sensors

MDPI

Article

Identification of Crop Type in Crowdsourced Road View Photos with Deep Convolutional Neural Network

Fangming Wu [1], Bingfang Wu [1,2,*], Miao Zhang [1], Hongwei Zeng [1,2] and Fuyou Tian [1,2]

[1] State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China; wufm@radi.ac.cn (F.W.); zhangmiao@radi.ac.cn (M.Z.); zenghw@radi.ac.cn (H.Z.); tianfy@radi.ac.cn (F.T.)

[2] College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: wubf@radi.ac.cn

Abstract: In situ ground truth data are an important requirement for producing accurate cropland type map, and this is precisely what is lacking at vast scales. Although volunteered geographic information (VGI) has been proven as a possible solution for in situ data acquisition, processing and extracting valuable information from millions of pictures remains challenging. This paper targets the detection of specific crop types from crowdsourced road view photos. A first large, public, multiclass road view crop photo dataset named iCrop was established for the development of crop type detection with deep learning. Five state-of-the-art deep convolutional neural networks including InceptionV4, DenseNet121, ResNet50, MobileNetV2, and ShuffleNetV2 were employed to compare the baseline performance. ResNet50 outperformed the others according to the overall accuracy (87.9%), and ShuffleNetV2 outperformed the others according to the efficiency (13 FPS). The decision fusion schemes major voting was used to further improve crop identification accuracy. The results clearly demonstrate the superior accuracy of the proposed decision fusion over the other non-fusion-based methods in crop type detection of imbalanced road view photos dataset. The voting method achieved higher mean accuracy (90.6–91.1%) and can be leveraged to classify crop type in crowdsourced road view photos.

Keywords: crop type; crowdsourced road view photo; deep convolutional neural network; automatic photo identification; ensemble classification

check for updates

Citation: Wu, F.; Wu, B.; Zhang, M.; Zeng, H.; Tian, F. Identification of Crop Type in Crowdsourced Road View Photos with Deep Convolutional Neural Network. *Sensors* **2021**, *21*, 1165. https://doi.org/10.3390/s21041165

Academic Editor: Sonya A. Coleman
Received: 26 November 2020
Accepted: 4 February 2021
Published: 7 February 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Zero Hunger has been recognized as one of the core sustainable development goals [1,2]. Although global food production is increasing, some countries have still been short of food in recent years [3]. Against the background of global climate change, the frequency of extreme weather further increases the uncertainty of food production. Timely, transparent, and accurate information on global agricultural monitoring is essential for ensuring the proper functioning of food commodity markets and limiting extreme food price volatility [4]. Accurate and reliable crop type information is vital for many applications such as crop area statistics, yield estimation, land use planning, and food security research.

Remote sensing techniques have been proven to be an efficient, objective, and cost-effective method of agricultural monitoring at global, national, and sub-national scales. With more remote sensing data being made public and the development of cloud computing, it is possible to use these data for the large-scale classification of farmland types [5–8]. However, influenced by environmental factors such as elevation distribution, farmland area, land cover richness and cloud cover frequency, the overall accuracy of the four farmland products was below 65% and the standard deviations among all four cropland datasets varied from 0 to 50% [9]. In order to improve the accuracy of future cropland products, cropland classification methods require more and richer training or verification data collected from ground surveys to build a robust model for crop identification [10].

The crowdsourcing method is available for collecting field data and managing public access (such as Geo-WiKi.org and iNaturalist.org). Crowdsourcing geo-tagged images from Flickr and Geograph were used to create a binary land cover classification (developed/undeveloped) for an area of 100 × 100 km^2 in Great Britain, and the accuracy achieved was around 75% [11]. Roadside sampling strategies for cropland data collection that enable the sampling of large areas at a relatively low cost have been suggested [12,13] and integrated with remote sensing data to provide crop acreage estimations [14]. Quality control of crowdsourcing geo-tagged data is very important; otherwise, the user will be unable to assess the quality of the data or use it with confidence [15]. For crowdsourcing photo classification, visual interpretations by volunteers have been used in previous studies [16]. On the other hand, they also discussed the long time required for visual interpretation of many photos, and automatic approaches should be proposed instead of manual classification.

Deep learning is a recent, modern technique for photo processing and data analysis that has resulted from the continued development of computer hardware and the appearance of large-scale datasets [17]. Convolutional neural networks (CNNs), one of the most successful network architectures in deep learning methods, have been developed for photo recognition and applied to complex visual photo processing. Deep learning CNNs have entered the domain of agriculture, with applications such as plant disease and pest recognition [18], picking and harvesting automatic robots [19], weed-crop classification [20], and monitoring of crop growth [21]. Photo datasets are the most commonly used basic data in the field of deep learning. Some research data come directly from crowdsourced datasets, such as ImageNet [22], iNaturalist [23] and PlantVillage [24]. Most applied agriculture research collects sets of real photos based on the research needs of fine-grained photo categorization, such as DeepWeeds [25], CropDeep [26] and iCassava 2019 [27]. The DeepWeeds dataset consists of 17,509 labelled images of eight nationally significant weed species native to eight locations across northern Australia. The CropDeep species classification and detection dataset, consisting of 31,147 images with over 49,000 annotated instances from 31 different classes of vegetables and fruits grown in greenhouses. iCassava 2019 is a dataset consisting of 9436 labeled images covering healthy cassava leaves as well as 4 common diseases. These datasets do not focus on crop types in the field and all photos are close-ups of the identified objects.

However, if there are no public benchmark datasets specifically designed for crop type classification, this limits the further application of deep learning technology and the development of intelligence in acquiring accurate crop types, distributions, and proportions. A study in Thailand explored the potential of using deep learning to classify photos from Google Street View (GSV) for the identification of seven regionally common cultivated plant species along roads, and the overall accuracy of the multiclass classifier was 83.3% [28]. A total of 8814 GSV images with 7 classes of crop for the Central Valley and 1 class representing "other" were prepared for training the CNN model in the USA, and the overall classification accuracy was 92% [29]. Obviously, the timing of GSV photo capture may not be during the growing seasons, and the revisit frequency in most rural areas is low to nonexistent. It has been demonstrated that GSV survey detected fewer plants than car surveys in Portugal countrywide [30]. The dataset of the first study is not public, and, while the dataset of the second can be download freely, the size of the dataset is relatively small. In addition, more state-of-the-art classification networks and model ensemble could be compared and selected to improve performance.

The objectives of this paper are (1) to build a large road view crop photo dataset to support automatic fine-grained classification with deep learning, and (2) to identify and fuse the optimal deep learning architectures for road view photo classification.

The remainder of this paper is organized as follows: Section 2 presents a specific description of the iCrop datasets and introduces the deep learning classification network and the selected data augmentation process. Section 3 presents the experimental performance

and results. Section 4 discusses which model is best for the different applications. Finally, we summarize the conclusions and propose our further research aims in Section 5.

