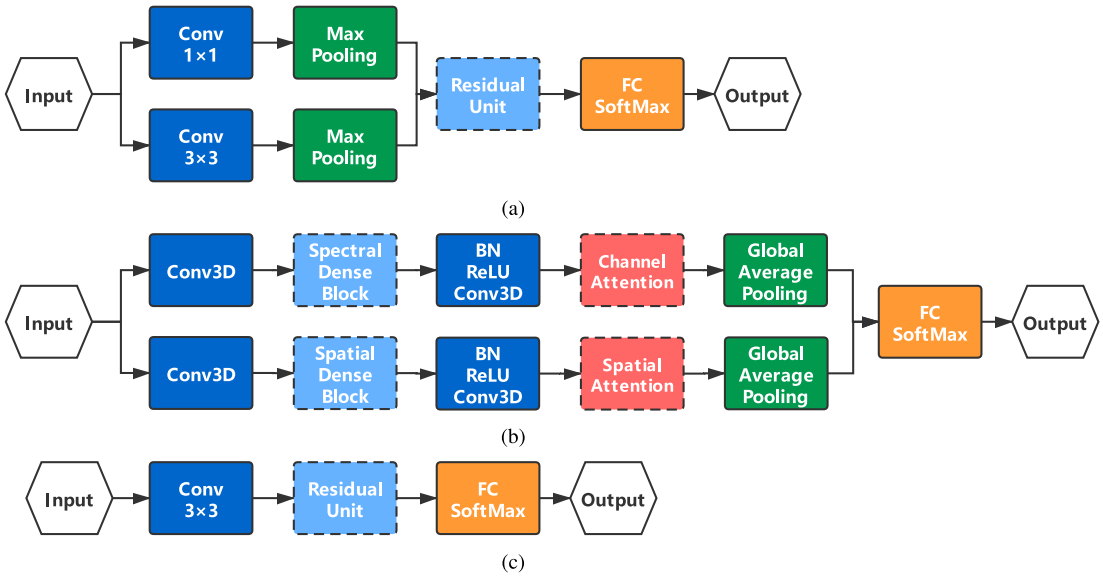


Fig. 5. Three CNNs tested for crop mapping. More details of the implementation regarding the number of kernels can be found in the Supplementary Material. (a) WCRN. (b) DBMA. (c) HResNet.

TABLE II
NUMBER OF TRAINING AND TESTING SAMPLES OF THE LANDIVISIAU DATA

Index	Name	#Train	#Test
1	Common Wheat	10000	73241
2	Corn Grain and Silage	10000	177815
3	Barley	10000	47640
4	Other Cereals	3342	4864
5	Rapeseed	3114	1656
6	Gel - Frozen Surfaces	960	1462
7	Fodder	7826	9310
8	Permanent Meadows	10000	84251
9	Temporary Meadows	10000	103882
10	Vegetables or Flowers	10000	67253
11	Drivers	2931	5456
12	Artificial	842	484

III. METHODOLOGY

A. Architectures of the Three CNNs

In this section, we first go through the three CNNs used in this study in general terms. We then illustrate in detail about their components in the following subsections. The three CNNs, i.e., the WCRN, the HResNet, and the DBMA, are tested in the experiments. The architectures of these CNNs are shown in Fig. 5. For the WCRN, the head of the network is a multiscale 2-D convolution to extract spatial and spectral information. The extracted features go through a max pooling layer to reduce the spatial dimension to one, followed by a residual unit with 1×1 convolutional layer. Finally, the fused features are fed into a fully connected (FC) layer with the SoftMax function to give the probability to each class.

The DBMA network contains two parallel parts: spectral and spatial. Each part includes a 3-D convolutional layer, a dense block focusing on the spectral or spatial domain, a dimension reduction component, a channelwise (spectral) or spatial attention unit, and a global average pooling layer. The features from the spectral and spatial branches are then concatenated together and fed into an FC+SoftMax layer.

As for the HResNet, the input 9×9 features first go through a 2-D 3×3 convolutional layer with 64 kernels. The output $7 \times 7 \times 64$ features are then fed into a residual unit with 3×3 convolutions. The extracted features are fed to the FC+SoftMax layer. Details of the three networks' implementation are shown in the Supplementary Material and will be discussed in the following sections.

B. Deep Network Elements

1) Convolution: The fundamental component of a CNN is the convolutional layer. A convolutional layer is capable of extracting high-level spatial features, such as the frequency and edge features, using local connections [53]. In a deep CNN, the output $y_{i,j}$ of a feature map of a convolutional layer at position (i, j) can be expressed as

$$y_{i,j} = b + \sum_{m=1}^{M} \sum_{p=1}^{P} \sum_{q=1}^{Q} w_{p,q,m} \times x_{i+p,j+q,m} \tag{1}$$

where P and Q are the height and width of the convolutional layer, M is the number of feature maps, and $w_{p,q,m}$ and b are the weight at (p, q) of the convolutional filter and the bias connected to the mth feature map. The complexity of a CNN can be modeled from two aspects, i.e., the number of parameters and the number of FLOPs. The number of parameters is the complexity of the network architecture. For better generalization, a network should have less parameters as possible. FLOPs, i.e., floating point operations, describe the calculations needed in a neural network. With more FLOPs, the network is more complex and time-consuming.

In this study, we selected three advanced CNNs for crop mapping, i.e., the WCRN, the HResNet, and the DBMA. The three networks are representatives of lightweight CNNs, 2-D

CNNs, and 3-D CNNs. In the literature, the DBMA achieved the best classification on hyperspectral benchmarks, and the WCRN is the fastest among the three.

2) Residual Learning and Batch Normalization: Two out of three models contain residual learning, which is a basic element to obtain deeper networks and to accelerate the training process [54]. Assuming that the input feature matrix is $\mathbf{x}$, the output feature matrix $\mathbf{y}$ after a residual unit is given by

$$\mathbf{y} = \mathbf{x} + \mathcal{F}(f(\mathbf{x}), \mathcal{W}) \tag{2}$$

where $\mathcal{F}$ is the residual function containing convolutions and batch normalization [55], f is the activation function, and $\mathcal{W}$ is a set of weights and bias of this layer. If we extend the d-dimensional feature matrix $\mathbf{x}$ in a channelwise fashion, i.e., $\mathbf{x} = (x^1, \ldots, x^d)$, each dimension after batch normalization can be expressed as

$$\hat{x}^k = \mathrm{BN}(x^k) = \frac{x^k - \mathbf{E}[x^k]}{\sqrt{\mathbf{Var}[x^k]}} \tag{3}$$

where the expectation, $\mathbf{E}[\cdot]$, and variance, $\mathbf{Var}[\cdot]$, are calculated over a small batch of training samples. The batch normalization layer keeps the gradients in backpropagation stable and brings generalization ability to the model.

3) Dense Connection: Dense connection claims to be an advanced skip connection technique that outperforms residual learning. Its basic idea is similar to residual learning to use skip connection and help information flow in the network, but in a more progressive way. Inside a residual unit, the residual function $\mathcal{F}$ is extended as two series of batch normalization, ReLU, and a convolutional layer. Each series as $\mathcal{G}$, the residual unit is reformulated as

$$\mathbf{y} = \mathbf{x} + \mathcal{G}_2(f(\mathcal{G}_1(f(\mathbf{x}), \mathcal{W}_1)), \mathcal{W}_2). \tag{4}$$

Dense connection is more progressive

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathcal{G}_1(f(\mathbf{x}_0), \mathcal{W}_1) \tag{5}$$

$$\mathbf{y} = \mathbf{x}_n = \mathbf{x}_0 + \mathcal{G}_1(f(\mathbf{x}_0), \mathcal{W}_1) + \cdots + \mathcal{G}_n(f(\mathbf{x}_{n-1}), \mathcal{W}_n). \tag{6}$$

The DBMA network used in this study contains three dense blocks, i.e., $n = 3$, as shown in the Supplementary Material.

