

Letter

# Deep Neural Networks and Transfer Learning for Food Crop Identification in UAV Images

Robert Chew <sup>1,\*</sup>, Jay Rineer <sup>1</sup>, Robert Beach <sup>1</sup>, Maggie O'Neil <sup>1</sup>, Noel Ujeneza <sup>2</sup>, Daniel Lapidus <sup>1</sup>, Thomas Miano <sup>1</sup>, Meghan Hegarty-Craver <sup>1</sup>, Jason Polly <sup>3</sup> and Dorota S. Temple <sup>1</sup>

<sup>1</sup> RTI International, Research Triangle Park, NC 27709, USA; jrin@rti.org (J.R.); rbeach@rti.org (R.B.); maggio@rti.org (M.O.); dlapidus@rti.org (D.L.); tmiano@rti.org (T.M.); mhegarty@rti.org (M.H.-C.); temple@rti.org (D.S.T.)

<sup>2</sup> Independent Agri-Consultant, Kigali 20093, Rwanda; nujeneza@gmail.com

<sup>3</sup> RTI International, Fort Collins, CO 80528, USA; jpolly@rti.org

\* Correspondence: rchew@rti.org; Tel.: +1-919-541-5823

Received: 25 December 2019; Accepted: 24 February 2020; Published: 26 February 2020



**Abstract:** Accurate projections of seasonal agricultural output are essential for improving food security. However, the collection of agricultural information through seasonal agricultural surveys is often not timely enough to inform public and private stakeholders about crop status during the growing season. Acquiring timely and accurate crop estimates can be particularly challenging in countries with predominately smallholder farms because of the large number of small plots, intense intercropping, and high diversity of crop types. In this study, we used RGB images collected from unmanned aerial vehicles (UAVs) flown in Rwanda to develop a deep learning algorithm for identifying crop types, specifically bananas, maize, and legumes, which are key strategic food crops in Rwandan agriculture. The model leverages advances in deep convolutional neural networks and transfer learning, employing the VGG16 architecture and the publicly accessible ImageNet dataset for pretraining. The developed model performs with an overall test set F1 of 0.86, with individual classes ranging from 0.49 (legumes) to 0.96 (bananas). Our findings suggest that although certain staple crops such as bananas and maize can be classified at this scale with high accuracy, crops involved in intercropping (legumes) can be difficult to identify consistently. We discuss the potential use cases for the developed model and recommend directions for future research in this area.

**Keywords:** remote sensing; crop analytics; crop mapping; UAVs; machine learning; convolutional neural networks; deep learning; smallholder systems

## 1. Introduction

### 1.1. Background and Motivation

Achieving food security for a growing global population will require significant advances in local capacity, market building, and technology. An important component of improving food security in the near term is better information on seasonal agricultural production, made available as early as possible during the growing season and updated as conditions change [1]. For instance, having timely access to information on crop progress by area can aid in the logistics of harvesting, processing, and marketing crops. Identifying regions where agricultural planting is delayed, or crop development is behind schedule, can help inform allocation of resources and improve preparation for mitigating food insecurity in those regions [2]. However, in many regions of the world, agricultural data lack the accuracy, centralization, structure, and consistency for farmers and government stakeholders to make timely decisions [3].

A lack of accurate and timely data is particularly pronounced for smallholder farms, the predominant agricultural system in traditionally food-insecure regions of Southeast Asia and sub-Saharan Africa [4,5]. Smallholder systems are not only the most common form of agriculture in the world, covering an estimated 75% of the world's agricultural area [4], but they also produce a large share of the food consumed by people living in the regions where they are grown [6]. For example, 50% of food calories consumed by people in sub-Saharan Africa are estimated to be from regional farms smaller than 5 ha [5]. Despite the importance of smallholder systems for addressing food security, important metrics such as crop productivity are often poorly measured and data at the subnational or field level are often unavailable [7]. Complicating this issue, smallholder plots in areas like sub-Saharan Africa have intense intercropping, with multiple different crop types being planted in close proximity [7–9] and large differences identified in planted crop distributions across different regions [8].

Remote sensing from satellites and unmanned aerial vehicles (UAVs) can augment ground surveys and improve the accuracy and timeliness of the agricultural information [10]. Modern publicly supported satellites, like the Sentinel series operated by the European Space Agency, provide wide-area coverage (100 km by 100 km image tiles) with revisit frequency of several days, but they have limited image resolution (ground resolution of 10 to 20 m depending on the band) [11]. UAVs can support satellite-based crop analytics by providing georeferenced images with much higher resolution, on the order of centimeters [12]. The analysis of UAV images has been used to provide information on crop types at a local scale [13] and to create ground-truth datasets for training of satellite-based models [14]. Thanks to the high resolution, crop identification has the potential to be effective not only for large monocropped fields but also in the smallholder agricultural systems described above.

In this case study, we use images collected from UAVs flown in Rwanda to develop a deep learning algorithm for identifying food crop types. We focus on bananas, maize, and legumes, which are key to food security in Rwanda. While most works in the literature using UAVs in smallholder agriculture focus on a single crop type (see Related Works below), this study modeled six common classes of land cover to help better understand the feasibility of a more comprehensive, high-resolution crop mapping for East African smallholder systems. Our objective is to better understand the promise and challenges of UAV agricultural classification methods in settings vastly different from large monocrop plots commonly adopted in industrial agriculture.

### 1.2. Related Works

The majority of remote sensing applications in the literature for smallholder systems rely on satellite data to classify crops. For East Africa, Jin et al. [8] used multispectral images from Sentinel-1 and Sentinel-2 to train a maize classifier and estimate crop yield in Kenya and Tanzania. Using a random forest model, they were able to classify satellite pixels of 10 m × 10 m ground area as “maize” or “non-maize” with an accuracy of 79% in Tanzania and 63% in Kenya. Likewise, Jin et al. [9] developed a three-class random forest model consisting of (1) maize crops, (2) other crops, and (3) non-crops for Kenya, resulting in an overall test set accuracy of 80%.

