

Article

Investigating the Use of Street-Level Imagery and Deep Learning to Produce In-Situ Crop Type Information

Fernando Orduna-Cabrera ^{1,*}, Marcial Sandoval-Gastelum ¹, Ian McCallum ¹, Linda See ¹, Steffen Fritz ¹, Santosh Karanam ¹, Tobias Sturn ¹, Valeria Javalera-Rincon ¹ and Felix F. Gonzalez-Navarro ²

¹ International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria; sandovalgaste@iiasa.ac.at (M.S.-G.); mccallum@iiasa.ac.at (I.M.); see@iiasa.ac.at (L.S.); fritz@iiasa.ac.at (S.F.); karanam@iiasa.ac.at (S.K.); sturn@iiasa.ac.at (T.S.); javalera@iiasa.ac.at (V.J.-R.)

² Instituto de Ingeniería, Universidad Autónoma de Baja California, Mexicali 21000, Mexico; fernando.gonzalez@uabc.edu.mx

* Correspondence: orduna@iiasa.ac.at; Tel.: +43-(0)-2236-807-227

Abstract: The creation of crop type maps from satellite data has proven challenging and is often impeded by a lack of accurate in situ data. Street-level imagery represents a new potential source of in situ data that may aid crop type mapping, but it requires automated algorithms to recognize the features of interest. This paper aims to demonstrate a method for crop type (i.e., maize, wheat and others) recognition from street-level imagery based on a convolutional neural network using a bottom-up approach. We trained the model with a highly accurate dataset of crowdsourced labelled street-level imagery using the Picture Pile application. The classification results achieved an AUC of 0.87 for wheat, 0.85 for maize and 0.73 for others. Given that wheat and maize are two of the most common food crops grown globally, combined with an ever-increasing amount of available street-level imagery, this approach could help address the need for improved global crop type monitoring. Challenges remain in addressing the noise aspect of street-level imagery (i.e., buildings, hedgerows, automobiles, etc.) and uncertainties due to differences in the time of day and location. Such an approach could also be applied to developing other in situ data sets from street-level imagery, e.g., for land use mapping or socioeconomic indicators.

Keywords: crop type recognition; deep learning; crowdsourcing; street-level imagery



Citation: Orduna-Cabrera, F.; Sandoval-Gastelum, M.; McCallum, I.; See, L.; Fritz, S.; Karanam, S.; Sturn, T.; Javalera-Rincon, V.; Gonzalez-Navarro, F.F. Investigating the Use of Street-Level Imagery and Deep Learning to Produce In-Situ Crop Type Information. *Geographies* **2023**, *3*, 563–573. <https://doi.org/10.3390/geographies3030029>

Academic Editor: Karoly Nemeth

Received: 11 July 2023

Revised: 18 August 2023

Accepted: 22 August 2023

Published: 30 August 2023



Copyright: © 2023 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

The spatial extent of cropland areas has been mapped extensively since the mid-eighties, as increasing numbers of satellites have been launched and higher-spatial- and -temporal-resolution imagery has become available. For example, cropland is provided as a land cover class in many global land cover products such as GLC-2000 [1], MODIS land cover [2] and ESA-CCI [3], among many others. More recently, a time series of higher-resolution cropland extent products has been produced using Landsat [4] at a 30 m resolution, while Sentinel-2 is now also being used for land cover mapping, including cropland extent at a 10 m resolution [5]. However, for monitoring food security at national-to-global scales, spatially explicit crop type maps are urgently needed [6]. With recent advances in analytical methods, data infrastructure and the availability of higher-resolution imagery, several recent studies have applied machine learning techniques for crop type recognition. Some of the most successful include support vector machines (SVMs) [7–9], random forests [9–12], decision trees [12–14], the maximum likelihood classifier (MLC) [11,15,16], artificial neural networks (ANNs) [11,17] and minimum distance (MD) [11]. Another example is the work undertaken by Mou et al. in [16], where the authors proposed a deep recurrent neural network (RNN) for hyperspectral image classification. The RNN model effectively analyzed hyperspectral pixels as sequential data and determined information categories via network reasoning. The specific application of

convolutional neural networks (CNN) in remote sensing for crop type recognition has also shown excellent performance [13,16,18–24].

Several recent research studies have achieved a higher accuracy in the learning phase because of the implementation of CNNs. For instance, Cai et al. in [18] introduced a methodology for the cost-effective and in-season classification of field-level crop types using common land units (CLUs) from the United States Department of Agriculture (USDA) to aggregate spectral information based on a time series. The authors built a deep-learning-based classification model based on deep neural networks (DNN). The research aimed to understand how different spatial and temporal features affected the classification performance. Their experiments also evaluated which input features were the most helpful in training the model and how various spatial and temporal factors affected the crop type classification. Castro et al. in [19] explored three approaches to improve the classification performance for land cover and crop type recognition in tropical areas using an image-stacking approach in combination with a CNN. Their findings outperformed the traditional system based on image stacking alone in terms of overall and class accuracy.