2. Materials and Methods

2.1. Dataset

To build a road view crop photo dataset to support automatic fine-grained classification with deep learning, the crowdsourced photos were collected, cleaned, labeled and divided as shown in Figure 1.

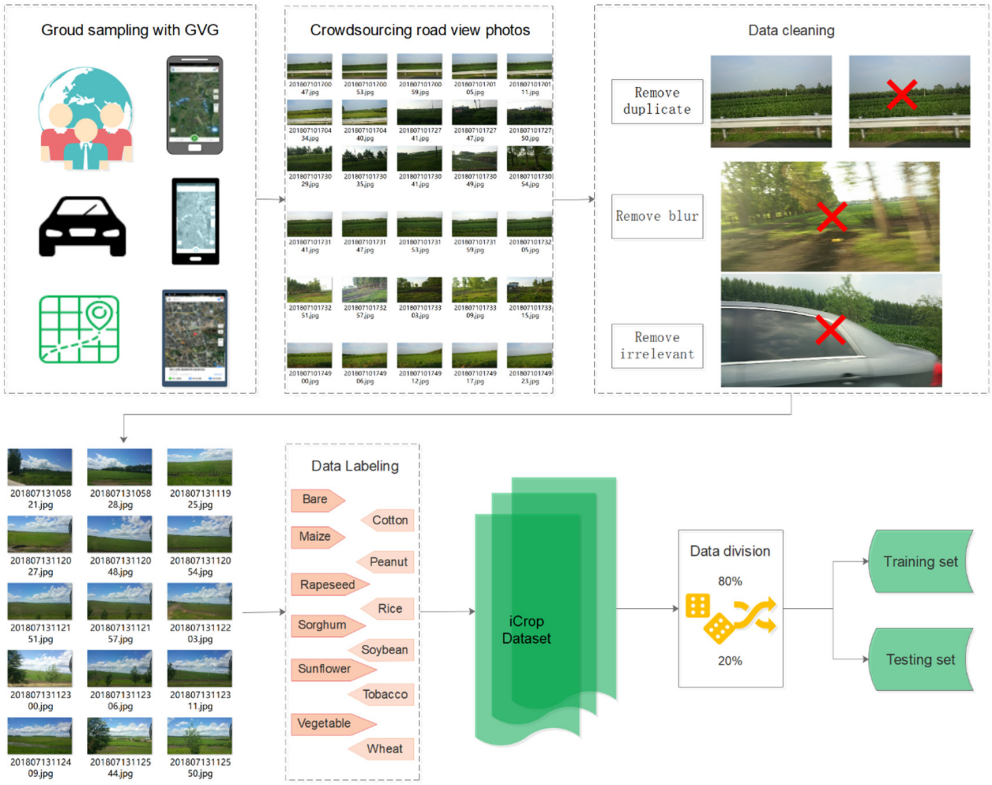


Figure 1. A workflow diagram of crowdsourced crop photos collection, cleaning, labeling and division.

2.1.1. Data Collection

The data were collected by a smartphone app named GVG as part of a crowdsourcing project initiated by the CropWatch team since 2015 [31]. GVG is mobile phone software that can record the photos, location and time of crops at the same time, and users can mark the types of crops in the photos. It is freely downloaded from application marketplaces such as Google Play, the Apple App Store, the Huawei App Gallery and other app marketplaces. The GVG application is easy to use for non-professionals and reduces the amount of ground observation work. A tutorial of field data collection with GVG can be download from http://www.nwatch.top:8085/icrop/docs/gvg.pdf (accessed on 28 October 2020). As shown in Figure 2, the phone was mounted on the window for fast roadside sampling along the road with the help of vehicles. Hundreds of thousands of roadside view photos were collected automatically from the main grain-producing areas of China, including Liaoning, Hebei, Shandong, Jilin, Inner Mongolia, Jiangxi, Hunan, Sichuan, Henan, Hubei, Jiangsu, Anhui, and Heilongjiang. The sampling time was based on the crop phenology calendar. These data have been used to support the paddy field/dry land identification and other land cover mapping [6,32].

Figure 2. GVG was used for data collection along the road base from a vehicle.

2.1.2. Dataset Cleaning and Labeling

First, the photos without cropland or severely blurred field photos were deleted; then, all valid photos were annotated and submitted by the observers, consisting of photos, classes, locations and observation time. A web photo management system was built to easily view and manage the photos based on the Piwigo photo management tool. When a user logs in, they can check whether the photo classification is correct and mark the incorrectly classified photo with the correct classification. All users can rate the trustworthiness of the photo tags on a five-point scale. Photos with an average score less than 3 will be removed or re-tagged. With this method, 34,117 correct photos were divided into twelve types, representing the dominate crop types or farmland without crop, including cotton, maize, peanut, rape, rice, sorghum, soybean, sunflower, tobacco, vegetable, and wheat.

2.1.3. Training and Validation Subsets

Based on fine annotation, the photos were randomly divided into a training set and a test set with approximately 80% of the photos included in the training set. The 80/20 split rate of the training/test dataset is the most common in deep learning applications, and other similar split rates (e.g., 70/30) should not have a significant impact in the performance of the developed model [33]. These empirical proportions make up for the imbalance problem in the dataset. Even if there are some classes with a large number of training samples, their corresponding test sets also contain more samples; thus, these samples undergo more rigorous assessment. At this point, we have the final photo splits, with a total of 27,401 training photos and 6716 test photos. In Table 1, we list the numbers of training and test photos in each category in the iCrop dataset. Randomly selected sample photos from each category of the dataset can be viewed in Figure 1 in Appendix A. It can be seen from the picture that the weather, color, and angle of each photo are different, and some crops are partially obscured by trees and buildings along the road.

The largest crop class was “rice”, with 5850 photos, and the smallest was “sorghum”, with 149 photos. We can see that the photos of each class are imbalanced, which represents a long-tailed real-world challenge in classification problems.

Table 1. Number of training and test photos in each category in the iCrop dataset.

Categories	Train	Test	Total
Bare land	1543	327	1870
Cotton	249	59	308
Maize	4114	1044	5158
Peanut	608	139	747
Rapeseed	3695	871	4566
Rice	4646	1204	5850
Sorghum	112	37	149
Soybean	4257	1080	5337
Sunflower	199	44	243
Tobacco	297	70	367
Vegetable	3651	890	4541
Wheat	4030	951	4981

2.2. Method

The general structure of the presented method identifying and fusing optimal deep learning architectures for road view photo classification is shown in Figure 3.