While the literature on UAV for precision agriculture in general is large [15,16], using UAVs to study crops in smallholder systems is still limited. Yang et al. [17] used a combination of spectral features, digital surface models, and texture analysis captured by UAV flights to identify rice lodging in Chiayi County, Taiwan. Using a decision tree classifier, they were able to obtain an accuracy of 96%, while also demonstrating additional image processing steps that could help minimize commission error. Jiang et al. [18] used a scale-space filtering algorithm with a Lab color transformation to develop a papaya tree detection model. Using imagery trained from UAV flights on a papaya farm in the Guangdong province of China, their model was able to detect papaya trees with an F1 score of 0.94. Nhamo et al. [19] used a combination of satellite modeling and UAV post-processing correction to detect irrigated areas in South Africa. This UAV post-processing correction yielded a substantial increase in accuracy compared with using satellite data alone (from 71% to 96%), providing a prime

example of how different imagery sources can provide complementary benefits. The study that most closely resembles ours in its goals is the work by Hall et al. [20], in which they used object-based image analysis (OBIA) image classification methods on UAV imagery to classify maize on smallholder farms in Ghana. Using both RGB and near-infrared (NIR) bands, they found classification accuracies for both single and mosaic images above 94%.

### 1.3. Our Approach

The objective of this study is to demonstrate a classification algorithm for identifying selected crops and other types of land cover in RGB images acquired by UAVs. In this paper, we leverage advances in deep convolutional neural networks (CNNs) [21] to identify selected crop types in UAV images. Because of their ability to effectively capture both local and global patterns in images, CNNs have advanced several areas of remote sensing for which high-resolution imagery is available, including hyperspectral image analysis [22–24], terrain surface classification with synthetic-aperture radar images [25–27], and 3D reconstruction [28,29]. In particular, CNNs are becoming the established method for scene classification [30–34], a task in which the goal is to assign an entire image into one of several distinct semantic classes. Due to this being analogous to our goal of classifying UAV images representing small areas (roughly 5 m × 5 m on the ground) to classes relevant to agriculture in Rwanda, we adopt CNNs and transfer learning as the modeling approach for this work. Though scene classification for identifying crops is rare in the literature (see [35] for a notable exception), the approach offers two operational advantages in comparison with more granular supervised segmentation-based models: (1) labeling images is more straightforward and less time consuming than creating bounding polygons around areas of interest (particularly in the presence of intercropping) and (2) CNNs designed for image recognition tasks are significantly less computationally expensive, an advantage in resource-constrained settings.

## 2. Materials and Methods

### 2.1. Study Area

The broad study area for this work is the country of Rwanda. Agriculture plays an important role in Rwanda's economy, accounting for an estimated 30.9% of the country's gross domestic product and 75.3% of the nation's labor force in 2017 [36]. Fields in Rwanda are often small (<1 ha) and heavily intercropped [37]; major crops include maize, beans, bananas, cassava, potatoes, and sweet potatoes. Rwanda has two main growing seasons: Season A extends from September through February, and Season B extends from March through June [38]. The start and end of the agricultural seasons can fluctuate, depending on the type of crop, region, and rainfall.

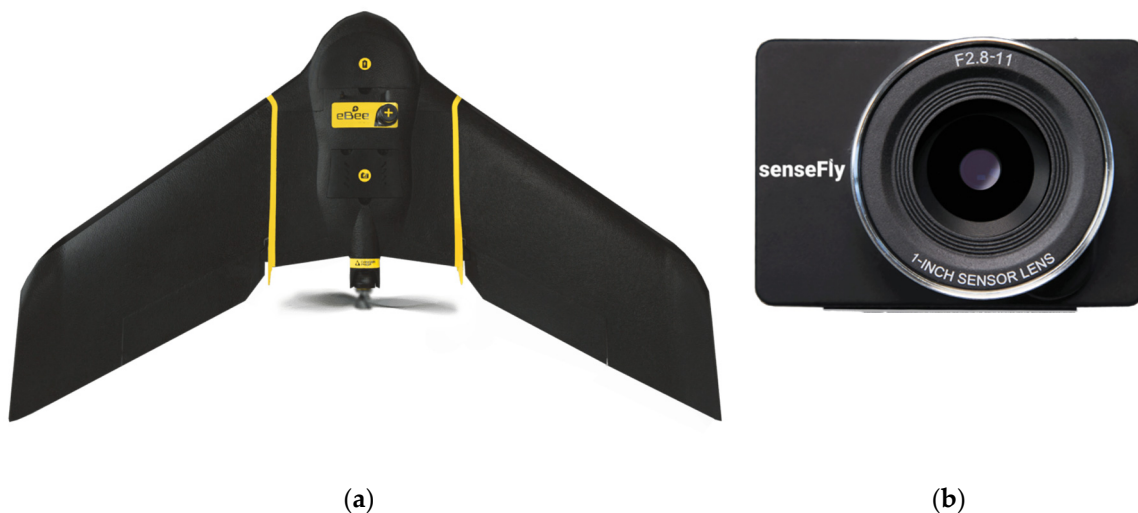
Table 1 shows the percentage of cultivated land occupied by each crop of interest for the districts in which the six UAV flights were conducted (see Section 2.4 for a full list and description of classes). The percentage of cultivated land for each crop type was determined from the 2019A Seasonal Agricultural Survey [38] and varies by district. The percentages are provided for the districts where UAV flights were conducted, as well as for the entire country for reference. The other labeled categories (forest, structure, and other) do not fall under cultivated land and are not described at a district level in the survey. For the entire country of Rwanda, 11% is forest and woodlands (excluding national parks) and 2.2% is urban areas or rural settlements.

**Table 1.** Percentage of cultivated land occupied by each crop of interest (Maize, Beans, and Bananas) for the districts in which the six UAV flights were conducted. Other classes of interest (Forest, Structure, and Other) are not reported at the district level in the Rwandan 2019A Seasonal Agricultural Survey.