However, all of these applications require training data on the presence of different crop types, where a lack of such data is one challenge identified in many research papers. Many studies collect field-based data as part of the development of a training data set (see, e.g., [25]), which is costly and often not shared with the broader remote sensing community. For example, in the study by Wang et al. [20], local farmers contributed by utilizing a mobile application to take pictures and assign a label according to the crop type as the training data for crop type mapping. However, these examples are generally limited to small data sets. Another source of in situ data is from the LUCAS survey, which collects information at around 300 k locations across Europe [26]. However, the data are only collected every three years, they cover all land cover and land use types, not just crop types, and the data collection exercise is costly [27].

More recently, street-level imagery has become available for many areas around the world, e.g., from Google Street View and Baidu or as crowdsourced contributions through sites like Mapillary. However, much of the research involving street-level imagery has involved applications related to urban areas [28]. In contrast, D'Andrimont et al., (2018) [29] compared the amount of street-level imagery available for Europe to imagery available from the LUCAS survey as a potential source of training data, in particular for cropland mapping. They found that street-level imagery was available within 300 m of a LUCAS survey point for 9.4% of the EU territory, so it could provide additional training data. Focusing on the Netherlands, the authors then examined photographs from the Mapillary database for in situ crop type information. Of the 785 K photographs available, it was possible to identify some crops and to link these to agricultural parcels. However, the authors did not attempt to automatically classify the photographs for crop type.

The use of computer vision and the segmentation of street-level photographs is the subject of an active area of research. For example, Kang et al., (2018) [30] used images from Google Street View to classify building types, Cao et al., (2018) [31] created a land cover map of New York from combining street-level and aerial imagery, while a detailed urban map (of local climate zones) was developed by Cao et al., (2023) [32] using Google Street View. These and other similar studies are largely focused on the mapping of urban areas or features and have used pre-trained DL networks such as Places-CNN to first classify the images. The outputs from the pre-trained network are then often further classified for the urban features of interest. However, such a pre-trained network that predicts crop types or features that allow for crop types to be identified does not exist. Moreover, what is also missing is the crop type labels that would allow for such an existing pre-trained network to potentially be used.

To fill this gap, the aim of this paper is to determine the feasibility of using a tool like Picture Pile for the rapid labelling of geo-tagged street-level photographs for crop types in combination with a CNN utilizing a deep learning architecture [33] to classify the images. The novelty lies in combining these two tools, where the first provides high-

quality image labels and the second uses this information for the automatic classification of street-level photos for crop types. The images were labelled using Picture Pile as part of the Earth Challenge Food Insecurity crowdsourcing campaign [34]. In terms of crop species of global importance to food security, both maize and wheat (and related wheat crops) are crucial to meeting the global food demand [35]. Hence, the CNN was trained to recognize maize and wheat, referred to here as the Maize–Wheat–Other CNN (MWO CNN). Geo-tagged street-level images are noisy because in addition to maize and wheat, many objects such as cars, streets, buildings, and people present in the images make crop type classification more complex. Finally, we present the results from the CNN model regarding the performance in predicting crop types. Such a trained model could potentially generate a large in situ training data set on crop types given the large volume of street-view-level imagery now available. This, in turn, could then be used in classification algorithms to produce wall-to-wall crop type maps. The model is openly available at: <https://github.com/iiasa/CropTypeRecognition> (accessed on 25 August 2023).

2. Materials and Methods

2.1. Crowdsourced Labelling of Street-Level Imagery from Google Street View and Mapillary

A total of 10,776 street-level photographs were selected for use in this study, the majority of which were taken from Google Street View and a small amount taken from Mapillary. The bulk of the images were from France (Figure 1), as France is both representative of central European agriculture and provides an openly available land parcel information database for benchmarking. These images were then placed into the Picture Pile rapid image classification app [36] and labelled by volunteers. The quality of the images varied across the data set. There were excellent images that contained very clear, unobstructed pictures of roadside crops. There were also poorer-quality or noisier images, which contained objects such as cars, houses, etc., in addition to a crop field (Figure 2). Table 1 lists the total number of images used in this study along with the number of images used in the model training, test, and validation data sets.