4) Usage of 1 × 1 Convolution: The usage of 1×1 convolutional layer reduces the number of parameters of a deep network and is useful in deep-learning-based remote sensing image classification by easing overfitting [56]. In the pre-deep-learning era, remote sensing image classification consisted of two major procedures [57], [58], which often included designing some filter to extracting spatial features (e.g., GLCM) and the fusion of multisource data (e.g., spectral, spatial, temporal, and backscattering) with dimensional reduction methods (e.g., principal component analysis and manifold learning [59]). If we treat the network into the same two components, then the 1×1 convolutional layer is responsible for the fusion process, and a larger convolutional layer at the head of a network is responsible for spatial feature extraction. The usage of 1×1 convolutional layer significantly accelerates the WCRN.

5) Max Pooling and Average Pooling: Max pooling and average pooling are often used to further reduce the number of parameters in a deep network [60]. For each dimension, the output feature z_{avg} after an average pooling layer can be expressed as

$$z_{\text{avg}} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j}. \tag{7}$$

The output of a max pooling layer is given by

$$z_{\text{max}} = \max(x_{i,j}) \tag{8}$$

where i and j indicate the location of a feature map. When H and W are equal to the height and width of a feature map, respectively, pooling is squeezed to a feature value. In a sense, pooling can be regarded as a special spatial filter.

6) Attention: Attention is a mechanism to assign weights on the feature maps [61], [62]. This technique has been popular since its invention and is used in the DBMA network. Given an extracted feature $\mathbf{x}$, the output features $\mathbf{y}$ after attention are multiplied by a weight matrix $\mathbf{s}$

$$\mathbf{y} = f_m(\mathbf{x}, \mathbf{s}) \tag{9}$$

where $\mathbf{s}$ is given by

$$\mathbf{s} = \sigma(\mathbf{W}_2 f(\mathbf{W}_1 \mathbf{z})) \tag{10}$$

where $\mathbf{z}$ is the channelwise statistic obtained via global average pooling, and $\mathbf{W}_1$ and $\mathbf{W}_2$ are two FC layers.

C. Experimental Details

We implemented the deep learning experiments using Keras with TensorFlow back-end on a personal PC equipped with an Intel I5-8500 CPU, 32-GB RAM, and an Nvidia GTX1080Ti GPU. We used the AdaDelta optimizer to train the network [63]. The learning rate was set as 1.0 for the first 170 epochs and then as 0.1 for another 30 epochs. If the training loss does not decrease for five epochs, the training will enter to the next phase immediately. The batch size was 20. We augmented the training samples horizontally, vertically, and diagonally to make full use of the training samples. For experiments with random forest, we used the scikit-learn library [64]. The number of trees was searched from 100 to 1000 with a step of 100. The number of features was searched from 1 to the square root of the number of features. Threefold cross validation was used to prevent overfitting. As for experiments with LightGBM [65], we used the LightGBM official Python implementation. The boosting mode was set as gradient boosted decision trees, and the number of trees was searched from 100 to 500 with a step of 50. Threefold cross validation was used to prevent overfitting.

IV. Results and Analysis

A. Results on the Chongqing Data

The results obtained on the Chongqing data are shown in Table III. For random sampling, the best result in terms of OA (95.00%) was obtained via DBMA, though the difference among WCRN, DBMA, and HResNet is marginal (93.07%,

TABLE III
CLASSIFICATION ACCURACY OF THE CHONGQING REGION

Index	Random sampling					Regional sampling				
	RF	LightGBM	WCRN	DBMA	HResNet	RF	LightGBM	WCRN	DBMA	HResNet
1	88.68	84.92	85.23	91.04	91.04	71.69	77.71	54.92	54.14	56.87
2	69.67	66.67	80.51	86.62	93.89	32.87	23.88	14.07	17.19	13.84
3	66.63	73.95	91.66	90.85	84.98	50.43	51.13	68.30	65.23	66.63
4	91.42	87.62	94.46	95.66	92.18	10.21	12.92	15.09	5.43	19.22
5	33.66	42.57	73.27	77.23	79.21	0.00	3.96	8.91	26.73	20.79
6	94.13	93.84	99.41	99.41	100.0	30.79	58.65	55.72	59.82	38.42
7	82.15	86.45	97.67	99.00	98.38	64.41	65.15	70.16	69.85	64.46
8	99.02	99.10	99.18	99.51	99.84	99.75	99.67	98.28	96.89	99.10
9	95.20	97.28	99.84	100.0	100.0	51.84	60.64	63.04	53.76	69.60
10	63.50	76.00	90.50	93.50	94.50	0.00	0.00	24.50	0.00	14.00
OA (%)	82.55±0.85	84.38±0.38	93.07±0.96	95.00±0.86	94.23±1.01	58.83±0.87	58.08±0.58	59.57±1.79	57.53±1.53	58.07±1.72
Kappa×100	79.23±0.99	81.33±0.42	91.64±1.04	93.97±0.99	93.05±1.21	51.33±1.02	50.58±0.59	52.18±2.01	50.05±1.84	50.53±1.66
Error (%)	17.54±1.49	11.88±1.78	5.50±0.61	2.58±0.42	7.62±1.43	31.36±1.40	28.55±1.46	33.66±3.23	36.99±4.10	33.31±3.98

TABLE IV
CLASSIFICATION ACCURACY OF THE LANDIVISIAU DATA

Index	Random sampling					Regional sampling				
	RF	LightGBM	WCRN	DBMA	HResNet	RF	LightGBM	WCRN	DBMA	HResNet
1	93.55	94.61	97.63	97.03	96.99	92.41	93.07	92.13	91.40	88.94
2	94.23	93.56	93.17	93.68	94.24	93.33	92.51	91.54	92.69	91.41
3	95.25	95.39	97.23	98.05	97.80	92.86	91.95	88.71	88.52	88.14
4	80.90	88.75	98.19	98.66	99.05	12.71	19.06	29.38	26.64	34.09
5	99.58	99.70	99.94	99.94	99.82	97.83	98.79	96.26	95.41	94.57
6	48.50	57.25	92.41	93.71	94.19	0.00	0.00	1.37	2.53	1.71
7	87.62	92.45	99.08	99.28	99.42	42.60	41.69	37.47	37.05	40.42
8	70.54	72.41	79.79	88.89	87.13	61.82	62.14	61.77	60.94	61.75
9	72.54	72.19	81.13	84.51	87.23	61.66	62.01	62.96	63.99	59.65
10	94.67	95.13	95.63	96.49	95.64	92.63	92.44	91.33	87.81	88.72
11	70.38	69.45	82.81	86.99	83.12	24.96	22.64	21.21	23.24	25.05
12	99.79	99.17	100.0	100.0	100.0	72.31	66.74	48.14	38.43	61.57
OA (%)	86.41±0.38	86.74±0.41	90.30±0.87	92.54±0.81	92.79±0.85	80.31±0.56	80.27±0.38	79.56±1.82	79.44±1.27	78.29±1.73
Kappa×100	83.37±0.48	83.80±0.42	88.17±0.98	90.89±0.75	91.18±0.87	75.83±0.66	75.81±0.41	74.97±1.61	74.76±1.26	73.44±1.92
Error (%)	3.95±1.79	4.92±1.37	4.68±1.19	5.49±1.14	3.25±0.89	5.90±0.99	5.42±1.10	3.60±0.87	3.23±0.49	4.75±0.91

95.00%, and 94.23%). DBMA was also the best model in terms of mapping error. RF achieved an OA of 82.55%, which is lesser than other models in terms of all three metrics. For regional sampling, the WCRN achieved the best result, but the OA dropped dramatically compared to random sampling, from 93.07% to 59.57%. The five classification methods were comparable to each other within one standard deviation. The major influential factor is the sampling strategy. For regional sampling (real-case scenarios), using RF and LightGBM has similar performances as using CNNs.