Crop	District						Country
	Musanze	Karongi	Gakenke	Kamonyi	Nyaruguru	Gatsibo	Rwanda
Maize	27%	13%	18%	11%	11%	23%	16%
Beans	5%	28%	28%	24%	11%	25%	19%
Bananas	10%	15%	19%	24%	17%	29%	23%

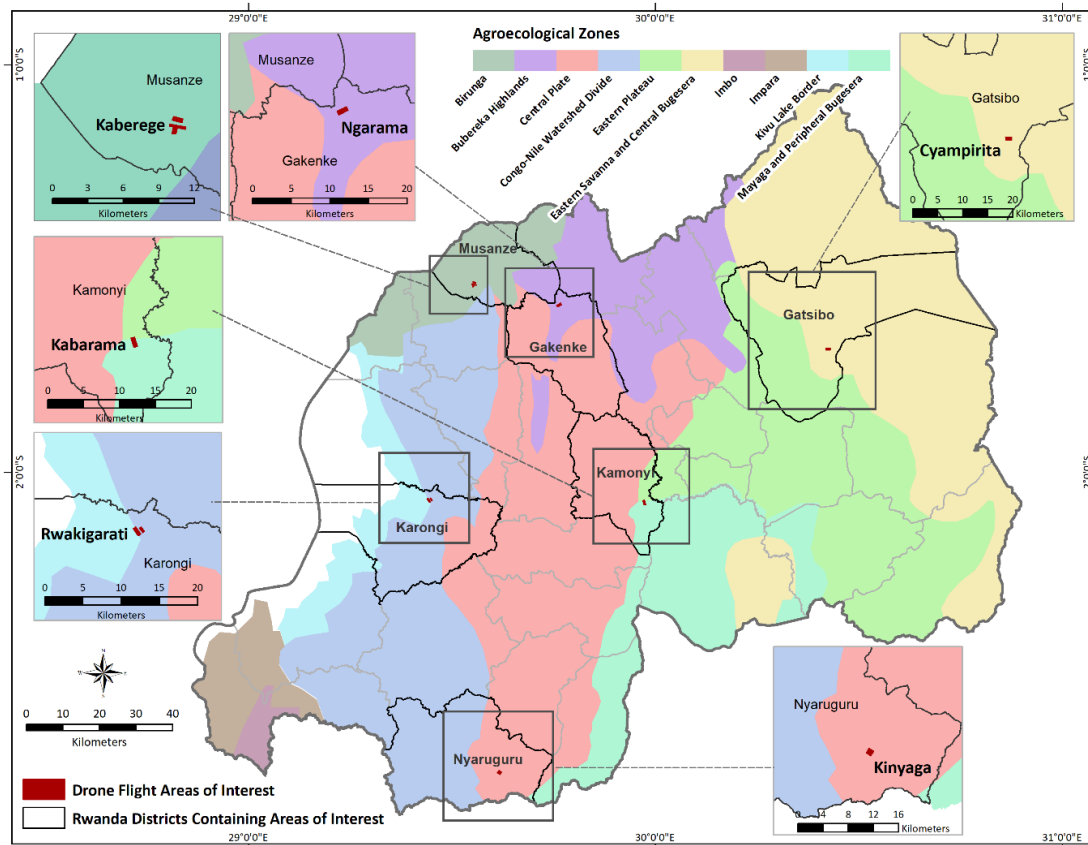
## 2.2. Data Collection

To develop training data, an in-country service provider, Charis Unmanned Aerial Solutions, used an eBee Plus UAV (senseFly SA, Cheseaux-sur-Lausanne, Switzerland) to capture UAV images (Figure 1). The eBee Plus was equipped with a GPS correction system based on the real time kinematic and post-processed kinematic technology that made it possible to georeference UAV-acquired images with survey-grade accuracy of 10 cm without the need for ground control points [39]. The UAV was equipped with a senseFly S.O.D.A. camera (senseFly SA, Cheseaux-sur-Lausanne, Switzerland), designed specifically for drone applications. This small, ultra-light, and fully configurable camera with built-in dust and shock protection features a 20 megapixel RGB sensor [40]. The flight plans were developed by Charis to obtain images with the ground resolution of 3 cm whenever possible; achieving this resolution required the UAV to fly at an altitude of 122 m above the ground level.



**Figure 1.** (a) A eBee Plus unmanned aerial vehicles (UAV) equipped with a (b) senseFly S.O.D.A. camera was used to collect RGB high-resolution images in the field, using the manufacture’s standard camera mount. Both the eBee Plus and S.O.D.A camera are manufactured by senseFly and are designed for cross-device compatibility (photos courtesy of senseFly.com).

To obtain training data, UAV flight sites were selected to represent a diversity in agroecological zones and cropping practices (both intercropping and monocropping) (Figure 2). The flights covered approximately 80 ha in each location and covered a mix of consolidated land use areas (relatively large, monocropped regions (Consolidated land use areas in Rwanda entail participating farmers consolidating some aspects of their production with neighboring farmers through cooperatives. They agree to grow a single priority crop, identified by the Ministry of Agriculture and Animal Resources, while retaining ownership of their individual parcels.)) and smaller, intercropped fields. The resulting georeferenced RGB images had a target resolution of 3 cm, although actual resolution varied as a result of terrain constraints requiring different flight heights.



**Figure 2.** Map of Rwanda with district boundaries (black), UAV flight sites (solid red polygons within zoomed area), and agroecological zones (various colors).