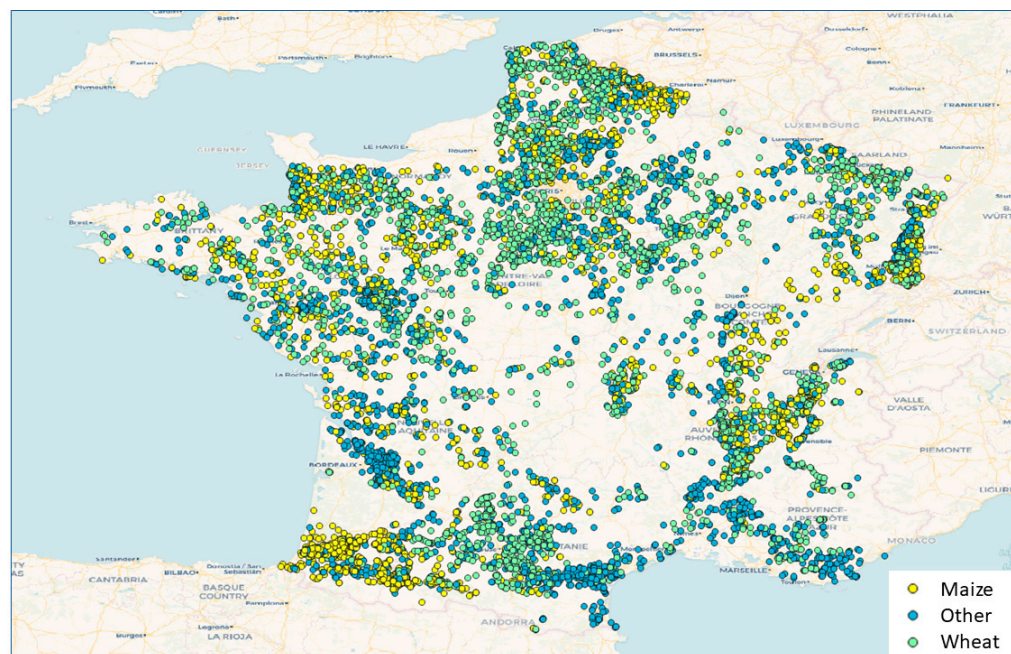


Figure 1. Locations of the majority of the 10,776 street-level images classified via crowdsourcing in Picture Pile, classified as either maize, wheat or other, in France.



Figure 2. Typical noisy street-level images containing crop species and additional non-crop objects such as roads, buildings, vehicles and trees (a–f). These images were classified by the crowd as (a) maize; (b) maize; (c) wheat; (d) wheat; (e) other and (f) wheat.

Table 1. The total number of classified street-level photographs used in the study, separated by crop type and usage by the CNN.

Crop Type	Total Images	Test	Training	Validation
Maize	3592	359	2873	360
Wheat	3592	359	2873	360
Other	3592	359	2873	360

In order to ensure accuracy of the crowdsourced image classifications, we created a set of 867 control-point images for the crop types of wheat, maize, sunflower, vineyards, sorghum, olive trees and other crops. Each of these images was classified between 5 and 8 times by different individuals. If a minimum of 5 classifications agreed, then we marked that image as a crowdsourced control image. At the end of the campaign, we compared the crowdsourced results with the Land Parcel Information System (LPIS) of France.

2.2. Development of a Deep Learning Model for Crop Type Detection

CNNs are a popular data mining technique for image recognition, first introduced by Fukushima [37]. The use of CNNs for object classification has been implemented in many domains, achieving a high efficiency and accuracy [38–40]. Figure 3 presents the CNN architecture used in the MWO model. A CNN (and any neural network) requires what is referred to as hyperparameter tuning. This is the determination of parameters such as the number of convolution layers (and hence the number of filters applied), the size of the filter, the stride length and the pooling method. These different settings for the MWO model are explained in the sections that follow.

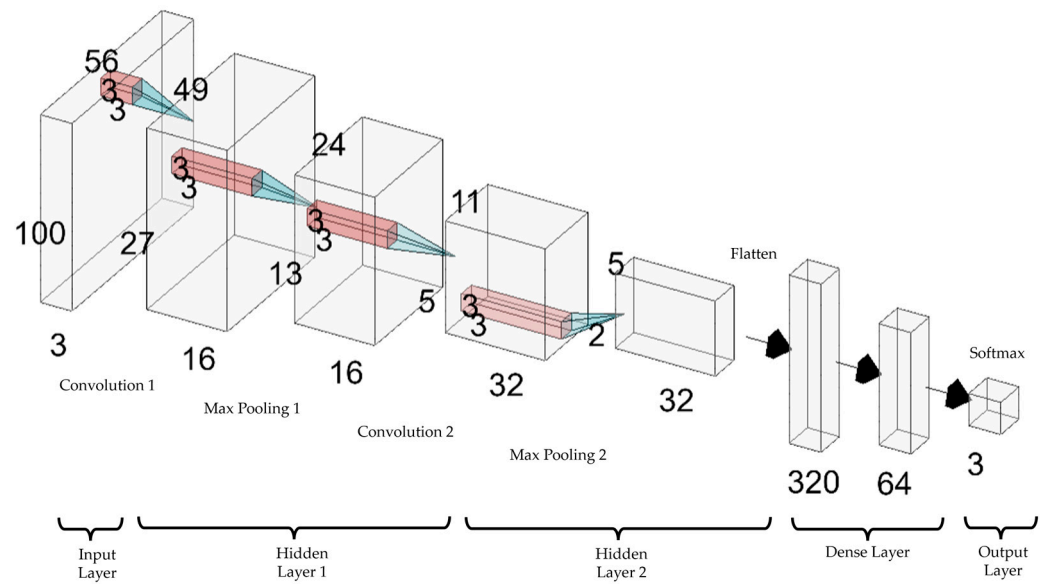


Figure 3. The CNN architecture used in the Maize–Wheat–Other (MWO) model.