B. Results on the Landivisiau Data

As for the Landivisiau, France data, we show the classification results in Table IV. For random sampling, the best classification was achieved via HResNet in terms of all metrics, followed by DBMA. The difference between these two models was marginal. The OAs were 92.79% and 92.54%, respectively. RF and LightGBM obtained significantly worse results. With regional sampling, the best two models were, surprisingly, RF and LightGBM. The three CNNs were comparable and slightly worse than the ensemble methods.

Based on the two experiments, we observed that there was no significant difference between ensemble methods and CNNs in distinguishing crop types under the regional sampling strategy. However, in random sampling, the CNNs were better. The reason is straightforward. CNNs are capable of extracting spatial information, where nearby pixels are more dependent on each other than the distant ones. Such reliance provides additional benefit for CNNs in smoothing the classification results.

Although the two sampling strategies had an impact on the classification accuracy, it was not caused (only) by the overlap of spatial filtering. The features we used in RF and lightGBM were single-pixel radar scatters only. They did not contain spatial information from neighborhood pixels. However, it is fair to say that some overestimation was caused in opening a window, since the accuracy gap between RF and CNNs was smaller in regional sampling than random sampling.

C. Comparison Between SAR and Optical Data

To answer the question of whether we can produce crop maps with SAR data that are as good as optical data, we show results on the two study sites in Table V using optical only, SAR only, and both optical and SAR data.

For the Chongqing data, the best classification under random sampling is obtained by using optical data only via HResNet, followed by the combined use of optical and SAR data. Under regional sampling, the best classification is obtained by using optical data only via RF. The difference between the results obtained with optical and SAR data is significant. SAR data can achieve classification with over 90% OA under random sampling and over 55% OA under regional sampling, but is worse than that using optical data only. It is also interesting that the combined used of both data cannot improve the classification.

TABLE V
CLASSIFICATION COMPARISON WITH DIFFERENT CLASSIFIERS AND DIFFERENT SETS OF DATA IN TERMS OF OA (%)

Sampling Strategy		Random sampling					Regional sampling				
Feature Set		Optical	SAR	SAR	Optical+SAR	Optical+SAR	Optical	SAR	SAR	Optical+SAR	Optical+SAR
SAR Terrian Correction		NA	No	Yes	No	Yes	NA	No	Yes	No	Yes
Chongqing	RF	92.83±0.21	82.67±0.62	82.55±0.85	93.08±0.33	92.96±0.29	73.54±0.56	55.66±0.78	58.83±0.87	72.39±0.87	74.23±0.58
	LightGBM	92.80±0.18	83.52±0.56	84.38±0.38	93.12±0.15	93.12±0.19	72.50±0.54	57.33±0.53	58.08±0.58	70.71±0.73	72.51±0.72
	WCRN	95.43±0.49	93.60±0.88	93.07±0.96	95.67±0.95	95.18±0.87	72.72±1.11	58.17±2.25	59.57±1.79	68.56±1.85	70.29±1.88
	DBMA	95.12±0.53	93.79±0.46	95.00±0.86	94.90±0.53	95.41±0.64	71.65±1.03	57.78±1.70	57.53±1.53	64.25±2.01	69.69±1.87
	HResNet	96.03±0.39	94.54±0.81	94.23±1.01	96.12±0.80	95.99±0.75	71.59±0.89	55.49±1.69	58.07±1.72	69.68±1.53	70.33±1.37
Landivisiau	RF	83.46±0.31	86.44±0.42	86.41±0.38	89.32±0.31	89.14±0.25	72.32±0.48	80.01±0.42	80.31±0.56	82.68±0.36	89.32±0.40
	LightGBM	83.65±0.22	86.95±0.30	86.74±0.41	90.33±0.37	90.22±0.26	73.11±0.50	80.20±0.41	80.27±0.38	82.51±0.56	82.60±0.43
	WCRN	87.53±0.54	90.80±0.76	90.30±0.87	93.20±0.65	93.00±0.81	76.27±0.81	80.25±1.64	79.56±1.82	82.20±1.18	82.39±1.46
	DBMA	92.07±0.59	93.25±0.79	92.54±0.81	93.39±0.68	93.29±0.68	78.12±1.27	78.99±1.19	79.44±1.27	81.08±1.03	81.51±1.11
	HResNet	89.97±0.43	92.89±0.65	92.79±0.85	94.29±0.59	94.28±0.61	76.69±1.00	78.63±1.61	78.29±1.73	80.76±1.54	80.52±1.32

For the Landivisiau data, the best classifications under random sampling and regional sampling are obtained using optical and SAR data. The obtained OAs using optical only and SAR data only are similar, which means multitemporal SAR data could be as good as optical data for crop mapping.

The different conclusions via the two study sites are caused by the fact that Landivisiau, France, is a flat agricultural region, as shown in Fig. 4, whereas Chongqing, China, is a mountain agricultural region, as shown in Fig. 2. For the Landivisiau region, the SAR signals did contain different information from the ground that is caused by crop phenology and ground objects. We can easily identify and separate the airport from others. The high values in SAR signals indicate the existence of man-made ground objects, e.g., the airport terminal and downtown Landivisiau. However, for the Chongqing region, it is apparent that the mountains play a very important role in determining the reflected radar signals. In this case, seasonal crop phenology might be too minor in radar signals compared with huge mountains. By comparing the classifications from terrain-corrected SAR and nonterrain-corrected SAR in Table V, we can see that terrain correction improved the classification in regional sampling by about 1–3% for the Chongqing data.

D. Visualization of the Samples Sets via Random Sampling and Regional Sampling

We here show the tSNE visualization of latent features of the HResNet in Fig. 6. The goal is to show that the sample set's variation contributes to the classification gap in the two sampling strategies. We first check with the Chongqing data, as shown in Fig. 6(a) and (b). The tSNE visualization is a dimension reduction tool (similar to principal component analysis) to visualize high-dimensional data, which is helpful in clustering and classification [66], [67]. The x and y axes of the tSNE visualization have no special physical meaning. Each point in the figure represents a sample instance. The closer the points are, the similar they are, which is why instances from the same class always cluster together. We can see some difference regarding the random sampling and regional sampling strategies. Take the point cluster (Lotus) of the Chongqing data on the left as an example. Points from regional sampling are more compact than those from random sampling. The closer these training samples are, the less diverse the training set is. Compared to a diverse training set, a compact training set fails to capture the whole picture of the real world. In this case, the classifier is forced to make classification decisions based on a bias sample set, resulting in overfitting and poor performance of a machine learning model.

This situation is lightened in the Landivisiau data, as shown in Fig. 6(d) and (e), because the training set contains a massive amount of samples (10 000 per class). The massive amount of training data help the classifier capture the data's real-world distribution and enhances the classification performance in regional sampling. Still, sample points are more compact in regional sampling than in random sampling, e.g., the yellow dots representing the rapeseed class.

E. Classification Maps Among the Four Methods With Two Sampling Strategies

To tell the difference, we show the Landivisiau's classification maps obtained with all four methods in both random and regional sampling strategies in Fig. 7. The first row shows classification maps under random sampling, and the second row shows classification under regional sampling. We can see that there is no huge difference in the final classification maps for random sampling with different methods. Some variations exist in the lower left of the image, in which HResNet has a slightly better classification as the classification map is more compact. For regional sampling, the major difference is in the largest cropland in the middle of the image. DBMA and HResNet misclassified some areas inside this cropland, and RF, surprisingly, produced the most accurate classification. However, RF still failed to classify crop types in the lower left of the image.