### 2.3. Data Labeling

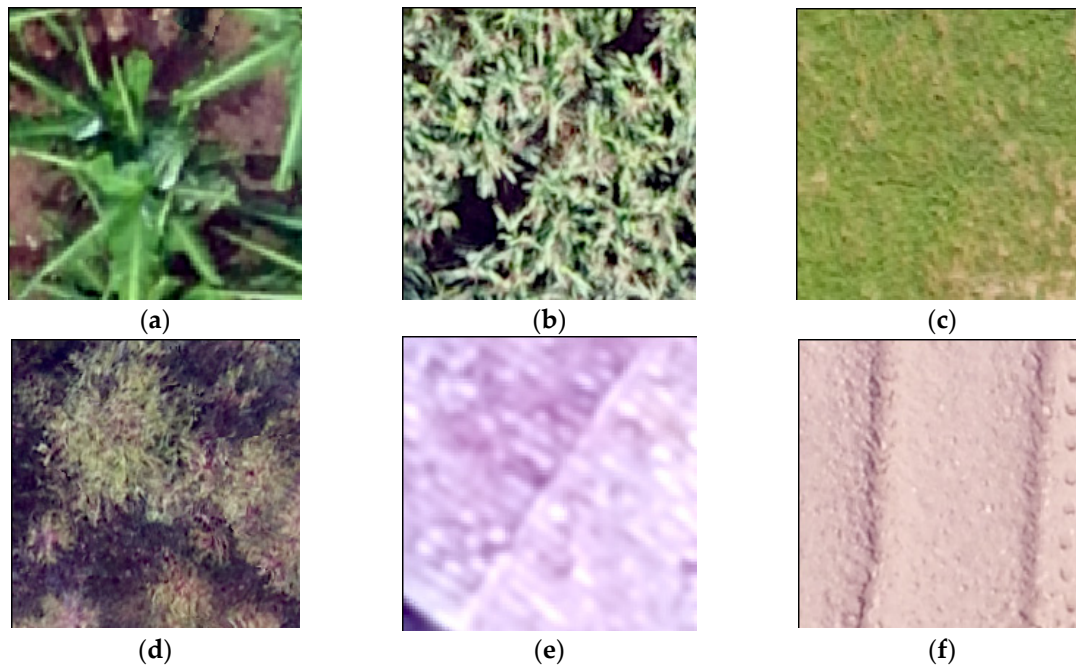
Traditionally, the process of crop labeling requires visiting agricultural areas using an electronic survey instrument with GPS location capture. Although laborious, this effort is often required because visual identification of crop types is difficult or impossible with satellite imagery. However, given the high resolution of our UAV images, we were able to use a web-based system to remotely label crops at greatly reduced effort. The viewer, constructed using ESRI's geographic information system platform, was designed to support multiple users simultaneously, tracking user and date of entry for all collected labels. Tools were provided within the viewer to support capture of labels by point location and by polygon delineation. For each point or polygon added by the user, a preconfigured menu of attribute options was provided. Polygon delineations were principally used to capture large monocrop areas, in which points were randomly sampled to stay consistent with direct point observations. To help ensure quality, a local Rwandan agricultural expert performed initial labeling of crops in the viewer and supervised a team of three independent labelers remotely.

For use in the classification models, the collected crop instances in the viewer were further processed into discrete images using ArcGIS, with the labeled point at the center of the new image. The resulting exported PNG images were  $200 \times 200$  pixels, with each pixel representing 2.5 cm to retain the resolution of the original UAV imagery. Prior to training the classification model, the final images were quality-checked by our in-country agricultural expert.

### 2.4. Data Description

Our final dataset consisted of six distinct classes: Banana, Maize, Legume, Forest, Structure, and a catch-all "Other" category (Figure 3). Each image is labeled with one of the six classes and represents roughly  $5 \text{ m}^2$  on the ground. The three agricultural classes (Banana, Maize, and Legume) were

chosen to represent priority food security crops that are both prevalent and important to livelihoods in Rwanda [41,42]. Common land cover types prevalent in the Rwandan countryside were included as additional classes (Forest and Structure). In cases when more than one class is present within the same image, labelers were instructed to label for the class occupying the majority of the image; implications for this choice are further expanded on in the Discussion section.



**Figure 3.** Example images of the six classes used for training and validating the CNNs; (a) Banana, (b) Maize, (c) Legume, (d) Forest, (e) Structure, and (f) Other.

After labeling, the images were randomly divided into a training set for model building (80.0%) and a test set for model evaluation (20.0%). The sampling into training and test sets was stratified to preserve the class ratios present in the full labeled dataset. Table 2 depicts the number of each class contributing to the training and test sets, respectively. Overall, the most heavily represented classes are Maize (32.2%), Banana (25.8%), and Forest (19.7%), while the Other (11.6%), Legume (5.6%), and Structure (5.1%) classes comprise relatively smaller shares.

**Table 2.** Training and test dataset split. The quantities in this table describe the number of images for each class across all six sites after being partitioned into training and test sets.

Class	# Training	# Test
Maize	1660	415
Banana	1329	332
Forest	1016	254
Other	600	150
Legume	290	73
Structure	265	66
<b>Total</b>	<b>5160</b>	<b>1290</b>

For modeling, RGB values were extracted for each pixel in the training and test images. The RGB values for each pixel in each image were extracted using the Python Imaging Library and were resized from 200 pixels  $\times$  200 pixels to 150 pixels  $\times$  150 pixels to match the pre-processing steps outlined in the paper for the model architecture explained below. Radiometric corrections were not performed on the RGB values in preparation for the algorithm development. In the literature, deep learning models

using high-resolution satellite or drone imagery for patch-based classification tend not to include a radiometric correction [30,32–34], likely because the algorithm relies on localized patterns of contrast (e.g., edges) rather than direct pixel-wise comparisons of color for analysis. Additionally, there is growing evidence that increasing the variation and distortion within training data images (a practice known as data augmentation) tends to help deep learning models improve performance [43].

### 2.5. Agricultural Classification Model

In this study, we used a machine learning approach for distinguishing UAV images that contain at least one of our six target classes. Specifically, we used a deep neural network (DNN), a type of artificial neural network that includes several chained layers of processing between the input (i.e., an image) and the output (i.e., a classification/label of the input image). Each processing layer amounts to a mathematical function that takes a tensor (i.e., n-dimensional matrix) from a previous layer as input, transforms it, and then outputs a new tensor. Various types of layers are commonly used in deep learning research. For example, convolutional layers create summary feature tensors (i.e., activation maps) of their input via convolution matrix operations. Pooling layers down-sample feature tensors to reduce their spatial size and reduce the total amount of parameters (i.e., weights) in the network. A common final layer for DNNs is the fully connected layer, which maps a feature tensor to a probability distribution of the target classes.

At a high level, a DNN is simply a series of functions that takes an input and returns a predicted label. The training process for a supervised DNN entails repeatedly passing labeled data through the network, using a loss function to evaluate how well the model performed at correctly identifying the true classes. The model optimizes for this loss function by computing the gradient of the loss function with respect to the model parameters, updating the model parameters iteratively during training to minimize the loss. A single test, evaluation, and update pass through the model is called an epoch, and the total training process typically requires several epochs to reach a point where the loss has reached a stable local minimum.