In the proposed CNN architecture used in this study, we applied two convolution layers and the maximum operator for pooling. To arrive at this configuration, we tested different-sized filters, different stride lengths and different numbers of convolution layers (from two to a maximum of four due to the computational cost). The final architecture with the best performance had two convolution layers. Figure 3 shows the first convolutional output layer with 16 units followed by a second convolutional layer with 32 units. The experiments also gave better results with a small filter and a smaller stride size. The final filter size used was a 2 by 2 matrix, and the stride length was set to 2. For our architecture, we set the total number of hidden units in the dense layer to double the output size of the last convolution layer, i.e., with 64 units. Finally, we experimented with different loss and activation functions to find the combination that yielded the best performance, as shown in Table 2. After experimentation, we chose SOFTMAX as the last activation function and Y as the loss function to provide the best performance. Taking the output from the last dense layer as the input, the SOFTMAX function normalizes it into a probability distribution consisting of K probabilities proportional to the exponentials of the input numbers, where K is the number of classes (i.e., 3). After applying SOFTMAX, each component will be in the interval (0,1), and the sum of all the values is equal to 1. Once normalized, they can be interpreted as probabilities. Therefore, the more significant input components will correspond to larger probabilities.

Table 2. Different activation and loss functions for the experiments.

Activation Function	Loss Function
Relu	Mean squared error (MSE)
Identity	Poisson
Tanh	Mean squared logarithmic error
SOFTMAX	Cross entropy

Finally, we experimented with the number of training epochs at which the learning process is stabilized. We found that this occurred after 15 epochs and therefore used this as the maximum value.

2.3. Evaluation Methods

We evaluated the results from the crowdsourcing exercise using a standard confusion matrix and the overall accuracy. We evaluated the MWO CNN using three measures of

accuracy: the precision, which evaluates the number of positive classifications made by the model that were actually correct; the recall, which measures the amount of actual positive classifications that were correctly identified by the model [41] and the F1-score, which combines the precision and recall into an aggregated accuracy measure [42]. We also assessed the area under the receiver operating characteristic (ROC) curve [41], or the AUC, where the ROC is a plot of the false-positive rate (or specificity) against the recall. It allows one to determine how well the model performs across all classifications. We split the data set into 80% for training, 10% for testing and 10% for validation.

3. Results

3.1. Crowdsourcing

A total of 10,776 street-level images were classified for this study during a crowdsourcing campaign with the intention to create an accurate crop type training dataset. Approximately 600 people contributed around 76.5 K classifications, with many images classified multiple times in order to measure agreement. The participants classified the following crop types: wheat, maize, sunflower, vineyards, sorghum, olive trees and other crops.

Using the crowdsourcing classifications and the parcel information from the official French 2016–2019 Land Parcel Information System (LPIS), we computed a confusion matrix to examine the performance of the crowd [43]. Since each image was labelled by more than one person, we selected a sample of classifications from the database in which a minimum of eight classifications per location were collected and where there was a majority agreement, i.e., at least five classifications were of the same crop type. Finally, a total of 2049 images were used for comparison for an overall accuracy of 98.7%. The final confusion matrix is shown in Table 3.

Table 3. Confusion matrix with land parcel information in columns and volunteer classifications as rows.

	Wheat-Type Crop	Maize	Sunflower	Vineyard	Sorghum	Olive Trees	Other Crop	Total
Wheat-type crop	468	2	0	1	1	0	3	475
maize	2	589	0	0	0	0	0	591
Sunflower	0	1	46	0	0	0	2	49
Vineyard	0	1	0	939	0	0	0	940
Sorghum	0	4	0	0	1	0	0	5
Olive trees	0	0	0	0	0	0	0	0
Other crop	9	1	0	0	0	0	6	16
Total	484	598	46	941	2	0	11	0.986

3.2. MWO CNN

Table 4 shows the overall evaluation results for the MWO CNN, using noisy street-level images and recognizing three kinds of crop types (i.e., maize, wheat and other). The overall accuracy was 75.93%. Examining the F1-score (combining precision and recall), we found that wheat had the highest value (82.04), followed by maize (79.72) and other (64.88).

Table 4. Evaluation results for the MWO CNN by the maize, wheat and other classes.

Crop	Precision	Recall	F1	AUC
Maize	79.18	80.28	79.72	0.85
Wheat	77.67	86.94	82.04	0.87
Other	69.87	60.56	64.88	0.73

Figure 4 depicts the ROC values for each class. The ROC curve shows the trade-off between the precision and the specificity. As a reference, an ideal classifier should have a

high precision and a low specificity. The area under the ROC curve (AUC) is a measure of a classifier's overall performance, where a value of 1 indicates a perfect classifier (e.g., no wrong classification in all the test samples) and a value of 0.5 indicates that the classifier performs no better than random chance. Figure 4 depicts the model outcome; this was more accurate when classifying pictures of wheat crops, which was the class with the highest AUC of 0.87. The second-best performance corresponded to the class "Maize", with an AUC value of 0.85. As can be seen, the model struggled to classify images of the "Other" class, but still had values above 0.5.