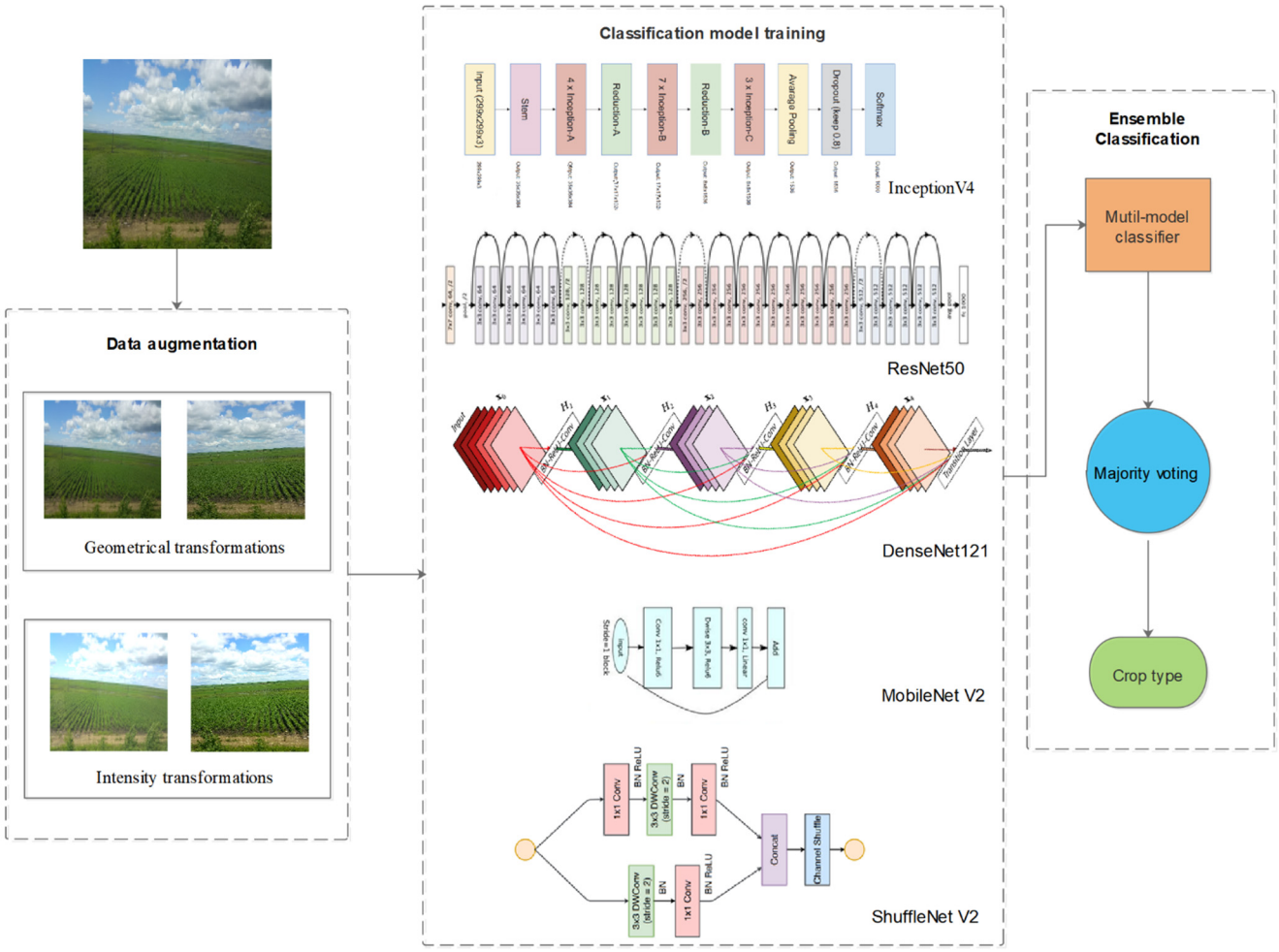


Figure 3. The overall model framework of CNN model for crowdsourcing crop photo classification.

2.2.1. Data Augmentation

The data scale has a great influence on the accuracy of a deep learning network model. A small amount of data will lead to the overfitting of the model, making the training error

very small and the testing error extremely large. To avoid overfitting of the network, some augmented preprocessing was applied to enhance a large number of photos in the dataset before training [34]. The augmentation process has been reported to improve classification accuracy in many studies [35–40]. We used two primary ways to generate new photos from raw photo data with very little computation before training, and the transformed photos only need to be stored in memory. The first method is geometrical transformations consisting of resizing, random cropping, rotation and horizontal flipping [41]. The second method is intensity transformations consisting of contrast, saturation, brightness, and color enhancement [42].

2.2.2. Convolutional Neural Networks

Deep learning allows computing models from multiple processing layers to learn data that have multiple abstract levels. Although a series of CNN models have shown outstanding performance on plant disease detection and diagnosis, the challenges related to addressing other agricultural tasks online or offline are still difficult to overcome [43]. Therefore, to characterize the classification difficulty of iCrop, we ran experiments with five state-of-the-art CNN models, including Resnet50, InceptionV4, Densenet121, MobileNetV2 and ShuffleNetV2.

ResNet50, which was the champion of the ImageNet Large Scale Visual Recognition Challenge (ILSVCR) 2015, introduces a new residual structure and solves the problem that the accuracy rate decreases as the network deepens [44]. Once it was established, InceptionV4 was improved from InceptionV3 (the winner of ILSVRC2014) with resumed connectivity, greatly accelerating training and improving performance through ResNet's structure [45]. The best paper of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017 proposed a densely connected network structure named DenseNet121, which is more conducive to the transmission of information flow [46]. At present, most deep learning networks run on computers with strong floating-point computing power. However, MobileNetV2 and ShuffleNetV2 are designed for mobile and embedded visual applications. The finer tuning of MobileNetV2 based on the MobileNet structure, skipped linking directly on a thinner bottleneck layer, and no ReLu nonlinear processing on the bottleneck layer can achieve better results [47]. ShuffleNet-V2 is a lightweight CNN network that balances speed and accuracy [48]. At the same complexity, it is more accurate than ShuffleNet and more suitable for mobile and unmanned vehicles.

These networks were implemented in PaddlePaddle deep learning frameworks, which is an open-source platform with advanced technologies and rich features [49]. An excellent RMSprop optimizer proposed by Geoff Hinton was used in training the adaptive learning rate [50]. Training batches of size 30 were created by uniformly sampling from all available training photos as opposed to sampling uniformly from the classes.

2.2.3. Ensemble Classification

Ensemble is the process of fusing information from several sources after the data have undergone preliminary classification to improve the final decision [51]. To improve crop identification accuracy, deep learning networks with good accuracy and fast speed will be selected to decision fuse. These models can be seen as different experts focusing on different point of views, whose decisions are complementary and could be fused as a more accurate and stable one. Ensemble classification based on majority voting is proposed in this paper. Majority Voting is one of the most popular, fundamental and straightforward combiners for the predictions from multiple deep learning algorithms [52]. Every individual classifier vote for one class label. The class label that most frequently appears in the output of individual classifiers is the final output. Majority voting was applied on the individual classification results of all classifiers without a reliability check.