Training an exceedingly deep network from scratch was prohibitive for our sample size because most state-of-the-art deep learning models require fitting millions of model weights; our dataset sample size was insufficient for robustly fitting this many parameters. To address this challenge, we used a transfer learning approach [44,45] to initialize our model with weights from a CNN trained on a much larger dataset. The aim of transfer learning is to use a model trained in one source domain to help accelerate model building in a related target domain. In our case, we used the ImageNet dataset [46], a labeled image dataset consisting of over 14 million high-resolution images in 1000 categories as our source domain, and our labeled UAV images as the target domain. By using pretrained weights, our model was initialized with latent image features useful for distinguishing complex classes learned during the training process of the source model. We built off these by then training a model specifically for agricultural classification in Rwanda.

For our pretrained model, we used the VGG16 architecture [47] originally trained on the aforementioned ImageNet dataset. A DNN architecture is a blueprint of specific layers and parameters for those layers. VGG16 is a deep CNN model architecture first introduced in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC 2014), where it placed second on the “classification and localization” challenge task. This architecture remains popular today because of its relatively simple construction that involves alternating sets of convolutional and max pooling layers with a final set of fully connected layers.

To develop our agricultural classification model, we first ran our UAV images through the pretrained VGG16 model without the final layers to generate feature tensors for each image. This method is commonly referred to as feature extraction [19], because the outputs are feature tensors rather than class predictions. These feature tensors are created by applying operations useful for discriminating between classes in the source domain, creating a transformed representation of the original images that often improves classification. We used this as input to a shallow feed-forward network to classify

our specific categories. This smaller network consisted of one fully connected layer with sigmoid activation [48], a dropout layer to help with overfitting (probability of dropout = 0.5) [49], and a final output layer with a softmax activation function [48] to produce class probabilities for the six classes. At test time, the class with the highest modeled probability was assigned as the predicted class. The Adam optimizer [50] was used for gradient updates, and categorical cross-entropy was used as the loss function. The final model was trained with a batch size of 215 images and with the loss stabilizing at roughly 20 epochs.

### 3. Results

Table 3 summarizes the results of the classification model on the test set. The overall model recorded an F1, precision, recall, and accuracy each at 0.86, whereas the kappa coefficient was slightly lower at 0.82. The Banana, Maize, Forest, and Structure classes all performed well with an F1 score near or exceeding 0.90. However, images labeled as Legume or Other by human coders were more difficult to consistently classify, having test set F1 scores of 0.49 and 0.62, respectively. For a visual representation of this model's predictions over a UAV flight area, Figure 4 juxtaposes a stitched drone image panel of the Kabarama site (right) with the overlaid model predictions (left).

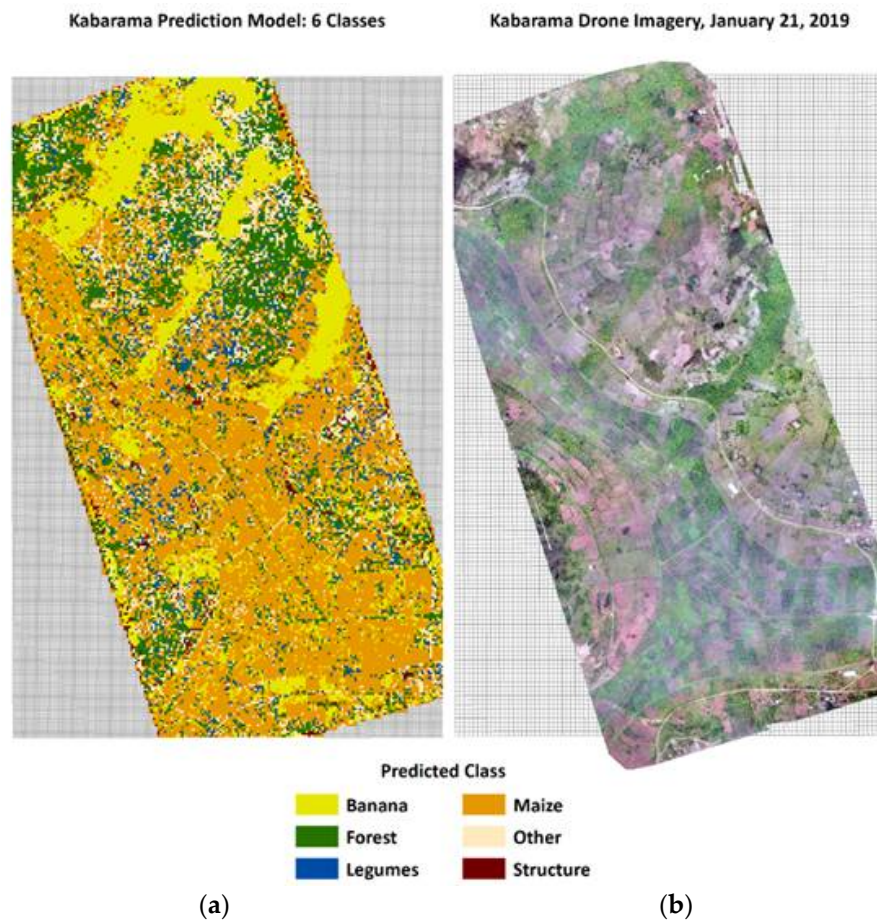
**Table 3.** Model test set evaluation metrics. To calculate accuracy and kappa scores per class, the multi-class confusion matrix (Table 4) was decomposed into six individual binary confusion matrices by changing all labels that were not the positive class to the negative class.