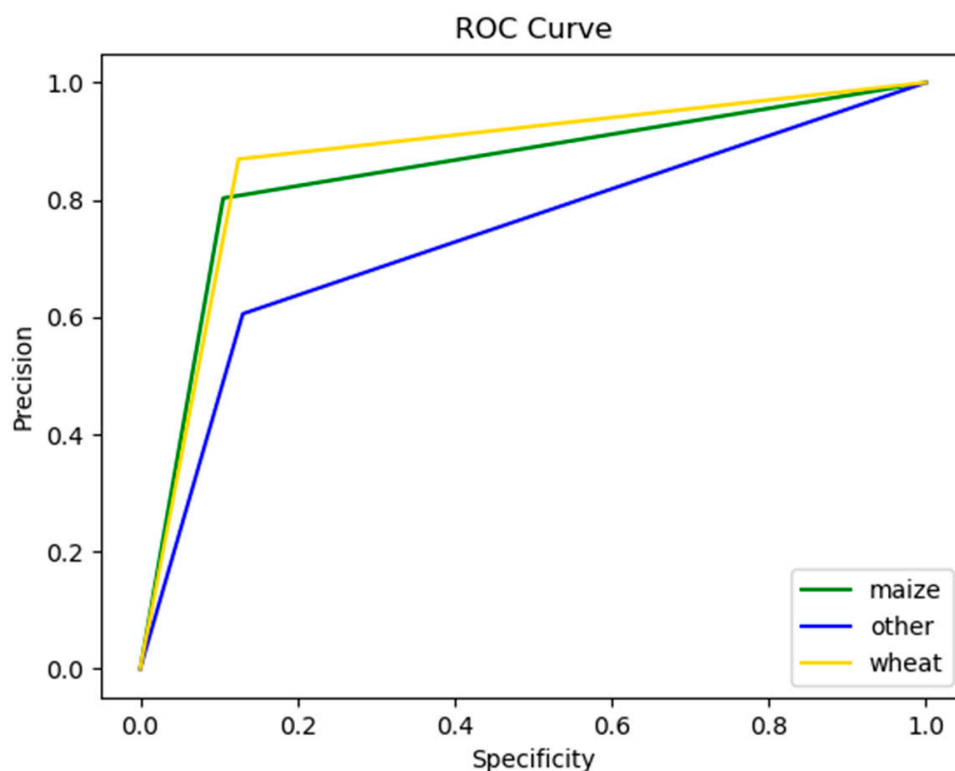


Figure 4. The receiver operating characteristic (ROC) curve for the MWO CNN model for maize, wheat and other classes.

After analyzing the nature of the images corresponding to the other class, we noticed that many contained non-crop objects and crop types other than maize or wheat. The similarity between the additional crops and those of interest (i.e., maize and wheat) made it difficult for the model to distinguish between them. As a result, we achieved a lower AUC performance of 0.73 for the other class.

4. Discussion

As crop type detection from satellite data has proven challenging due to a lack of training data, we explored the use of alternative methods for generating in situ data from street-level imagery. The first part of the method involved using Picture Pile to rapidly label the images using crowdsourcing. Picture Pile has been used in many different rapid image classification crowdsourcing campaigns [34], so considerable experience has been gained in producing a high-quality labelled data set. Hence, a high level of accuracy was achieved using the crowdsourcing approach (>95%). This rapid labeling approach could be used to build a very large image data set and then be used to create pre-trained networks such as those that already exist. The advantage would be that rather than the quite generic features that are currently identified by these pre-trained networks (e.g., grass), this network could focus specifically on major crop types.

We then introduced a deep learning architecture to classify noisy street-level images according to the following three classes: maize, wheat and other objects. In addition to the crops of interest, street-level imagery may include objects such as cars, roads, buildings, trees, people, other crops and more. Because of the nature of the viewing angle for street-level imagery, automatic classification can prove challenging, as the above-mentioned objects often obscure the view. This study differs from many others that have used street-level imagery because they first used a pre-trained classifier to segment the images into features. These features are then input to a neural network to learn other specific features of interest, e.g., building types [30] or local climate zones [32]. In contrast, in this study, the images were fed directly into a CNN and classified by crop type in one system. Moreover, such an open-source classifier does not currently exist, where much of the focus of street-level classification to date has been on urban areas [28].

Another limitation in the current approach is that the street-level imagery used here came from existing sources such as Google Street View and Mapillary opportunistically and were thus taken at different times of the day and from different geographical locations. Hence, there were additional uncertainties due to effects of shading, the sun angle, the camera angle or differences in brightness. However, the CNN model still produced a good performance despite these uncertainties. While it was not possible to replicate the level of accuracy achieved with the crowdsourcing approach, the MWO CNN model nevertheless produced initial results that are still promising (AUC of 0.87 for wheat and AUC of 0.85 for maize).

In the future, the model can be extended to other crop types, so this may improve the ability of the model to predict the 'other' class. Moreover, using a large labelled image set may help to further reduce these uncertainties and improve the model performance.