As for Chongqing, the classification is shown in Fig. 8. The variation between random sampling and regional sampling is high. Take the classification by HResNet as an example. It is apparent that with regional sampling, the classifier failed to capture general features to classify crop types, in this case, paddy. With random sampling, the classification is slightly better, but cannot be categorized as satisfactory, probably due to the fragmented land parcels and the significant difference in elevation. It might be better to use UAV remote sensing in this mountainous region to obtain high spatial and spectral resolution imagery and to avoid bad weather.

F. Importance of SAR Channels of Each Month and the Influential Factors

To understand which month's data are important in crop mapping, we show the importance of SAR channels of each

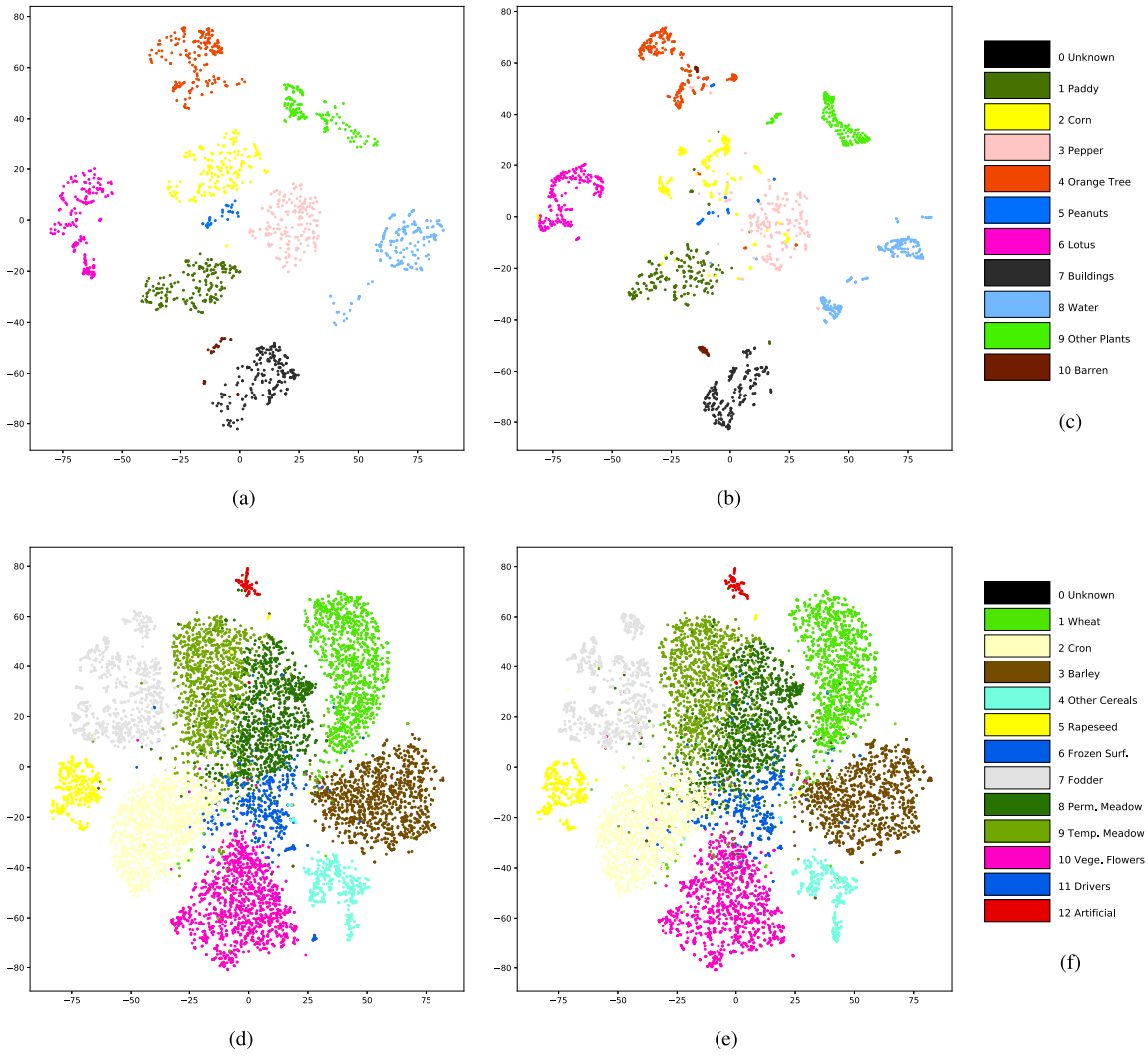


Fig. 6. Visualization of HResNet-extracted features of the training samples using tSNE. Random sampling is more diverse; regional sampling is more compact. A diverse sample set helps the machine learning model achieve better generalization. (a) Random sampling from the Chongqing data. (b) Regional sampling from the Chongqing data. (d) Random sampling from the Landivisiau data. (e) Regional sampling from the Landivisiau data.

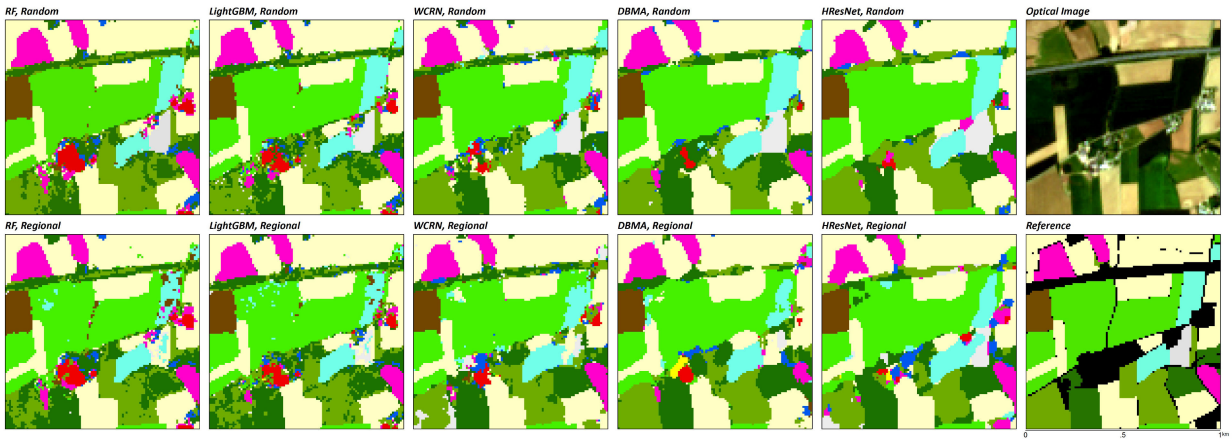


Fig. 7. Classification maps in Landivisiau, France, using five models with random sampling and regional sampling.