Class	F1 Score	Precision	Recall	Accuracy	Kappa
Banana	0.96	0.97	0.95	0.98	0.95
Forest	0.89	0.88	0.90	0.96	0.86
Legume	0.49	0.57	0.42	0.95	0.46
Maize	0.90	0.87	0.93	0.93	0.85
Other	0.62	0.67	0.58	0.92	0.58
Structure	0.89	0.84	0.95	0.99	0.89
<b>Overall</b>	<b>0.86</b>	<b>0.86</b>	<b>0.86</b>	<b>0.86</b>	<b>0.82</b>

**Table 4.** Confusion matrix for the test set. Correct predictions are highlighted in bold along the diagonal.

		Predicted						
		Banana	Forest	Legume	Maize	Other	Structure	
Actual	Banana	<b>315</b>	3	0	9	5	0	332
	Forest	0	<b>229</b>	4	6	12	3	254
	Legume	1	5	<b>31</b>	20	15	1	73
	Maize	3	6	9	<b>388</b>	9	0	415
	Other	5	17	10	23	<b>87</b>	8	150
	Structure	0	1	0	0	2	<b>63</b>	66
		324	261	54	446	130	75	





**Figure 4.** Comparison of (a) predicted crop classifications for the Kabarama site and (b) the stitched drone imagery for the Kabarama site.

We hypothesized that the lower model performance on these two groups is largely due to high within-class heterogeneity. Both the Legume and Other classes are an aggregation of multiple more specific categories; the Legume class contains instances of climbing beans, bush beans, and peas, and the Other class contains a diverse set of agricultural and land cover classes, including fallow land, water, cassava, and sweet potatoes. Furthermore, these classes have among the smallest number of training examples in the model (Legume,  $n = 290$ ; Other,  $n = 600$ ), due in large part to the low prevalence rates of the individual component classes in our study area. Even though there is no consensus for a minimum recommended sample size for effective transfer learning, classifiers tend to perform better with more labeled examples and balanced class ratios. Lastly, several images actually contain more than one class, preventing a clean single designation. This issue is particularly acute for the Legume class, in which images may also contain crops like maize in the same grid. The confusion matrix (Table 4) numerically demonstrates this interaction—20 of the images labeled legumes were “misclassified” as maize by the model. Figure 5 depicts such an example showing climbing beans sprouting between rows of maize crops.



**Figure 5.** Example images depicting intercropping for legumes. (a) Maize alone; (b) Legumes alone; (c) Legumes growing between rows of maize.

#### 4. Discussion

Our findings suggest that CNN-based classification models can be effective for identifying certain crops and land categories when trained on low-altitude UAV images. This finding is promising, given the challenging conditions posed by smallholder farm systems in Rwanda (e.g., intercropping, small plots, heterogeneous landscapes). In particular, our findings suggest that at least some important food security crops (bananas and maize), as well as traditional land cover and use categories (forested areas and built structures), can be detected with high accuracy. However, legumes were most difficult to consistently detect, possibly because of the diversity of legumes present in the labeled images, their less pronounced aerial profile when compared with above-ground crops such as maize, and/or their higher likelihood of intercropping. Likewise, the broad diversity of images in the Other class made consistent characterization difficult. While our initial hypothesis that dividing UAV imagery into small areas for modeling would help reduce misclassification error associated with intercropping, results from the confusion matrix suggest it still affects model performance even at this scale for certain key crops.

Though few studies closely resemble our work for direct comparison, our findings generally complement other related works in the literature. Lottes et al. [51] performed a classification of sugar beets and different weed types gathered from UAV flights in Germany and Switzerland. Using a random forest classifier trained on RGB images, they obtained an overall accuracy of 86% for predicted objects and 93% of the area correctly classified. Although they reported high detection rates for many plant types (e.g., 78% recall and 90% precision for sugar beets), they also experienced poor model performance for their catch-all class (“other-weeds”), obtaining a recall of 45%. Of the studies reviewed, the methodology of Hung et al. [52] is the most similar to our approach, although our categories of interest and geography differ. They used a feature learning–based approach on RGB images captured from low-flying UAVs to identify patches of different weed types (water hyacinth, serrated tussock, and tropical soda apple) in New South Wales, Australia. Searching over a grid of different pixel and window sizes, they found a best F1 score of 94.3% for water hyacinth, 92.9% for serrated tussock, and 72.2% for tropical soda apple. For studies focusing on UAV classification in smallholder systems, Hall et al. [20] used a combination of RGB and NIR imagery to classify maize in Ghana. Using an OBIA approach, they reported an overall accuracy above 94% compared with our F1 of 90% for maize. This finding suggests that incorporating additional sensor readings may help improve classification results, even in difficult smallholder environments.

##### 4.1. Study Limitations

Though promising, our study has several limitations. First, using high-resolution UAV imagery has its challenges. For instance, photogrammetry software can struggle to stitch overlapping UAV images in the presence of complex geometry (e.g., plants with thousands of branches and leaves). Generally, flights with high overlap and high flight altitude tend to help minimize distortion during reconstruction. Even though our flights had high overlap (75 to 80%), because of our relatively low flight

altitude, images for certain types of classes exhibit distortion (e.g., forests). This distortion at times made labeling more challenging, but we do not expect this problem to significantly affect the performance of the CNNs, because distortion is often added to input images purposely to prevent overfitting and aid in generalization [43]. Second, our results encompass images from only six nonrandom UAV flight sites totaling 480 ha. Although we selected sites for their diversity in agroecological zones and cropping patterns (both intercropping and monocropping), we cannot guarantee that these sites are fully representative of Rwandan farmland. Similarly, labeled crop instances were not chosen at random from the drone flight areas but, rather, were adaptively selected to ensure coverage of the crop types of interest. Even though this process was useful for generating training data, it may introduce selection bias if most areas in the drone flight areas are unlike the labeled images. A related caveat is that although we labeled only the classes that our in-country agricultural expert could identify from the UAV imagery, we did not compare our labels to independent ground truth from the field. This limitation is less severe for crop classification but will likely be important if this labeling approach is extended to yield estimation. Lastly, the issue of intercropping can make developing ground-truth labels and predictions challenging, even at the scale of  $5\text{ m} \times 5\text{ m}$  grids. Even though we required labelers to choose a single category for each image, as shown in Figure 4, several crops can and often appear within the same image. This problem is an ongoing issue noted by other research teams working in East Africa [8,9] and especially in Rwanda, which has among the most intensive intercropping systems and smallest plot sizes in the world. We believe that investigating effective methods of crop identification in the presence of intercropping is a fruitful area for future study.