These initial results are promising owing to the vast potential of this data as an in situ data set for crop types. With additional improvements, classified street-level imagery could provide a powerful training data set for global satellite mapping. Crop type information combined with the image acquisition dates could be ingested into various global land products. For example, the World Cereal system for the high-resolution mapping of cereals and maize globally [44] would greatly benefit from such a model, which currently lacks in situ data in many parts of the world, particularly from Africa, South America and parts of Asia. Street-level imagery is increasing in volume, and there are other providers such as Baidu that have yet to be used in such a context.

5. Conclusions

We here introduced a convolutional neural network (CNN) architecture for a crop-type-recognition application using deep learning to classify two specific crop types in street-level images. The architecture demonstrated the application of CNN methods to recognize maize, wheat and other classes in street-level images.

The MWO CNN model was trained using more than 8000 crowdsourced street-level images from a Picture Pile campaign over France, where citizens contributed to labeling more than 10,000 images. The crowdsourced images were classified with an accuracy of >95%, ensuring that the model was trained on high-quality data. The MWO CNN model achieved an AUC of 0.87 for wheat and 0.85 for maize, the two most predominant crops grown globally. The other class achieved an AUC of 0.73. Given the specific viewing angle of street-level imagery, various non-crop structures impeded the view, which could have confounded the algorithms. In addition, street-level imagery is an opportunistic form of data, which is collected infrequently at different times of the day with varying sun and sensor angles. Nonetheless, this method holds great potential to massively increase our ability to globally track important crop types as the amount of street-level imagery continues to increase globally.

Such an approach can also be used to classify other types of in situ features from street-level imagery, e.g., socioeconomic indicators. Although street-level imagery has been used in land cover mapping, in particular in the mapping of urban features, land use

remains a difficult area to classify from remote sensing (satellite) imagery alone. Given the possibility to recognize different types of land use from street-level images and the advent of new hyperspectral satellites coming online in the next few years, this may greatly improve our ability to create detailed land use maps of the world.

Author Contributions: Conceptualization, F.O.-C. and M.S.-G.; methodology, F.O.-C. and M.S.-G.; software F.O.-C. and M.S.-G.; validation, I.M., L.S., S.F., S.K., V.J.-R. and F.F.G.-N.; formal analysis, F.O.-C., M.S.-G. and I.M.; investigation, L.S., V.J.-R. and S.F.; resources, S.K. and T.S.; data curation, F.O.-C., M.S.-G. and S.K.; writing—original draft preparation, F.O.-C. and M.S.-G.; writing—review and editing, I.M. and L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Open-Earth-Monitor Cyberinfrastructure project with funding from the European Union’s Horizon Europe research and innovation program under grant agreement no. 101059548.

Data Availability Statement: The data presented in this study are available on <https://github.com/iiasa/CropTypeRecognition> (accessed on 25 August 2023).

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

References

1. Joint Research Centre (European Commission); Fritz, S.; Bartholomé, E.; Belward, A. *Harmonisation, Mosaicing and Production of the Global Land Cover 2000 Database (Beta Version)*; Publications Office of European Union: Luxembourg, 2004.
2. Sulla-Menashe, D.; Gray, J.M.; Abercrombie, S.P.; Friedl, M.A. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sens. Environ.* **2019**, *222*, 183–194. [[CrossRef](#)]
3. Bontemps, S.; Boettcher, M.; Brockmann, C.; Kirches, G.; Lamarche, C.; Radoux, J.; Santoro, M.; Vanbogaert, E.; Wegmüller, U.; Herold, M.; et al. Multi-year global land cover mapping at 300 m and characterization for climate modelling: Achievements of the Land Cover component of the ESA Climate Change Initiative. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 323–328. [[CrossRef](#)]
4. Potapov, P.; Turubanova, S.; Hansen, M.C.; Tyukavina, A.; Zalles, V.; Khan, A.; Song, X.-P.; Pickens, A.; Shen, Q.; Cortez, J. Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century. *Nat. Food* **2021**, *3*, 19–28. [[CrossRef](#)] [[PubMed](#)]
5. Venter, Z.S.; Barton, D.N.; Chakraborty, T.; Simensen, T.; Singh, G. Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover. *Remote Sens.* **2022**, *14*, 4101. [[CrossRef](#)]
6. Fritz, S.; See, L.; Bayas, J.C.L.; Waldner, F.; Jacques, D.; Becker-Reshef, I.; Whitcraft, A.; Baruth, B.; Bonifacio, R.; Crutchfield, J.; et al. A comparison of global agricultural monitoring systems and current gaps. *Agric. Syst.* **2019**, *168*, 258–272. [[CrossRef](#)]
7. Sonobe, R.; Tani, H.; Wang, X.; Kobayashi, N.; Shimamura, H. Discrimination of crop types with TerraSAR-X-derived information. *Phys. Chem. Earth* **2015**, *83–84*, 2–13. [[CrossRef](#)]
8. Guo, J.; Wei, P.L.; Liu, J.; Jin, B.; Su, B.F.; Zhou, Z.S. Crop Classification Based on Differential Characteristics of H α Scattering Parameters for Multitemporal Quad- and Dual-Polarization SAR Images. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6111–6123. [[CrossRef](#)]
9. Feng, S.; Zhao, J.; Liu, T.; Zhang, H.; Zhang, Z.; Guo, X. Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3295–3306. [[CrossRef](#)]
10. Vuolo, F.; Neuwirth, M.; Immitzer, M.; Atzberger, C.; Ng, W.-T. How much does multi-temporal Sentinel-2 data improve crop type classification? *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *72*, 122–130. [[CrossRef](#)]
11. Zhan, Y.; Muhammad, S.; Hao, P.; Niu, Z. The effect of EVI time series density on crop classification accuracy. *Optik* **2018**, *157*, 1065–1072. [[CrossRef](#)]
12. Sonobe, R.; Tani, H.; Wang, X.; Kobayashi, N.; Shimamura, H. Random forest classification of crop type using multioral TerraSAR-X dual-polarimetric data. *Remote Sens. Lett.* **2014**, *5*, 157–164. [[CrossRef](#)]
13. McNairn, H.; Kross, A.; Lapen, D.; Caves, R.; Shang, J. Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *28*, 252–259. [[CrossRef](#)]
14. McNairn, H.; Shang, J.; Jiao, X.; Champagne, C. The Contribution of ALOS PALSAR Multipolarization and Polarimetric Data to Crop Classification. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3981–3992. [[CrossRef](#)]
15. Chen, Y.; Lu, D.; Moran, E.; Batistella, M.; Dutra, L.; Sanches, I.; da Silva, R.F.B.; Huang, J.; Luiz, A.J.B.; de Oliveira, M.A.F. Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *69*, 133–147. [[CrossRef](#)]
16. Kenduiywo, B.K.; Bargiel, D.; Soergel, U. Higher Order Dynamic Conditional Random Fields Ensemble for Crop Type Classification in Radar Images. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4638–4654. [[CrossRef](#)]