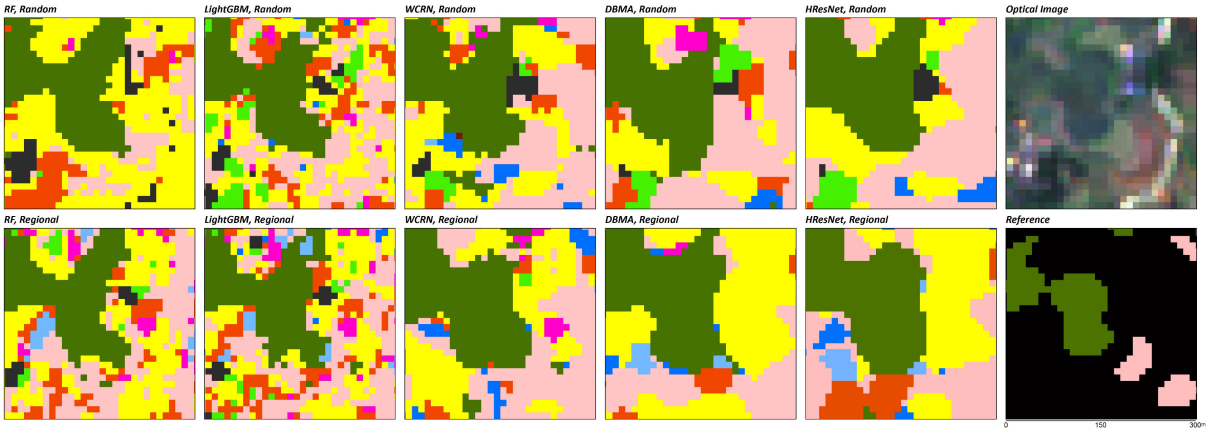


Fig. 8. Classification maps in Chongqing, China, using five models with random sampling and regional sampling.

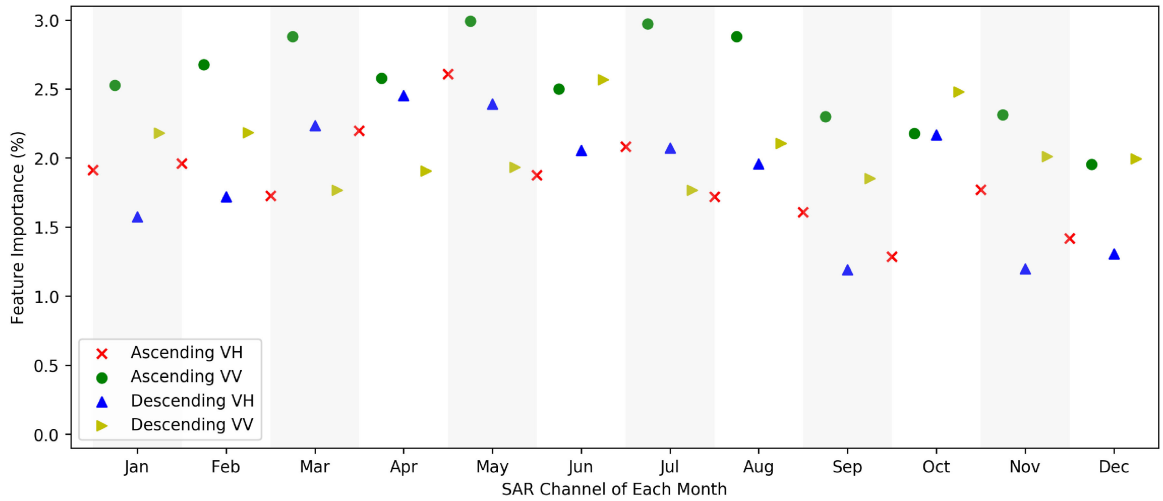


Fig. 9. Importance of SAR channels of each month (Chongqing).

month estimated by random forest in Figs. 9 (Chongqing) and 10 (Landivisiau). For Chongqing, ascending VV (green point) appears to be the most important channel in almost all months, while in some months, descending VV is more important. SAR signals from spring and summer months (January to August) are more important than those from the fall and winter months in distinguishing crop types. For Landivisiau, the most important feature is descending VH of July. SAR signals from the summer months (May to July) are significantly important. The two sites show the importance of summer SAR data in distinguishing crop types. The difference lies in the spring SAR data. Chongqing (29.4°) is located in a subtropical is located in a subtropical monsoon climate region, where the temperature in spring is enough to grow crops. We show the monthly average temperatures of Chongqing and Landivisiau in Fig. 11. In March, when the average high temperature in Landivisiau is below 10°, the average high temperature in Chongqing has already reached 15°. The temperature difference leads to different farming behaviors (Chongqing is two or even three harvests a year, while Landivisiau is one harvest a year), and therefore, the sensitivity of SAR signals is different (due to changes of croplands) in the two regions.

Such difference can also be inferred from the image's entropy, as shown in Fig. 12. Strong linear correlations were found between feature importance and each SAR channel's entropy ($p < 0.01$ for Chongqing, and $p < 0.05$ for Landivisiau).

Along with the SAR channel's entropy, we examined four climatological factors, including monthly average high and low temperatures, soil moisture, and rainfall. Monthly soil moisture data of their respective years (2019 for Chongqing and 2017 for Landivisiau) were obtained from the NOAA Physical Sciences Laboratory. Temperature and rainfall data were obtained from the Climate-Data.ORG. The result is shown in Table VI. These climatological factors are not universal factors correlated with feature importance. Only the high and low temperatures of Landivisiau are strongly correlated with feature importance, which is a reflection of its growing season.

G. *Effects of Growing Seasons*

The growing season should be considered when using full-year SAR data. According to Garonna *et al.* [68], the European countries have a growing season length of about 160 days, starting from the 100th day (March) of the year to the 260th

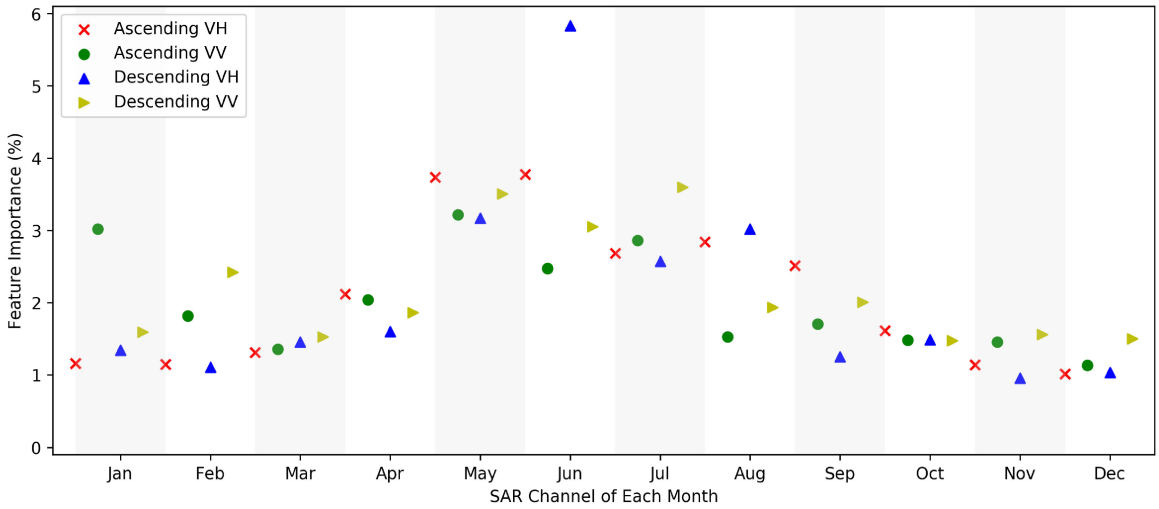


Fig. 10. Importance of SAR channels of each month (Landivisiau).