3. Results

Experiments with deep learning classification architectures were carried out. All models were trained and tested on an Intel(R) Xeon(R) Gold 6148 CPU @ 2.40 GHz with NVIDIA Tesla V100-SXM2 GPU and 16G RAM. The training proceeded on the training set, after which the evaluation was performed on the validation set to minimize overfitting. When the training process and parameter selection were achieved, the final evaluation was performed on the unknown testing set to evaluate the performance. During training and testing, the photo was adjusted to 224 px as the input of the network.

3.1. Accuracy of Single CNN

The classification test accuracy across all species of each model is shown in Table 2. Classification accuracy is mentioned as the Rank-1 identification rate per class [53]. The percentage of correct predictions where the top class (the one with the highest probability), as indicated by the deep learning model, is the same as the previously annotated target label. For multiclass classification problems, "Average accuracy" indicates the total number of correct prediction samples divided by the total number of testing photos. We observe a larger difference in accuracy and small difference in a average accuracy across the different crops.

Table 2. Test accuracy across all species computed by the five classification models.

Categories	InceptionV4	DenseNet121	ResNet50	MobileNetV2	ShuffleNetV2
Bare land	76.7	78.3	**83.2**	79.8	81.3
Cotton	62.7	76.2	**84.7**	69.5	81.4
Maize	91.1	**93.1**	92.1	92.9	92
Peanut	59.0	64.7	**73.4**	68.3	65.5
Rapeseed	89.2	**96.7**	93.5	88.9	93.2
Rice	93.5	87.0	88.9	**95.5**	93
Sorghum	59.5	73.0	73.0	70.3	**75.7**
Soybean	85.5	83.5	**85.4**	79.4	82.1
Sunflower	79.5	65.9	**72.7**	56.8	**72.7**
Tobacco	72.9	65.7	71.4	81.4	**82.9**
Vegetable	81.1	80.4	**81.8**	79.7	81.3
Wheat	88.4	84.5	91.7	**96.0**	89.3
Average accuracy	**86.2**	**86**	**87.9**	**87.5**	**87.5**

Numbers in bold represent the best classification accuracy for each cropland type.

As seen in Table 2, for rapeseed, the DenseNet121 model outperformed the other models with an accuracy of 96.7%. However, for wheat, the MobileNetV2 model outperformed the other models with an accuracy of 96.0%. For average accuracy, the ResNet50 model outperformed the other models with an average accuracy of 87.9%. MobileNetV2 and ShuffleNetV2 have similar average accuracies of 87.5%. However, the worst model was DenseNet121, which obtained an accuracy of only 86%.

3.2. Identification Efficiency of Single CNN

The original intention of the collected database was to construct an intelligent platform that could be operated online or offline on various mobile phones and other equipment; this database required not only accuracy but also real-time performance to further improve the overall timeliness and efficiency of precision crop structures. Thus, the frames per second (FPS) were selected as the evaluation indicator to evaluate the speed performance of each classification model on the same machine. Please note that the time taken to perform the required preprocessing steps was also measured. These steps include loading a photo and resizing it for input to the network. The evaluation results of the detection time are shown in Figure 4. The result shows that ShuffleNetV2 with approximately 13 FPS is fastest to meet the needs for real-time cropland classification.

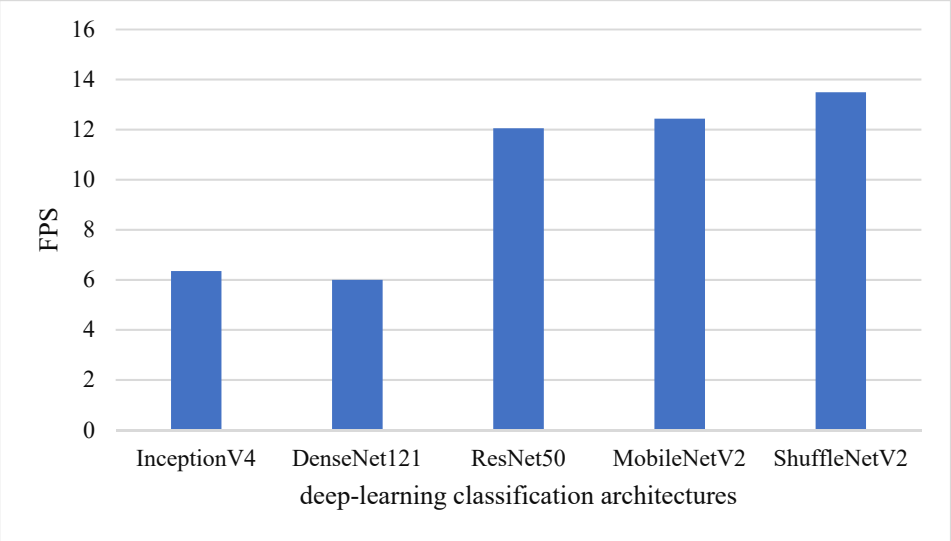


Figure 4. Speed performance of five deep learning classification networks.

3.3. Fusion Accuracy

To avoid the draw problem, the number of classifiers performed for voting is usually odd. We had two voting schemes, one named voting-5 which contained five CNN classifiers and another named voting-3 which contained three CNN classifiers. As shown in Figure 5, the ResNet50 model outperformed the other models on average classification accuracy, and the ShuffleNetV2 model outperformed in classification speed. However, the differences among ResNet50, MobileNetV2 and ShuffleNetV2 on average accuracy and speed were very small. Therefore, the classification results of ResNet50, MobileNetV2 and ShuffleNetV2 were selected in the voting-3 scheme. According to the comparisons presented in Table 3 and Figure 5, small differences in accuracy and average accuracy can be observed between voting-5 and voting-3 schemes.

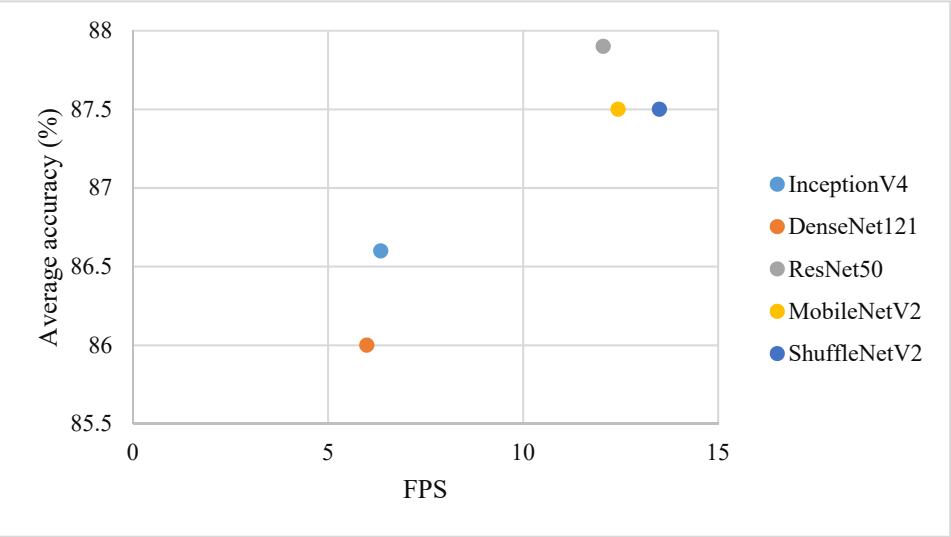


Figure 5. Average accuracy vs. FPS of five deep learning classification networks.