17. Júnior, C.C.; Schemmer, R.C.; Johann, J.A.; de Almeida Pereira, G.H.; Deppe, F.; Opazo, M.A.U.; Da Silva Junior, C.A. Artificial Neural Networks and Data Mining Techniques for Summer Crop Discrimination: A New Approach. *Can. J. Remote Sens.* **2019**, *45*, 16–25. [[CrossRef](#)]
18. Cai, Y.; Guan, K.; Peng, J.; Wang, S.; Seifert, C.; Wardlow, B.; Li, Z. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* **2018**, *210*, 35–47. [[CrossRef](#)]
19. Castro, J.D.B.; Feitoza, R.Q.; Rosa, L.C.L.; Diaz, P.M.A.; Sanches, I.D.A. A Comparative Analysis of Deep Learning Techniques for Sub-Tropical Crop Types Recognition from Multitemporal Optical/SAR Image Sequences. In Proceedings of the 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Niteroi, Brazil, 17–20 October 2017; pp. 382–389.
20. Wang, S.; Di Tommaso, S.; Faulkner, J.; Friedel, T.; Kennepohl, A.; Strey, R.; Lobell, D.B. Mapping Crop Types in Southeast India with Smartphone Crowdsourcing and Deep Learning. *Remote Sens.* **2020**, *12*, 2957. [[CrossRef](#)]
21. Wu, Y.; Wu, P.; Wu, Y.; Yang, H.; Wang, B. Remote Sensing Crop Recognition by Coupling Phenological Features and Off-Center Bayesian Deep Learning. *Remote Sens.* **2023**, *15*, 674. [[CrossRef](#)]
22. Pei, H.; Owari, T.; Tsuyuki, S.; Zhong, Y. Application of a Novel Multiscale Global Graph Convolutional Neural Network to Improve the Accuracy of Forest Type Classification Using Aerial Photographs. *Remote Sens.* **2023**, *15*, 1001. [[CrossRef](#)]
23. Li, G.; Han, W.; Dong, Y.; Zhai, X.; Huang, S.; Ma, W.; Cui, X.; Wang, Y. Multi-Year Crop Type Mapping Using Sentinel-2 Imagery and Deep Semantic Segmentation Algorithm in the Hetao Irrigation District in China. *Remote Sens.* **2023**, *15*, 875. [[CrossRef](#)]
24. Weilandt, F.; Behling, R.; Goncalves, R.; Madadi, A.; Richter, L.; Sanona, T.; Spengler, D.; Welsch, J. Early Crop Classification via Multi-Modal Satellite Data Fusion and Temporal Attention. *Remote Sens.* **2023**, *15*, 799. [[CrossRef](#)]
25. Fowler, J.; Waldner, F.; Hochman, Z. All pixels are useful, but some are more useful: Efficient in situ data collection for crop-type mapping using sequential exploration methods. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *91*, 102114. [[CrossRef](#)]
26. Eurostat. *LUCAS—Land Use and Land Cover Survey*; Eurostat Statistics Explained: Luxembourg, 2019.
27. Bayas, J.C.L.; See, L.; Bartl, H.; Sturn, T.; Karner, M.; Fraisl, D.; Moorthy, I.; Busch, M.; van der Velde, M.; Fritz, S. Crowdsourcing LUCAS: Citizens Generating Reference Land Cover and Land Use Data with a Mobile App. *Land* **2020**, *9*, 446. [[CrossRef](#)]
28. Biljecki, F.; Ito, K. Street view imagery in urban analytics and GIS: A review. *Landsc. Urban Plan.* **2021**, *215*, 104217. [[CrossRef](#)]
29. D’Andrimont, R.; Yordanov, M.; Lemoine, G.; Yoong, J.; Nikel, K.; der Velde, M. Crowdsourced Street-Level Imagery as a Potential Source of In-Situ Data for Crop Monitoring. *Land* **2018**, *7*, 127. [[CrossRef](#)]
30. Kang, J.; Körner, M.; Wang, Y.; Taubenböck, H.; Zhu, X.X. Building instance classification using street view images. *ISPRS J. Photogramm. Remote Sens.* **2018**, *145*, 44–59. [[CrossRef](#)]