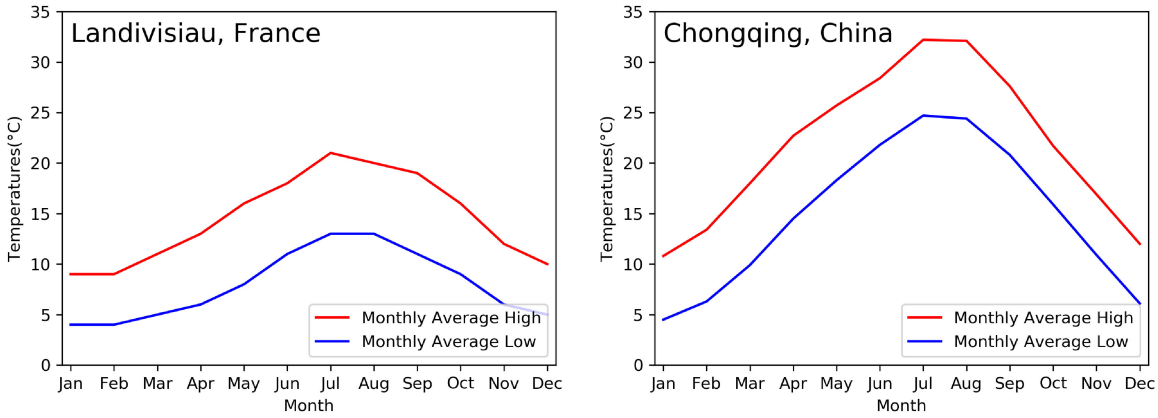


Fig. 11. Monthly average temperatures of Chongqing and Landivisiau.

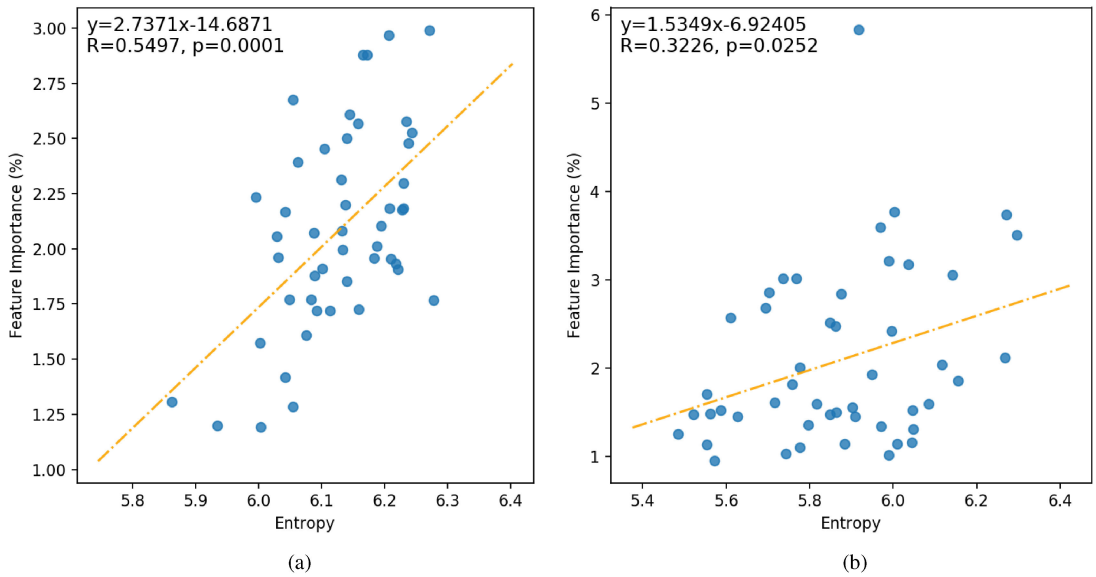


Fig. 12. Relationship between feature importance and the SAR channel's entropy. (a) Chongqing. (b) Landivisiau.

TABLE VI
CORRELATION WITH FEATURE IMPORTANCE

Region	Chongqing, China		Landivisiau, France		
Parameter	R value	p value	R value	p value	Note
Channel's entropy	0.5497	0.0001**	0.3226	0.0252*	on 48 channels
Temperature (High)	0.4101	0.1855	0.6286	0.0285*	on 12 months
Temperature (Low)	0.3534	0.2598	0.5983	0.0399*	on 12 months
Soil moisture	-0.2278	0.4763	0.2049	0.5229	on 12 months
Rainfall	0.5773	0.0493*	-0.5718	0.0521	on 12 months

*indicates p value < 0.05; **indicates p value < 0.01.

TABLE VII
CLASSIFICATION COMPARISON USING FULL-YEAR DATA AND GROWING-SEASON DATA

Test Site	Chongqing, China		Landivisiau, France	
Method	Full-year	Growing-season	Full-year	Growing-season
RF	58.83±0.87	56.33±0.90	80.31±0.56	79.19±0.57
LightGBM	58.08±0.58	56.32±0.28	80.27±0.38	79.22±0.24
WCRN	59.57±1.79	58.07±1.21	79.56±1.82	79.30±1.10
DBMA	57.53±1.53	57.02±1.48	79.44±1.27	78.93±1.29
HResNEt	58.07±1.72	57.17±1.80	78.29±1.73	78.75±1.65

For Chongqing, the growing season is February–November; for Landivisiau, the growing season is March–August.

TABLE VIII
EFFECTS OF DESPECKLING IN TERMS OF OVERALL ACCURACY (%)

Test Site	Chognqing		France	
Method	No Despeckling	Refined Lee	No Despeckling	Refined Lee
RF	57.19	60.88	80.05	81.16
LightGBM	58.08	58.87	80.27	81.18
WCRN	59.57	59.32	79.56	78.54
DBMA	57.53	58.73	79.44	78.21
HResNet	58.07	55.71	78.29	76.95

day (August). For south China (where Chongqing locates), the growing season is 280–320 days, starting as early as January and ending after November [69].

We selected the growing season in Landivisiau as from March to August, and the growing season in Chongqing as from February to November. The comparison between using full-year and growing-season data (via regional sampling) is shown in Table VII. When using CNNs, the OAs were similar or slightly increased using growing-season data. When using 1-D classifier such as RF or lightGBM, the OAs of growing season were smaller than the one of full year. This is because RF and lightGBM are insensitive to noise. The addition of nongrowing season data would not affect their performance. However, CNNs are not selective. If noisy channels or abundant channels were included, their performance could be degraded due to the lack of feature selection and high dimensionality.

H. Effects of Despeckling

In our main experiments, the SAR images used were without despeckling since this is the default setting in Google Earth Engine. To test its effects on crop mapping, we did an experiment on the data after the Refined Lee filtering with a 7×7 kernel size. The result is shown in Table VIII. We can see that usage of the Refined Lee filter led to an increase of classification accuracy for RF and LightGBM, as it included some spatial information that helped the classification. However, usage of the Refined Lee filter led to a decrease of the classification using CNNs, as it blurred the image.

TABLE IX
COMPUTATION TIME (S) AND MODEL COMPLEXITY IN TERMS OF THE NUMBER OF PARAMETERS AND FLOPS

		RF	LightGBM	WCRN	DBMA	HResNet
#Params		NA	NA	58K	67K	96K
#FLOPs		NA	NA	1.16M	1.34M	1.91M
Chongqing	Train	18	30	75 (41)	567	82
	Predict	41	96	175 (111)	1091	260
Landivisiau	Train	716	242	576 (604)	4711	640
	Predict	85	42	78 (58)	341	89

For WCRN, we also report the run time on CPU (Intel I5-8500) in brackets. We did not report this metric on HResNet and DBMA because they are extremely slow on CPU. It took 7 s to run an epoch on CPU with the WCRN, but took 4 min to run an epoch on CPU with the HResNet.

I. Computation Time

Finally, we show the computation time of the five models and the number of parameters and FLOPs of neural networks in Table IX. We can see that DBMA, the 3-D CNN, is exceptionally time-consuming and nearly ten times slower compared to other models. The three deep learning models have similar numbers of parameters and FLOPs. Still, WCRN and HResNet are faster than DBMA because 1) 3-D convolutional layers require a larger memory than the 2-D version, and 2) the implementation of 2-D convolutions is well optimized in TensorFlow, while the 3-D version is not.