Table 3. Test accuracy across all species computed by the fusion of three or five CNN classifiers.

Categories	Voting-3	Voting-5
Bare land	85	85.3
Cotton	84.7	84.7
Maize	94.7	95.3
Peanut	74.8	72.7
Rapeseed	94.4	96.1
Rice	94.9	95.7
Sorghum	75.7	73.0
Soybean	85.9	86.8
Sunflower	70.5	70.5
Tobacco	84.3	77.1
Vegetable	85.2	86.5
Wheat	94.3	92.8
Average accuracy	90.6	91.1

4. Discussion

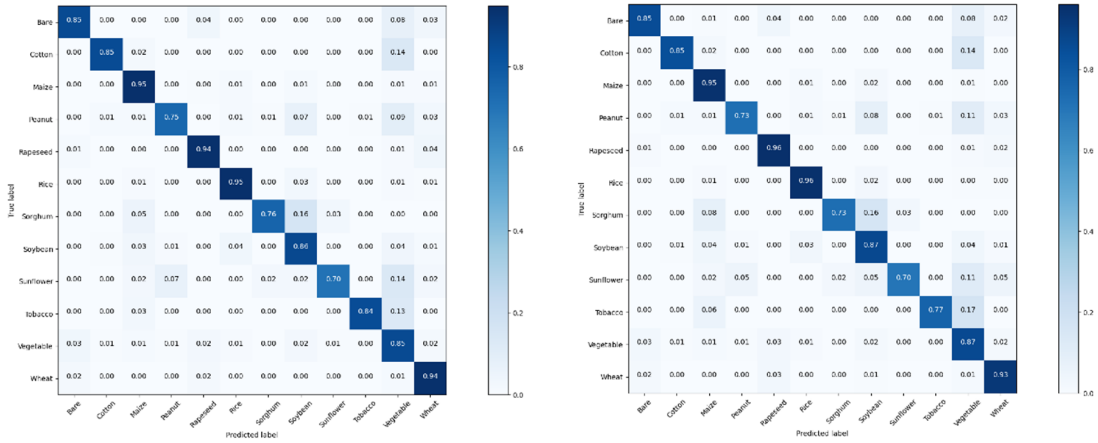
We present a road view crop type dataset named iCrop for the development of crop type classification systems to support remote sensing crop distribution mapping as well as crop area estimation. To the best of our knowledge, no comparable, publicly available dataset exists that is the basis for deep learning research, and the datasets that are currently available (the most influential is ImageNet) do not have many photographs of arable crop, and the angles and distances of the shots are different. Therefore, the photos taken by GVG were sorted, classified and corrected, and the training set and test set were divided.

Unlike GSV, our photos were collected during the crop growing season, capturing the differences in the field as the places and mobile phones change. The angles, heights and directions of the photos are different for each person, and photos also vary in resolution, color, contrast, and clarity. The photos in the datasets contain similar characteristics and imbalance among each class, which reflects the long-tailed real-world challenges in classification problems. Therefore, our photos are more challenging to classify than GSV photos.

The baseline classification results were determined from our experiments. We can see that state-of-the-art CNN models have room to improve when applied to imbalanced roadside view crop datasets. None of the CNN models have the best recognition accuracy for all kinds of crops. The test data, training environment, iteration times and other conditions are the same, but the complexity of the structure for each model is different. The accuracy and complexity of the model are not necessarily related. For wheat, rice, tobacco, and sunflower lightweight models such as MobileNetV2 and ShuffleNetV2, the crop classification accuracy is high. There are different feature fusion methods between DenseNet121 and ResNet50, and the accuracies of classification for different kinds of crops are also similar, but in general, ResNet50 exhibits slightly higher accuracy. Different CNNs are "accurate" in certain aspects, so model fusion could improve the final prediction ability to a greater or lesser degree. According to the comparisons in Tables 1 and 3, the ensemble classification accuracy is higher than individual models for most species, and the average accuracy is also higher than that of each model; in particular, voting with five classifiers increased the overall accuracies by up to 3.5%.

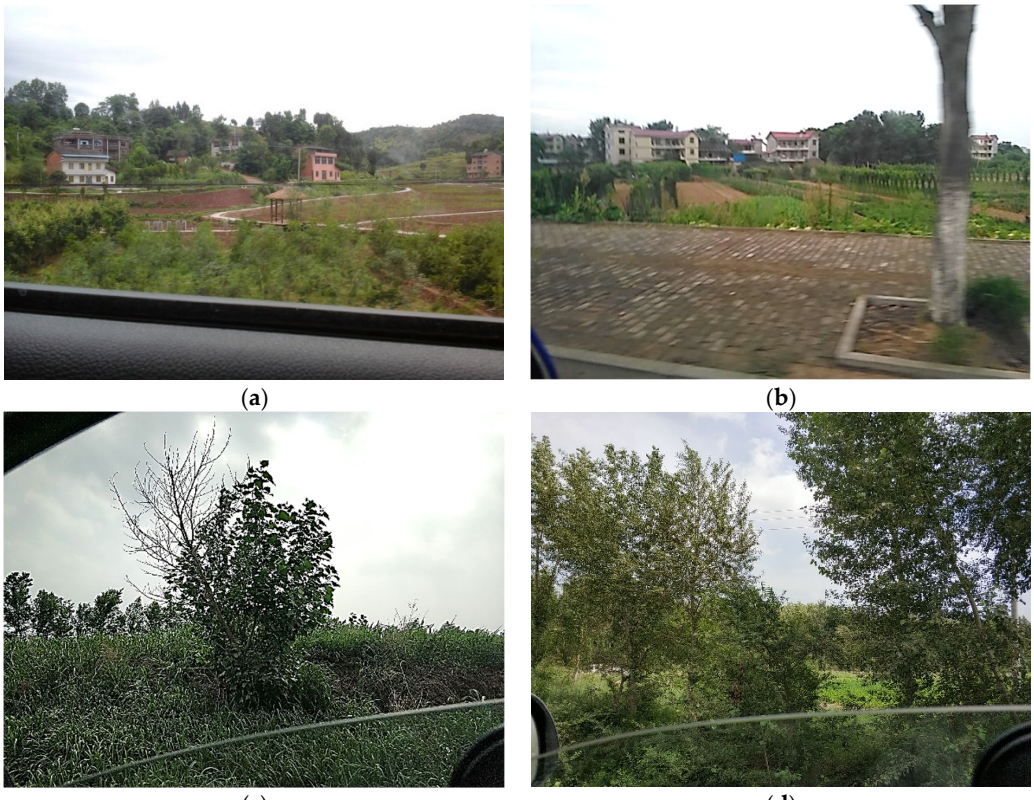
Figure 6 shows the normalized confusion matrix resulting from combining the voting-3 and voting-5 models' performances across the 12 cross-validated test subsets. The model confuses 8% of bare land images with vegetable, and 3% vice versa. Reviewing these particular samples shows that bare land with weeds looks strikingly similar to cropland with small leafed vegetables, while some vegetable gardens contain small patches of bare land. This is illustrated in the sample misclassification of bare land in Figure 7a,b. This likeness is the reason for these false positives in our model. Figure 7c,d also show that crops badly shaded by trees or grasses in photo also affect the classification accuracy. It is necessary to add segmentation information to photos for deep learning to extract farmland