31. Cao, R.; Zhu, J.; Tu, W.; Li, Q.; Cao, J.; Liu, B.; Zhang, Q.; Qiu, G. Integrating Aerial and Street View Images for Urban Land Use Classification. *Remote Sens.* **2018**, *10*, 1553. [[CrossRef](#)]
32. Cao, R.; Liao, C.; Li, Q.; Tu, W.; Zhu, R.; Luo, N.; Qiu, G.; Shi, W. Integrating satellite and street-level images for local climate zone mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *119*, 103323. [[CrossRef](#)]
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
34. Fraisl, D.; See, L.; Sturn, T.; MacFeely, S.; Bowser, A.; Campbell, J.; Moorthy, I.; Danylo, O.; McCallum, I.; Fritz, S. Demonstrating the potential of Picture Pile as a citizen science tool for SDG monitoring. *Environ. Sci. Policy* **2021**, *128*, 81–93. [[CrossRef](#)]
35. Erenstein, O.; Chamberlin, J.; Sonder, K. Estimating the global number and distribution of maize and wheat farms. *Glob. Food Secur.* **2021**, *30*, 100558. [[CrossRef](#)]
36. Danylo, O.; Moorthy, I.; Sturn, T.; See, L.; Bayas, J.-C.L.; Domian, D.; Fraisl, D.; Giovando, C.; Girardot, B.; Kapur, R.; et al. The picture pile tool for rapid image assessment: A demonstration using hurricane matthew. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *4*, 27–32. [[CrossRef](#)]
37. Fukushima, K. Neocognitron: A hierarchical neural network capable of visual pattern recognition. *Neural Netw.* **1988**, *1*, 119–130. [[CrossRef](#)]
38. Ciresan, D.C.; Meier, U.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. Flexible, High Performance Convolutional Neural Networks for Image Classification. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
39. Simard, P.Y.; Steinkraus, D.; Platt, J.C. Best practices for convolutional neural networks applied to visual document analysis. In Proceedings of the 7th International Conference on Document Analysis and Recognition, Edinburgh, UK, 3–6 August 2003; pp. 958–962.
40. Wiesner-Hanks, T.; Wu, H.; Stewart, E.; DeChant, C.; Kaczmar, N.; Lipson, H.; Gore, M.A.; Nelson, R.J. Millimeter-Level Plant Disease Detection From Aerial Photographs via Deep Learning and Crowdsourced Data. *Front. Plant Sci.* **2019**, *10*, 1550. [[CrossRef](#)] [[PubMed](#)]
41. Bowers, A.J.; Zhou, X. Receiver Operating Characteristic (ROC) Area Under the Curve (AUC): A Diagnostic Measure for Evaluating the Accuracy of Predictors of Education Outcomes. *J. Educ. Stud. Placed Risk JESPAR* **2019**, *24*, 20–46. [[CrossRef](#)]
42. Rahman, S.S.M.M.; Rafiq, F.B.; Toma, T.R.; Hossain, S.S.; Biplob, K.B.B. Performance Assessment of Multiple Machine Learning Classifiers for Detecting the Phishing URLs. In *Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19*; Springer: Singapore, 2020; pp. 285–296. [[CrossRef](#)]

43. Fritz, S.; Sturn, T.; Karner, M.; Karanam, S.; See, L.; Bayas, J.C.L.; McCallum, I. Crowdsourcing In-Situ Data Collection Using Gamification. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 254–257.
44. Van Tricht, K.; Degerickx, J.; Gilliams, S.; Zanaga, D.; Battude, M.; Grosu, A.; Brombacher, J.; Lesiv, M.; Bayas, J.C.L.; Karanam, S.; et al. WorldCereal: A dynamic open-source system for global-scale, seasonal, and reproducible crop and irrigation mapping. *Earth Syst. Sci. Data Discuss.* **2023**, *2023*, 1–36.

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