With the acceleration of a GTX1080Ti GPU, HResNet and WCRN cost a similar time as RF. For the Landivisiau, France data, since the training samples are greatly enlarged, the time to train a neural network is short than RF. RF is an ensemble learning method of multiple decision trees, and its training time complexity is $\mathcal{O}(ntree \times mtry \times n \log n)$. When $ntree$ and $mtry$ are constant, the training time is linear with $n \log n$, where n is the number of training samples. For LightGBM, its time complexity is $\mathcal{O}(0.5 \times \#feature \times \#bin)$, which is insensitive to the number of training samples [65]. For neural networks, the time complexity is directly linear with the number of training samples. As a result, with a large amount of training data, neural networks can be faster than RF.

It is surprising but still reasonable that the WCRN runs faster on CPU than GPU. This is because: 1) there is some conversion time of the training data between RAM and GPU; 2) the I5-8500 is an advanced CPU; and, most importantly, 3) WCRN processes with only 5×5 input patches and has a simple and elegant architecture with only 58K parameters and 1.16M FLOPs. We only report the computation time of the WCRN because the other two networks are extremely slow. For example, it took 7 s to run an epoch on CPU with WCRN, but took 4 min to run an epoch on CPU with the HResNet. Based on the experiments, the WCRN not only has a similar performance with other deep networks, but also can run on CPU. For those with limited budgets, it is highly recommended to use WCRN as their first choice for crop mapping.

V. Conclusion

In this study, we aim at using multitemporal SAR data to produce crop type classification maps. We throw three questions at the beginning of this article. The first question is whether we can produce crop maps using multitemporal SAR data as accurately as those using optical data. The answer is yes, as we did produce better crop maps in terms of OA, Kappa, and mapping error in the Landivisiau study site. However, it should be noted that Landivisiau is a flat agricultural region located in northwest France. If the agricultural region is not flat enough, such as Chongqing, southwest China, mountains will become a major issue in using radar signals to distinguish crop types. One should note that this was based on two study sites in China and France. In our setting, the SAR signals were averaged monthly. Such setting, although provides a general scenario for crop mapping using SAR data, reduces the potentials of Sentinel-1 dual-polarized SAR data since its maximum capability is one visit per six days. Additionally, the ascending data were obtained in the evening and the descending data in the morning, which could lead to different characteristics of the SAR signals that might benefit crop mapping. The second question is about the random sampling and regional sampling. Classification with random sampling is better compared to regional sampling. The better classification of random sampling is not overestimation caused by overlapping of training and testing samples. It is a result of domain bias. With the same amount of training samples, the training set via random sampling has a larger variation than regional sampling. Therefore, it has a better estimation of the data's real-world distribution, which will bring good generalization to the machine learning models. As for the last question, 3-D CNNs such as DBMA are too time-consuming compared to 2-D CNNs, random forest, and LightGBM. Their performances were barely as competitive as 2-D CNNs. Fewer parameters are not the guarantee of faster computation for 3-D CNNs. It is better to use lightweight 2-D CNNs, such as the WCRN that can run on CPU with small FLOPs, in real-world crop mapping with multitemporal SAR data.

We also examined the potential factors that affect the capability of SAR data in distinguishing crop types. Strong correlations were found between each channel's entropy and its feature importance in distinguishing crop types. Monthly temperature, which is related to the growing season, is a major factor for the Landivisiau region (length of the growing season: 160 days), but not a factor for the Chongqing region (length of the growing season: 300 days). Other factors, including soil moisture and rainfall, did not show significant correlations with feature importance. Future studies should explore more about the influential factors of SAR data to understand the underling physical meaning.

Acknowledgment

The authors would like to thank the two anonymous reviewers for their careful reading of the work and insightful comments, which greatly improved the quality of this manuscript.

References

[1] K. Von Grebmer *et al.*, *2016 Global Hunger Index: Getting to Zero Hunger*. Washington, DC, USA: International Food Policy Research Institute, 2016.

[2] B. D. Wardlow and S. L. Egbert, "Large-area crop mapping using time-series MODIS 250 M NDVI Data: An assessment for the U.S. Central Great Plains," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 1096–1116, 2008.

[3] N. Kussul, S. Skakun, A. Shelestov, M. Lavreniuk, B. Yailymov, and O. Kussul, "Regional scale crop mapping using multi-temporal satellite imagery," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 40, no. 7, pp. 45–52, 2015.

[4] A. Shelestov, M. Lavreniuk, N. Kussul, A. Novikov, and S. Skakun, "Exploring Google Earth engine platform for big data processing: Classification of multi-temporal satellite imagery for crop mapping," *Front. Earth Sci.*, vol. 5, 2017, Art. no. 17.

[5] Z. Pan *et al.*, "Mapping crop phenology using NDVI time-series derived from HJ-1 A/B data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 34, pp. 188–197, 2015.

[6] F. Gao *et al.*, "Toward mapping crop progress at field scales through fusion of landsat and modis imagery," *Remote Sens. Environ.*, vol. 188, pp. 9–25, 2017.

[7] S. Skakun, E. Vermote, J.-C. Roger, and B. Franch, "Combined use of Landsat-8 and Sentinel-2A images for winter crop mapping and winter wheat yield assessment at regional scale," *AIMS Geosci.*, vol. 3, no. 2, 2017, Art. no. 163.

[8] S. Wang, G. Azzari, and D. B. Lobell, "Crop type mapping without field-level labels: Random forest transfer and unsupervised clustering techniques," *Remote Sens. Environ.*, vol. 222, pp. 303–317, 2019.

[9] A. I. De Castro, J. Torres-Sánchez, J. M. Peña, F. M. Jiménez-Brenes, O. Csillik, and F. López-Granados, "An automatic random Forest-OBIA algorithm for early weed mapping between and within crop rows using UAV imagery," *Remote Sens.*, vol. 10, no. 2, 2018, Art. no. 285.

[10] C. Pelletier, S. Valero, J. Inglada, N. Champion, and G. Dedieu, "Assessing the robustness of random forests to map land cover with high resolution satellite image time series over large areas," *Remote Sens. Environ.*, vol. 187, pp. 156–168, 2016.

[11] T. Le Toan *et al.*, "Rice crop mapping and monitoring using ERS-1 data based on experiment and modeling results," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 1, pp. 41–56, Jan. 1997.

[12] X. Jiao *et al.*, "Object-oriented crop mapping and monitoring using multi-temporal polarimetric RADARSAT-2 data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 96, pp. 38–46, 2014.

[13] S.-W. Chen, Y.-Z. Li, and X.-S. Wang, "Crop discrimination based on polarimetric correlation coefficients optimization for PolSAR data," *Int. J. Remote Sens.*, vol. 36, no. 16, pp. 4233–4249, 2015.

[14] H. Tamiminia, S. Homayouni, H. McNairn, and A. Safari, "A particle swarm optimized Kernel-based clustering method for crop mapping from multi-temporal polarimetric L-band SAR observations," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 58, pp. 201–212, 2017.

[15] H. Skriver *et al.*, "Crop classification using short-revisit multitemporal SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 2, pp. 423–431, Jun. 2011.

[16] H. Skriver, "Crop classification by multitemporal C-and L-band single- and dual-polarization and fully polarimetric SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2138–2149, Jun. 2012.

[17] D. Bargiel, "A new method for crop classification combining time series of radar images and crop phenology information," *Remote Sens. Environ.*, vol. 198, pp. 369–383, 2017.

[18] R. Sonobe, "Parcel-based crop classification using multi-temporal terraSAR-X dual polarimetric data," *Remote Sens.*, vol. 11, no. 10, 2019, Art. no. 1148.