#### 4.2. Future Research

Although identifying crop types from UAV images is useful for understanding local agricultural trends, scaling to entire districts or countries in the near future will likely require input from satellite data because flying drones multiple times across the extent of a large administrative unit may be cost prohibitive. However, we believe UAVs may provide a low-cost, high-throughput option for creating labeled data for machine learning models trained on lower resolution satellite imagery. This approach seems particularly promising given the effort required for developing ground-truth data using traditional field enumeration techniques. Future research could use computer-labeled UAV images as “noisy” ground-truth labels for crop classification models and compare the accuracy of such hybrid models with the accuracy of models based solely on labeling by human observers [10]. As the resolution of satellite imagery improves, similar approaches to remote labeling combined with deep learning models should become even more attractive for crop predictions of complex agricultural systems at scale.

Although a popular standard in the remote sensing literature, the ImageNet dataset used in the pretrained model does not contain aerial images. Future research could test the marginal benefit of using a model pretrained on a large dataset of satellite imagery, such as the Functional Map of the World [53]. Additionally, although classification using just RGB bands was effective for certain crops and land use categories in our model, future work can better understand how multispectral bands improve classification performance in this setting. An important operational consideration is how much labeled data are required to train models that generalize well across the intended population area. Although not covered in this study, performing cross-site validation and using diagnostics like learning curves (see [31] for an example) can help stakeholders better plan for future studies.

Lastly, future research should expand the relevant crop types for modeling to include others of strategic importance to countries in sub-Saharan Africa and prioritize modeling approaches that address the unique challenges of intercropping. This focus is key for many nations with a high proportion of smallholder farms, such as our study area in Rwanda, where intercropping systems account for 75% of the food production systems [38].

**Author Contributions:** Conceptualization, J.R., D.S.T., R.B., T.M., and R.C.; Methodology, R.C. and T.M.; Validation, N.U., M.H.-C., M.O., J.P., and D.L.; Formal Analysis, R.C.; Data Curation, M.O., N.U., M.H.-C., and J.R.; Writing—Original Draft Preparation, R.C., R.B., D.S.T., and T.M.; Writing—Review and Editing, N.U., M.H.-C., D.L., R.C., R.B., D.S.T., T.M., J.R., and M.O.; Visualization, R.C. and M.O.; Project Administration, D.S.T., R.B., D.L., and J.R.; Funding Acquisition, D.S.T., R.B., D.L., T.M., and J.R. All authors have read and agree to the published version of the manuscript.

**Funding:** This research received internal funding from RTI International under the RTI Grand Challenge initiative and the Social, Statistical, and Environmental Sciences unit publications fund.

**Acknowledgments:** We would like to acknowledge Jamie Cajka and Justine Allpress for helping process imagery for use in the AgViewer and model inputs, as well as Naomi Taylor, Justin Shelton, and Elizabeth Brown for their labeling and QC efforts.

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

## References

- Pérez-Escamilla, R. Food Security and the 2015–2030 Sustainable Development Goals: From Human to Planetary Health. *Curr. Dev. Nutr.* **2017**, *1*, e000513. [[CrossRef](#)] [[PubMed](#)]
- Brown, M.E.; Funk, C.C. Early Warning of Food Security Crises in Urban Areas: The Case of Harare, Zimbabwe, 2007. In *Geospatial Techniques in Urban Hazard and Disaster Analysis*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 229–241.
- Weersink, A.; Fraser, E.; Pannell, D.; Duncan, E.; Rotz, S. Opportunities and Challenges for Big Data in Agricultural and Environmental Analysis. *Annu. Rev. Resour. Econ.* **2018**, *10*, 19–37. [[CrossRef](#)]
- Lowder, S.K.; Scoet, J.; Raney, T. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Dev.* **2016**, *87*, 16–29. [[CrossRef](#)]
- Samberg, L.H.; Gerber, J.S.; Ramankutty, N.; Herrero, M.; West, P.C. Subnational distribution of average farm size and smallholder contributions to global food production. *Environ. Res. Lett.* **2016**, *11*, 124010. [[CrossRef](#)]
- HLPE. *Investing in Smallholder Agriculture for Food Security: A Report by the High Level Panel of Experts on Food Security and Nutrition of the Committee on World Food Security*; FAO: Roma, Italy, 2013.
- Burke, M.; Lobell, D.B. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 2189–2194. [[CrossRef](#)] [[PubMed](#)]
- Jin, Z.; Azzari, G.; You, C.; Di Tommaso, S.; Aston, S.; Burke, M.; Lobell, D.B. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sens. Environ.* **2019**, *228*, 115–128. [[CrossRef](#)]
- Jin, Z.; Azzari, G.; Burke, M.; Aston, S.; Lobell, D. Mapping smallholder yield heterogeneity at multiple scales in Eastern Africa. *Remote Sens.* **2017**, *9*, 931. [[CrossRef](#)]
- Temple, D.S.; Polly, J.S.; Hegarty-Craver, M.; Rineer, J.I.; Lapidus, D.; Austin, K.; Woodward, K.P.; Beach, R.H. The View From Above: Satellites Inform Decision-Making for Food Security. *RTI Press* **2019**, 10109. [[CrossRef](#)]
- Radiometric-Resolutions-Sentinel-2 MSI-User Guides-Sentinel Online. Available online: <https://earth.esa.int/web/sentinel/user-guides/sentinel-2-%20msi/resolutions/radiometric> (accessed on 25 December 2019).
- Turner, D.; Lucieer, A.; Wallace, L. Direct Georeferencing of Ultrahigh-Resolution UAV Imagery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2738–2745. [[CrossRef](#)]
- Tripicchio, P.; Satler, M.; Dabisias, G.; Ruffaldi, E.; Avizzano, C.A. Towards Smart Farming and Sustainable Agriculture with Drones. In Proceedings of the 2015 International Conference on Intelligent Environments, Prague, Czech Republic, 15–17 July 2015; pp. 140–143.
- Polly, J.; Hegarty-Craver, M.; Rineer, J.; O’Neil, M.; Lapidus, D.; Beach, R.; Temple, D.S. The use of Sentinel-1 and -2 data for monitoring maize production in Rwanda. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI. *Int. Soc. Opt. Photonics* **2019**, *11149*, 111491.
- Zhang, C.; Kovacs, J.M. The application of small unmanned aerial systems for precision agriculture: A review. *Precis. Agric.* **2012**, *13*, 693–712. [[CrossRef](#)]
- Adão, T.; Hruška, J.; Pádua, L.; Bessa, J.; Peres, E.; Morais, R.; Sousa, J.J. Hyperspectral Imaging: A Review on UAV-Based Sensors, Data Processing and Applications for Agriculture and Forestry. *Remote Sens.* **2017**, *9*, 1110. [[CrossRef](#)]
- Yang, M.D.; Huang, K.S.; Kuo, Y.H.; Tsai, H.P.; Lin, L.M. Spatial and Spectral Hybrid Image Classification for Rice Lodging Assessment through UAV Imagery. *Remote Sens.* **2017**, *9*, 583. [[CrossRef](#)]