types more accurately. Another way to protect crops from being obscured by other objects is to use drones to take pictures over farmland. Figure 6 and Table 1 illustrate that the accuracy rates are higher for rapeseed, which did not have the largest training sample size. This relatively high performance on distinct crop color features is likely because CNN exhibits better performance for capturing the color characteristics of these crop types.



(**a**) (**b**)

Figure 6. Normalized confusion matrices of the (**a**) voting-3 and (**b**) voting-5 models' performance across the 12 cross-validated test subsets.



(**a**) (**b**)

(**c**) (**d**)

Figure 7. Example images highlighting confusions between classes of species. Specifically, (**a**) bare land falsely classified as vegetable, (**b**) correctly classified vegetable, (**c**) sorghum falsely classified as vegetable, (**c**) peanut falsely classified as vegetable.

The top test accuracies against the number of training photos for each class from the five classification models and ensemble models are plotted in Figure 8. It is shown that there is a positive correlation between the number of training images and the test accuracy. The consensus of most current studies is that for deep learning, the performances will increase with growing data size [54–57]. However, we still observe a variance in the accuracy for classes with a similar amount of training data, revealing opportunities for algorithmic and dataset improvements in both the low data and high data regimes. The ensemble classification accuracy is significantly higher than individual model for low data species. The training images of peanut, sunflower and sorghum are all lower than those of other crops, and the performances for these crops are even low after fusion. it is caused by dataset imbalance. No imbalance-correcting technique can match adding more training data when it comes to measuring precision and recall. We suggest corresponding data collection efforts for classes with few photo samples should be also underway use a hybrid method that fuses GSV images and crowdsourcing data.

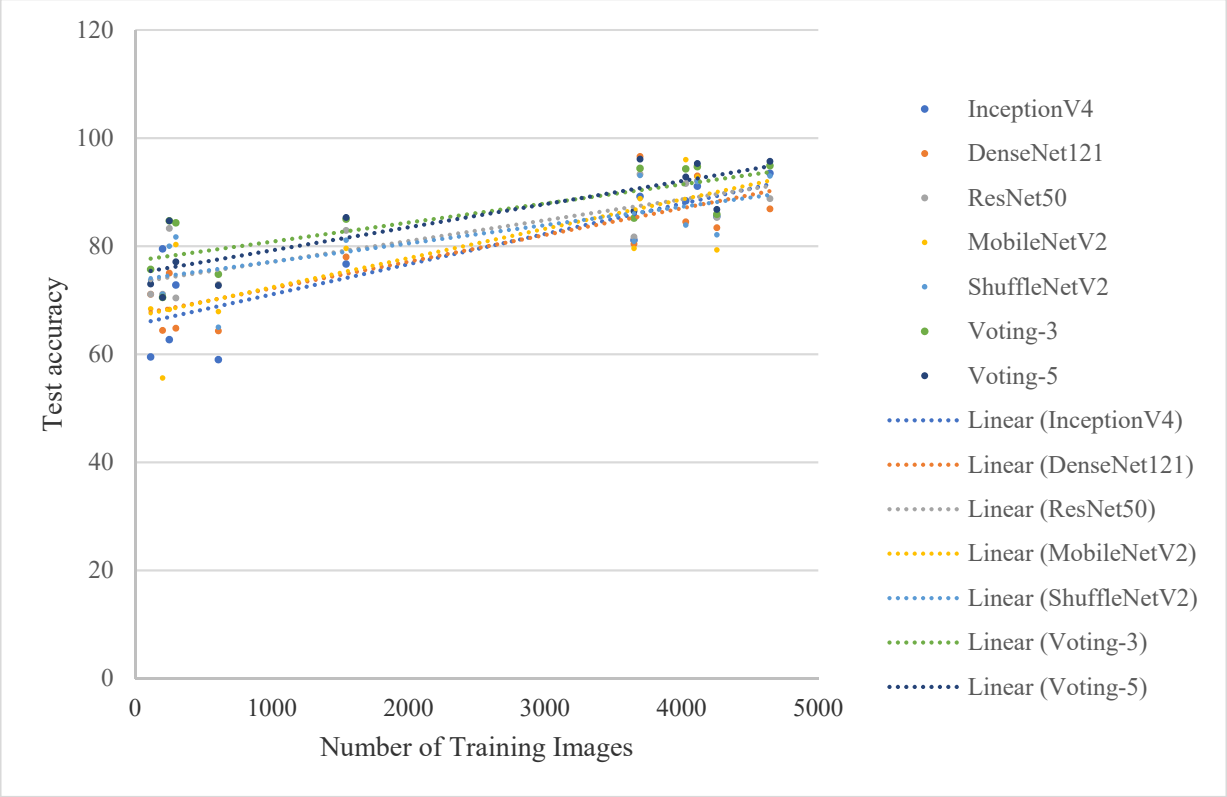


Figure 8. Top one-test accuracy per class against the number of training photos.

5. Conclusions

In this paper, a first large, public, multiclass road view crop photo dataset named iCrop is established for the development of crop type detection with deep learning. The iCrop dataset contains 12 types, representing the most popular crop types or farmland without crops, using 34,117 photos, and outlines the baseline performance for state-of-the-art deep convolutional neural networks. The results show that DCNN has good potential application in crop type detection from road view photos, and these computer vision models have room to improve when applied to imbalanced crop datasets. Small efficient ShuffleNetV2 models designed for mobile applications and embedded devices have better real-time performance (13FPS) and average accuracy (87.5%).

The deep learning network models with good accuracy and fast speed were selected, and the major voting decision fusion method was used to improve crop identification

accuracy. The results clearly demonstrate the superior accuracy of the proposed decision fusion over the other non-fusion-based methods in crop type detection of imbalanced road view photos dataset. The voting method achieved a mean accuracy of 91.1%, which can be leveraged to classify crop type in crowdsourcing road view photos online. The proposed fusion strategy increased overall accuracies by up to 3.5% compared to the best single CNN model.