[19] M. Busquier, J. M. Lopez-Sanchez, and D. Bargiel, "Added value of coherent copolar polarimetry at X-band for crop-type mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 819–823, May 2020.

[20] H. Li, C. Zhang, S. Zhang, and P. M. Atkinson, "Crop classification from full-year fully-polarimetric L-band UAVSAR time-series using the random forest algorithm," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 87, 2020, Art. no. 102032.

[21] E. Tomppo, O. Antropov, and J. Praks, "Cropland classification using Sentinel-1 time series: Methodological performance and prediction uncertainty assessment," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2480.

[22] X. Xiao and Y. Lu, "Temporal series crop classification study in rural China based on Sentinel-1 SAR data," in *Proc. IEEE 6th Asia-Pacific Conf. Synthetic Aperture Radar*, 2019, pp. 1–4.

[23] M. Arias, M. Á. Campo-Bescós, and J. Álvarez-Mozos, "Crop classification based on temporal signatures of Sentinel-1 observations over Navarre province, Spain," *Remote Sens.*, vol. 12, no. 2, 2020, Art. no. 278.

[24] J. Zhang, Z. Su, and G. Liu, "Effects of terracing and agroforestry on soil and water loss in hilly areas of the Sichuan Basin, China," *J. Mountain Sci.*, vol. 5, no. 3, pp. 241–248, 2008.

[25] A. W. Jacob *et al.*, "Sentinel-1 InSAR coherence for land cover mapping: A comparison of multiple feature-based classifiers," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, no. 1, pp. 535–552, Jan. 2020.

[26] A. Mestre-Quereda, J. M. Lopez-Sanchez, F. Vicente-Guijalba, A. W. Jacob, and M. E. Engdahl, "Time-series of Sentinel-1 interferometric coherence and backscatter for crop-type mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4070–4084, 2020.

[27] Z. Qi, A. Gar-OnYeh, X. Li, and Z. Lin, "A novel algorithm for land use and land cover classification using RADARSAT-2 polarimetric SAR data," *Remote Sens. Environ.*, vol. 118, pp. 21–39, 2012.

[28] Z. Guo, X. Wang, J. Zhou, and J. You, "Robust texture image representation by scale selective local binary patterns," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 687–699, Feb. 2016.

[29] U. Kandaswamy, D. A. Adjeroh, and M.-C. Lee, "Efficient texture analysis of SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2075–2083, Sep. 2005.

[30] J. Liang, J. Zhou, Y. Qian, L. Wen, X. Bai, and Y. Gao, "On the sampling strategy for evaluation of spectral-spatial methods in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 862–880, Feb. 2016.

[31] R. Hänsch, A. Ley, and O. Hellwich, "Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3672–3675.

[32] J. Lange, G. Cavallaro, M. Götz, E. Erlingsson, and M. Riedel, "The influence of sampling methods on pixel-wise hyperspectral image classification with 3D convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2087–2090.

[33] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, 2019.

[34] X. Cao, H. Lu, M. Ren, and L. Jiao, "Non-overlapping classification of hyperspectral imagery with superpixel segmentation," *Appl. Soft Comput.*, vol. 83, 2019, Art. no. 105630.

[35] M. Molinier and J. Kilpi, "Avoiding overfitting when applying spectral-spatial deep learning methods on hyperspectral images with limited labels," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5049–5052.

[36] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021.

[37] X. Liu, L. Jiao, X. Tang, Q. Sun, and D. Zhang, "Polarimetric convolutional network for PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 3040–3054, May 2019.

[38] Z. Wen, Q. Wu, Z. Liu, and Q. Pan, "Polar-spatial feature fusion learning with variational generative-discriminative network for PoLSAR classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8914–8927, Nov. 2019.

[39] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.

[40] W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 766–785, Mar. 2021.

[41] S. Liu, Z. Qi, X. Li, and A. G.-O. Yeh, "Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and SAR data," *Remote Sens.*, vol. 11, no. 6, 2019, Art. no. 690.

[42] S. Liu, H. Luo, Y. Tu, Z. He, and J. Li, "Wide contextual residual network with active learning for remote sensing image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 7145–7148.

[43] S. Liu and Q. Shi, "Multitask deep learning with spectral knowledge for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2110–2114, Dec. 2020.

[44] W. Ma, Q. Yang, Y. Wu, W. Zhao, and X. Zhang, "Double-branch multi-attention mechanism network for hyperspectral image classification," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1307.

[45] P. Du, A. Samat, B. Waske, S. Liu, and Z. Li, "Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features," *ISPRS J. Photogrammetry Remote Sens.*, vol. 105, pp. 38–53, 2015.

[46] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 26, no. 10, pp. 4843–4855, Oct. 2017.

[47] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral-spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.

[48] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral-spatial convolution network framework for hyperspectral images classification," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1068.

[49] R. Yao, Q. Luo, Z. Luo, L. Jiang, and Y. Yang, "An integrated study of urban microclimates in Chongqing, China: Historical weather data, transverse measurement and numerical simulation," *Sustain. Cities Soc.*, vol. 14, pp. 187–199, 2015.

[50] Y. Yang, N. Zhao, X. H. Hao, and C. Li, "Decreasing trend of sunshine hours and related driving forces in north China," *Theor. Appl. Climatol.*, vol. 97, no. 1-2, pp. 91–98, 2009.

[51] C. Robinet, P. Baier, J. Pennerstorfer, A. Schopf, and A. Roques, "Modelling the effects of climate change on the potential feeding activity of thaumetopoea pityocampa (Den. & Schiff.) (Lep., Notodontidae) in France," *Global Ecol. Biogeography*, vol. 16, no. 4, pp. 460–471, 2007.

[52] M. Rußwurm, C. Pelletier, M. Zollner, S. Lefévre, and M. Körner, "BreizhCrops: A time series dataset for crop type mapping," *ISPRS-Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. pp. 1545–1551, 2019.

[53] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 818–833.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[55] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 448–456

[56] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 145, pp. 120–147, 2018.

[57] H. Murray, A. Lucieer, and R. Williams, "Texture-based classification of sub-antarctic vegetation communities on heard island," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 12, no. 3, pp. 138–149, 2010.

[58] B. Ren, B. Hou, Z. Wen, W. Xie, and L. Jiao, "PolSAR image classification via multimodal sparse representation-based feature fusion," *Int. J. Remote Sens.*, vol. 39, no. 22, pp. 7861–7880, 2018.

[59] C. He, M. Tu, D. Xiong, and M. Liao, "Nonlinear manifold learning integrated with fully convolutional networks for PolSAR image classification," *Remote Sens.*, vol. 12, no. 4, 2020, Art. no. 655.

[60] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A theoretical analysis of feature pooling in visual recognition," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 111–118.

[61] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[62] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[63] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*.

[64] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[65] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, vol. 30, pp. 3146–3154.

[66] L. v. d. Maaten and G. Hinton, "Visualizing data using T-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, 2008.

[67] Z. Zhong, J. Li, L. Ma, H. Jiang, and H. Zhao, "Deep residual networks for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 1824–1827.

[68] I. Garonna, R. De Jong, A.J. De Wit, C.A. Mücher, B. Schmid, and M. E. Schaepman, "Strong contribution of autumn phenology to changes in satellite-derived growing season length estimates across Europe (1982-2011)," *Global Change Biol.*, vol. 20, no. 11, pp. 3457–3470, 2014.

[69] Y. Song, H. W. Linderholm, D. Chen, and A. Walther, "Trends of the thermal growing season in China, 1951-2007," *Int. J. Climatol. A. J. Roy. Meteorol. Soc.*, vol. 30, no. 1, pp. 33–43, 2010.