18. Jiang, H.; Chen, S.; Li, D.; Wang, C.; Yang, J. Papaya Tree Detection with UAV Images Using a GPU-Accelerated Scale-Space Filtering Method. *Remote Sens.* **2017**, *9*, 721. [CrossRef]
19. [1403.6382] CNN Features Off-the-Shelf: An Astounding Baseline for Recognition. Available online: <https://arxiv.org/abs/1403.6382> (accessed on 14 November 2019).
20. Hall, O.; Dahlin, S.; Marstorp, H.; Archila Bustos, M.F.; Öborn, I.; Jirstrom, M. Classification of Maize in Complex Smallholder Farming Systems Using UAV Imagery. *Drones* **2018**, *2*, 22. [CrossRef]
21. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; pp. 1097–1105.
22. Zhang, M.; Li, W.; Du, Q. Diverse region-based CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2018**, *27*, 2623–2634. [CrossRef]
23. Chen, Y.; Jiang, H.; Li, C.; Jia, X.; Ghamisi, P. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [CrossRef]
24. Gao, Q.; Lim, S.; Jia, X. Hyperspectral image classification using convolutional neural networks and multiple feature learning. *Remote Sens.* **2018**, *10*, 299. [CrossRef]
25. Geng, J.; Wang, H.; Fan, J.; Ma, X. Deep Supervised and Contractive Neural Network for SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2442–2459. [CrossRef]
26. Geng, J.; Fan, J.; Wang, H.; Ma, X.; Li, B.; Chen, F. High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2351–2355. [CrossRef]
27. Zhou, Y.; Wang, H.; Xu, F.; Jin, Y.Q. Polarimetric SAR image classification using deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1935–1939. [CrossRef]
28. Zbontar, J.; LeCun, Y. Computing the stereo matching cost with a convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1592–1599.
29. Fischer, P.; Dosovitskiy, A.; Brox, T. Descriptor matching with convolutional neural networks: A comparison to sift. *arXiv* **2014**, arXiv:1405.5769.
30. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [CrossRef]
31. Chew, R.F.; Amer, S.; Jones, K.; Unangst, J.; Cajka, J.; Allpress, J.; Bruhn, M. Residential scene classification for gridded population sampling in developing countries using deep convolutional neural networks on satellite imagery. *Int. J. Health Geogr.* **2018**, *17*, 12. [CrossRef] [PubMed]
32. Cheng, G.; Han, J.; Lu, X. Remote sensing image scene classification: Benchmark and state of the art. *Proc. IEEE* **2017**, *105*, 1865–1883. [CrossRef]
33. Han, X.; Zhong, Y.; Cao, L.; Zhang, L. Pre-trained alexnet architecture with pyramid pooling and supervision for high spatial resolution remote sensing image scene classification. *Remote Sens.* **2017**, *9*, 848. [CrossRef]
34. Nogueira, K.; Penatti, O.A.B.; dos Santos, J.A. Towards better exploiting convolutional neural networks for remote sensing scene classification. *Pattern Recognit.* **2017**, *61*, 539–556. [CrossRef]
35. Kussul, N.; Lavreniuk, M.; Skakun, S.; Shelestov, A. Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 778–782. [CrossRef]
36. Africa: Rwanda—The World Factbook-Central Intelligence Agency. Available online: <https://www.cia.gov/library/publications/resources/the-world-factbook/geos/rw.html> (accessed on 25 September 2019).
37. Ali, D.A.; Deininger, K. *Is There a Farm-Size Productivity Relationship in African Agriculture? Evidence from Rwanda*; The World Bank: Washington, DC, USA, 2014.
38. National Institute of Statistics of Rwanda. *Seasonal Agricultural Survey: Season A*; National Institute of Statistics of Rwanda: Kigali City, Rwanda, 2019.
39. senseFly-eBee Plus. Available online: <https://www.sensefly.com/drone/abee-plus-survey-drone> (accessed on 25 December 2019).
40. senseFly-senseFly S.O.D.A. Available online: <https://www.sensefly.com/camera/sensefly-soda-photogrammetry-camera> (accessed on 25 December 2019).
41. Bank, T.W. *Rwanda—Fourth Transformation of Agriculture Sector Program and Second Phase of Program for Results Project*; The World Bank: Washington, DC, USA, 2018; pp. 1–115.
42. Cantore, N. *The Crop Intensification Program in Rwanda: A Sustainability Analysis*; ODI: London, UK, 2011.

43. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [[CrossRef](#)]
44. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
45. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 9. [[CrossRef](#)]
46. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Li, F.F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
47. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* **2014**, arXiv:1409.1556.
48. Duch, W.; Jankowski, N. Survey of neural transfer functions. *Neural Comput. Surv.* **1999**, *2*, 163–212.
49. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
50. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
51. Lottes, P.; Khanna, R.; Pfeifer, J.; Siegwart, R.; Stachniss, C. UAV-based crop and weed classification for smart farming. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 29 May–3 June 2017; pp. 3024–3031.
52. Hung, C.; Xu, Z.; Sukkarieh, S. Feature Learning Based Approach for Weed Classification Using High Resolution Aerial Images from a Digital Camera Mounted on a UAV. *Remote Sens.* **2014**, *6*, 12037–12054. [[CrossRef](#)]
53. Christie, G.; Fendley, N.; Wilson, J.; Mukherjee, R. Functional Map of the World. *arXiv* **2018**, arXiv:1711.07846.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).