We anticipate that our proposed method will save researchers valuable time that they would otherwise spend on the visual interpretation of larger number of photos. In the future, we plan to update the dataset and include more diverse crop types from worldwide areas to expand the scope of the iCrop. With the increasing use of drones in agriculture, drones can also be used as a tool to collect photos of crops to solve the problem of crops being obscured. Meanwhile, mobile technology has developed at an astonishing rate in the past few years and will continue to do so. With ever-improving computing performance and storage capacity on mobile devices, we consider it likely that highly accurate real-time crop classification via smartphones is just around the corner.

Author Contributions: Conceptualization, B.W. and F.W.; Data curation, F.W., M.Z., H.Z. and F.T.; Funding acquisition, B.W. and M.Z.; Methodology, B.W. and F.W.; Software, F.W.; Validation, M.Z.; Visualization, F.W., H.Z.; Writing—original draft, B.W. and F.W.; Writing—review & editing, F.W., M.Z., H.Z. and F.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research & Development Program of China (2016YFA0600301 & 2019YFE0126900) and the National Natural Science Foundation of China (41561144013 & 41861144019).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study can be found here: http://www.nwatch.top:8085/icrop.

Acknowledgments: Thanks to all volunteers Zhongyuan Li, Liang Zhu, Weiwei Zhu, Qiang Wang, Xinfeng Zhao, Qifeng Zhuang, and so on for providing data for this article.

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

Appendix A

Figure 1. A random sample of crop photos from the iCrop dataset. From the first row to the last row in turn, each photo contains: bare land, cotton, maize, peanut, rape, rice, sorghum, soybean, sunflower, tobacco, vegetable, wheat.

References

1. United Nations. Transforming Our World: The 2030 Agenda for Sustainable Development. Available online: https://sustainabledevelopment.un.org/post2015/transformingourworld (accessed on 19 March 2020).
2. United Nations. The Sustainable Development Goals Report 2019. Available online: https://unstats.un.org/sdgs/report/2019/ (accessed on 19 March 2020).
3. FAO. Investing in Data for the SDGs: Why Good Numbers Matter. 2019. Available online: http://www.fao.org/partnerships/resource-partners/news/news-article/en/c/1200471/ (accessed on 19 March 2020).
4. Rahman, M.; Di, L.; Yu, E.; Zhang, C.; Mohiuddin, H. In-Season Major Crop-Type Identification for US Cropland from Landsat Images Using Crop-Rotation Pattern and Progressive Data Classification. *Agriculture* **2019**, *9*, 17. [CrossRef]
5. Xiong, J.; Prasad, S.T.; Murali, K.G.; Pardhasaradhi, T.; Justin, P.; Russell, G.C.; Kamini, Y.; David, T. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. *ISPRS J. Photogramm. Remote Sens.* **2017**, *126*, 225–244. [CrossRef]
6. Zhang, X.; Wu, B.; Guillermo, P.C.; Zhang, M.; Chang, S.; Tian, F. Mapping up-to-date paddy rice extent at 10 m resolution in china through the integration of optical and synthetic aperture radar images. *Remote Sens.* **2018**, *10*, 1200. [CrossRef]
7. Fabrizio, R.; Fabrizio, P.; Olivier, A. S2 prototype LC map at 20 m of Africa 2016. Users Feedback Compendium Esa, 2018. Available online: https://un-spider.org/links-and-resources/data-sources/cci-land-cover-s2-prototype-land-cover-20m-map-africa (accessed on 10 October 2019).
8. Adam, J.O.; Prasad, S.T.; Pardhasaradhi, T.; Xiong, J.; Murali, K.G.; Russell, G.C.; Kamini, Y. Mapping cropland extent of Southeast and Northeast Asia using multi-year time-series Landsat 30-m data using a random forest classifier on the Google Earth Engine Cloud. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *81*, 110–124. [CrossRef]
9. Nabil, M.; Zhang, M.; Bofana, J.; Wu, B.; Stein, A.; Dong, T.; Zeng, H.; Shang, J. Assessing factors impacting the spatial discrepancy of remote sensing based cropland products: A case study in Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *85*, 102010. [CrossRef]
10. Fritz, S.; See, L.; Perger, C.; McCallum, L.; Schill, C.; Schepaschenko, D.; Duerauer, M.; Karner, M.; Dresel, C.; Laso-Bayas, J.C.; et al. A global dataset of crowdsourced land cover and land use reference data. *Sci. Data* **2017**, *4*, 170075. [CrossRef] [PubMed]
11. Leung, D.; Newsam, S. Exploring geotagged images for land-use classification. In Proceedings of the ACM multimedia 2012 Workshop on Geotagging and Its Applications in Multimedia (GeoMM'12), Nara, Japan, 29 October–2 November 2012; ACM Press: New York, NY, USA, 2012; pp. 3–8.
12. Wu, B.; Li, Q. Crop planting and type proportion method for crop acreage estimation of complex agricultural landscapes. *Int. J. Appl. Earth Obs. Geoinf.* **2012**, *16*, 101–112. [CrossRef]
13. Waldner, F.; Bellemans, N.; Hochman, Z.; Newby, T.; de Diego, A.; Santiago, R.V.; Sergey, B.; Mykola, L.; Nataliia, K.; Guerric, L.M.; et al. Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *80*, 82–93. [CrossRef]
14. Wu, B.; Meng, J.; Li, Q. Remote sensing-based global crop monitoring: Experiences with China's CropWatch system. *Int. J. Digit. Earth* **2013**. [CrossRef]
15. Wu, B.; Tian, F.; Zhang, M.; Zeng, H.; Zeng, Y. Cloud services with big data provide a solution for monitoring and tracking sustainable development goals. *Geogr. Sustain.* **2020**, *1*, 25–32. [CrossRef]
16. Antoniou, V.; Fonte, C.C.; See, L.; Estima, J.; Arsanjani, J.J.; Lupia, F.; Minghini, M.; Foody, G.; Fritz, S. Investigating the feasibility of geo-tagged photographs as sources of land cover input data. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 64. [CrossRef]
17. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
18. Mohanty, S.P.; Hughes, D.P.; Salathé, M. Using Deep Learning for Image-Based Plant Disease Detection. *Front. Plant Sci.* **2016**, 1419. [CrossRef] [PubMed]
19. Chebrolu, N.; Lottes, P.; Schaefer, A.; Winterhalter, W.; Burgard, W.; Stachniss, C. Agricultural robot dataset for plant classification, localization and mapping on sugar beet fields. *Int. J. Robot. Res.* **2017**. [CrossRef]
20. Raja, R.; Nguyen, T.T.; Slaughter, D.C.; Fennimore, S.A. Real-time weed-crop classification and localisation technique for robotic weed control in lettuce. *Biosyst. Eng.* **2020**, *192*, 257–274. [CrossRef]
21. Kamilaris, A.; Prenafeta-Boldu, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
22. Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; Li, F. ImageNet: A Large-Scale Hierarchical Image Database. In Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009.
23. iNaturalist.org. iNaturalist Research-Grade Observations. Occurrence Dataset 2019. Available online: https://doi.org/10.15468/ab3s5x (accessed on 10 October 2019).
24. Hughes, D.P.; Salathe, M. An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv* **2015**, arXiv:1511.08060.
25. Olsen, A.; Konovalov, D.A.; Philippa, B.; Ridd, P.; Wood, J.C.; Johns, J.; Banks, W.; Girgenti, B.; Kenny, O.; Whinney, J.; et al. DeepWeeds: A Multiclass Weed Species Image Dataset for Deep Learning. *Sci. Rep.* **2019**, *9*, 2058. [CrossRef]
26. Zheng, Y.; Kong, J.; Jin, X.; Wang, X.; Su, T.; Zuo, M. CropDeep: The Crop Vision Dataset for Deep-Learning-Based Classification and Detection in Precision Agriculture. *Sensors* **2019**, *19*, 1058. [CrossRef]
27. Mwebaze, E.; Gebru, T.; Frome, A.; Nsumba, S.; Tusubira, J. iCassava 2019 fine-grained visual categorization challenge. *arXiv* **2019**, arXiv:1908.02900.

28. Ringland, J.; Bohm, M.; Baek, S. Characterization of food cultivation along roadside transects with Google Street View imagery and deep learning. *Comput. Electron. Agric.* **2019**, *158*, 36–50. [CrossRef]
29. Deus, E.; Silva, J.S.; Catry, F.X.; Rocha, M.; Moreira, F. Google street view as an alternative method to car surveys in large-scale vegetation assessments. *Environ. Monit. Assess.* **2016**, *188*, 560.1–560.14. [CrossRef] [PubMed]
30. Yan, Y.; Ryu, Y. Exploring Google Street View with deep learning for crop type mapping. *ISPRS J. Photogramm. Remote Sens.* **2021**, *171*, 278–296. [CrossRef]
31. Wu, B.; Gommes, R.; Zhang, M.; Zeng, H.; Yan, N.; Zou, W.; Zheng, Y.; Zhang, N.; Chang, S.; Xing, Q.; et al. Global Crop Monitoring: A Satellite-Based Hierarchical Approach. *Remote Sens.* **2015**, *7*, 3907–3933. [CrossRef]
32. Tian, F.; Wu, B.; Zeng, H.; Zhang, X.; Xu, J. Efficient Identification of Corn Cultivation Area with Multitemporal Synthetic Aperture Radar and Optical Images in the Google Earth Engine Cloud Platform. *Remote Sens.* **2019**, *11*, 629. [CrossRef]
33. Fine, T.L. *Feedforward Neural Network Methodology*; Springer Science Business Media: Berlin, Germany, 2006.
34. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In *NeurIPS Proceedings*; Curran Associates Inc.: Red Hook, NY, USA, 2012; Volume 25.
35. Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D.; Stefanovic, D. Deep neural networks based recognition of plant diseases by leaf image classification. *Comput. Intell. Neurosci.* **2016**, *2016*. [CrossRef]
36. Sørensen, R.A.; Rasmussen, J.; Nielsen, J.; Jørgensen, R. Thistle Detection Using Convolutional Neural Networks. In Proceedings of the EFITA Congress, Montpellier, France, 2–6 July 2017.
37. Namin, S.T.; Esmaeilzadeh, M.; Najafi, M.; Brown, T.B.; Borevitz, J.O. Deep Phenotyping: Deep Learning for Temporal Phenotype/Genotype Classification. *Plant Methods* **2018**, *14*, 14.
38. Chen, S.W.; Shivakumar, S.S.; Dcunha, S.; Das, J.; Okon, E.; Qu, C.; Kumar, V. Counting apples and oranges with deep learning: A data-driven approach. *IEEE Rob. Autom. Lett.* **2017**, *2*, 781–788. [CrossRef]
39. Hiroya, M.; Yoshihide, S.; Toshikazu, S.; Takehiro, K.; Hiroshi, O. Road damage detection using deep neural networks with images captured through a smartphone. *arXiv* **2018**, arXiv:1801.09454.
40. Liu, S.; Tian, G.; Xu, Y. A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter. *Neurocomputing* **2019**, *338*, 191–206. [CrossRef]
41. Xue, D.X.; Zhang, R.; Feng, H.; Wang, Y.L. CNN-SVM for microvascular morphological type recognition with data augmentation. *J. Med. Biol. Eng.* **2016**, *36*, 755–764. [CrossRef] [PubMed]
42. Montserrat, D.M.; Lin, Q.; Allebach, J.; Delp, E. Training object detection and recognition cnn models using data augmentation. *Electron. Imaging* **2017**, *2017*, 27–36. [CrossRef]
43. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [CrossRef]
44. He, K.; Zhang, X.; Ren, S.; Sun, J. ResNet: Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
45. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv* **2016**, arXiv:1602.07261.
46. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. *CVPR. IEEE Comput. Soc.* **2017**, arXiv:1608.06993v5.
47. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. MobileNetV2: Inverted Residuals and Linear Bottlenecks. *arXiv* **2019**, arXiv:1801.04381v4.
48. Ma, N.; Zhang, X.; Zheng, H.; Sun, J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. *arXiv* **2018**, arXiv:1807.11164.
49. PaddlePaddle. Available online: https://github.com/PaddlePaddle (accessed on 19 July 2019).
50. Ruder, S. An overview of gradient descent optimization algorithms. *arXiv* **2016**, arXiv:1609.04747.
51. Sridar, P.; Kumar, A.; Quinton, A.; Nanan, R.; Kim, J.; Krishnakumar, R. Decision Fusion-Based Fetal Ultrasound Image Plane Classification Using Convolutional Neural Networks. *Ultrasound Med. Biol.* **2019**, *45*, 1259–1273. [CrossRef]
52. Hall, D.; McCool, C.; Dayoub, F.; Sunderhauf, N.; Upcroft, B. Evaluation of features for leaf classification in challenging conditions. In *Winter Conference on Applications of Computer Vision (WACV)*; IEEE: Waikoloa Beach, HI, USA, 2015; pp. 797–804.
53. Hajdu, A.; Hajdu, L.; Jonas, A.; Kovacs, L.; Toman, H. Generalizing the majority voting scheme to spatially constrained voting. *IEEE Trans. Image Process.* **2013**, *22*, 4182–4194. [CrossRef] [PubMed]
54. Sun, C.; Shrivastava, A.; Singh, S.; Gupta, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. *arXiv* **2017**, arXiv:1707.02968.
55. Hestness, J. Deep Learning Scaling is Predictable, Empirically. *arXiv* **2017**, arXiv:1712.00409.
56. Joulin, A. Learning Visual Features from Large Weakly Supervised Data. *arXiv* **2015**, arXiv:1511.02251.
57. Lei, S.; Zhang, H.; Wang, K.; Su, Z. How Training Data Affect the Accuracy and Robustness of Neural Networks for Image Classification. In Proceedings of the ICLR Conference, New Orleans, LA, USA, 6–9 May 2019.